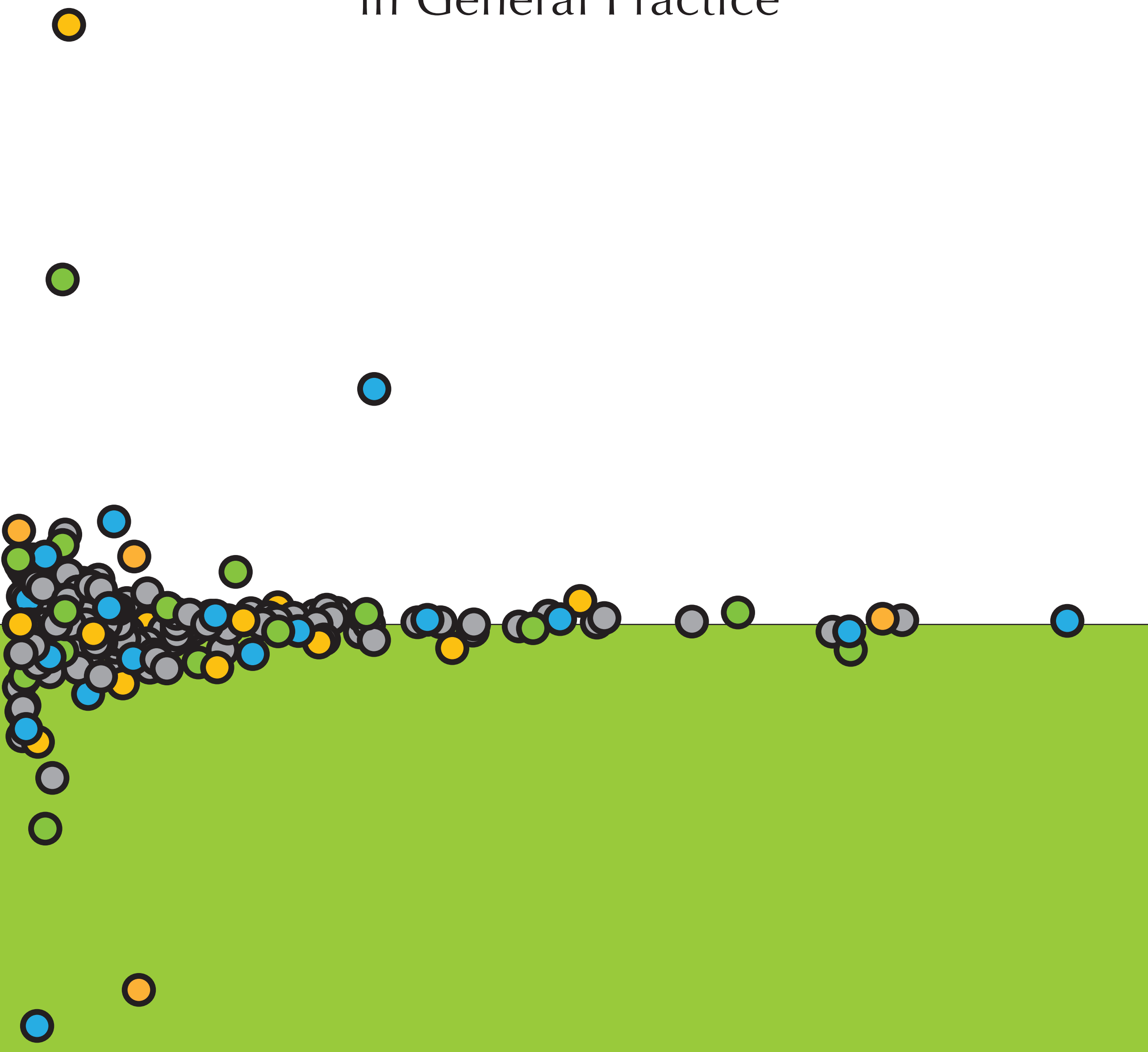
Epidemiologic Surveillance based on Electronic Medical Records in General Practice



Marion Biermans

Epidemiologic Surveillance based on Electronic Medical Records in General Practice

Epidemiologic Surveillance based on Electronic Medical Records in General Practice
Thesis, Radboud University Nijmegen, with summary in Dutch

Cover design and lay out: Janneke Tromp
Printed by: Gildeprint Drukkerijen, Enschede

ISBN: 978-90-9023762-6

The studies presented in this thesis were financially supported by ZonMw,
the Netherlands organisation for health research and development.

Financial support for the publication of this thesis was provided by PharmaPartners B.V.

Epidemiologic Surveillance based on Electronic Medical Records in General Practice

Een wetenschappelijke proeve op het gebied van de
Medische Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor aan de
Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
donderdag 8 januari 2009
om 15.30 uur precies

door

Marion Carolina Johanna Biermans

geboren op 16 maart 1967
te Eindhoven

Promotores

Prof. dr. P.F. de Vries Robbé

Prof. dr. ir. G.A. Zielhuis

Copromotor

Prof. dr. D.H. de Bakker (Universiteit van Tilburg)

Manuscriptcommissie

Prof. dr. J. van der Velden

Prof. dr. W.J.H.M. van den Bosch

Prof. dr. ir. A. Hasman (Universiteit van Amsterdam)

Contents

Chapter 1

General introduction

Surveillance is a core function of public health. It serves a range of public health needs, such as detecting unusual disease patterns, targeting interventions, evaluating prevention and control measures, and facilitating planning. Data derived from the electronic medical records (EMRs) of general practitioners (GPs) might be a valuable source for surveillance, because this data provide an indication of the health status of the general population. Furthermore, the approach is efficient since this data is already collected. Despite these advantages, currently there is no national, continuous surveillance system based on data from EMRs of general practices in the Netherlands. Why not? In contrast to the myth that data from EMRs is readily available by just pushing a button, complicated procedures are necessary to disclose and use data from EMRs for research purposes. In this thesis, we developed a strategy for continuous surveillance of health problems in general practice, based on data from EMRs of a national general practice database.

In this introductory chapter, we provide background information on epidemiologic surveillance and describe the history of surveillance. Then, we focus on surveillance based on EMRs in Dutch general practice and provide the aim, questions, and outline of this thesis.

BACKGROUND

Epidemiologic surveillance

The word surveillance is derived from the French *sur* (over) and *veiller* (to watch).[1] It was introduced into English at the time of the Napoleonic wars and meant: keeping a close watch over an individual or group of individuals in order to detect any subversive tendencies.[2] Within epidemiology, surveillance is the continuous and systematic process of collection, analysis, interpretation and dissemination of descriptive information for monitoring health problems.[3] The term 'epidemiologic surveillance' is used interchangeably with the term 'public health surveillance'. The first term emphasizes that surveillance is based on epidemiology, widely defined as the study of the distribution and determinants of disease frequency in man,[4] while the latter focuses on the use of the data for public health practice – i.e. information for action.[5]

Surveillance can be aimed both at the individual and at the population level. At the individual level, surveillance focuses on the identification of individuals with certain diseases in order to take action. In case of tuberculosis, for instance, surveillance aims to identify infectious individuals before they infect others, in an attempt to prevent an epidemic. At the population level, surveillance focuses on disease in populations, and serves to guide public health practice. In particular, surveillance is used for setting priorities, planning, and evaluation of public health programs. Furthermore, surveillance can serve as the starting point for more detailed epidemiologic investigations.[3;6] This thesis focuses on population-based epidemiologic surveillance.

The objective of population-based surveillance is to provide descriptive information regarding when and where health problems occur and who is affected, i.e. descriptive epidemiology of the time, place, and person. Basic epidemiologic measures that are used to compare disease occurrence at different times, in different locations or among different sub-groups, are the prevalence and incidence of disease. The prevalence is the proportion of a population that has a disease at a specific point in time, while the incidence refers to the occurrence of new cases of disease over a specified period of time. The incidence is either calculated as a proportion (the number of new cases divided by the population at risk) or as a rate (the number of new cases divided by the corresponding population time (a combination of population size and follow-up time)). The relation between incidence and prevalence can be understood by using a bathtub analogy in which the water level represents the stock (prevalence) which depends both on the inflow (incidence) and outflow (cure and mortality). When studying the etiology of disease, the incidence rather than the prevalence is useful, whereas the prevalence is more relevant than the incidence in applications such as planning for health resources.[7]

History of surveillance

The modern concept of surveillance has evolved along with changes in the way health information has been collected and used to guide public health practice. Table 1.1 provides a timeline of a number of key events in the history of surveillance.[2;3;5;8;9] The first public health use of surveillance occurred in 1348 during the bubonic plague, when public health authorities boarded ships in the port near the Republic of Venice to prevent persons suffering from plague illness from disembarking.[9] Beginning in the late 1600s, death reports were first used as a measure of the health of populations.[3]

Table 1.1 Key events in the history of surveillance

1300-1800	
1348	Public health authorities board ships near the Republic of Venice to prevent persons suffering from plague-like illness from disembarking.
1662	In London, John Graunt publishes *Natural and Political Observations Made upon the Bills of Mortality*, which defines disease-specific death counts and rates.
1680s	Gottfried Wilhelm von Leibnitz (Germany) calls for analysis of mortality in health planning.
1700s	The use of statistical methods to matters related to health spreads in Europe.
1741	Rhode Island requires tavern keepers to report contagious disease among their patrons.
1840-1850	Sir Edward Chadwick (UK) demonstrates that poverty and disease are closely related. Lemuel Shattuck (US) relates deaths, infant and maternal mortality, and communicable diseases to living conditions.
1839-1879	William Farr collects, analyses, and disseminates vital statistics for England and Wales to the authorities and to the general public.
1874	In Massachusetts, physicians reported prevalent diseases using a standard postcard reporting format.
Late 1800s	Compulsory reporting of communicable diseases, such as plague, cholera, and smallpox, in Europe, and in the US.
1900-1979	
1925	All states in the US begin participating in national morbidity reporting.
1935	First national health survey in the US.
1943	Danish Cancer Registry is established.
1946	Malaria Eradication Program starts in the US.
1955	Poliomyelitis vaccination program in the US.
1963	Alexander Langmuir (US) formulates the modern concept of surveillance in public health, focusing on populations rather than individuals.
1967	Networks of 'sentinel' general practitioners are established in England and Wales and in the Netherlands.
1968	Surveillance is used to target worldwide smallpox vaccination programs, leading to global eradication. The WHO broadens the concept of surveillance beyond communicable diseases.
1976	Surveillance in the US reveals a relation between Guillain-Barré syndrome and the swine influenza vaccine.
1979	Epidemic of toxic shock syndrome in the US; surveillance revealed that disease onset was typically during menstruation, providing a causal clue to tampon use.
1980-present	
1980	The introduction of microcomputers enables decentralized data analyses and electronic linkage of participants in surveillance networks.
Early 1980s	Surveillance of the acquired immunodeficiency syndrome (AIDS) in the US guides prevention activities before the etiology of disease is defined.
1990s, 2000s	Worldwide surveillance focuses on emerging and re-emerging diseases, such as AIDS, SARS, cholera, plague, and Ebola hemorrhagic fever. The internet is used increasingly to transmit and report data. Public concerns about privacy and confidentiality increase in parallel with the growth in information technology.
2001	Cases of anthrax associated with exposure of intentionally contaminated mail in the US lead to growth in 'syndromic surveillance' aimed at early detection of epidemics.
2002	Release of the National Electronic Disease Surveillance System (NEDSS) in the US, aimed at the ongoing, automatic capture and analysis of data that is already available electronically.

Two prominent names in the history of surveillance are Lemuel Shattuck and William Farr. In 1850, Shattuck published a landmark report that related death, infant and maternal mortality, and communicable diseases, to living conditions. William Farr is considered to be the founder of the basic principles of surveillance.[9] Box 1.1 provides a brief description of the work of William Farr.

Box 1.1 Where do we come from?

William Farr (1807-1883), a struggling general practitioner and medical journalist, was appointed compiler of statistical abstracts in 1838 when the General Register Office was established to collect information on births and deaths in England and Wales. Soon he became superintendent of the statistical department, a role in which he had access to unprecedented quantities of vital data, the aid of a staff, and even a mechanical computer, a model of a Babbage calculator. In this position, he developed the first national vital statistics system and assured its use as a surveillance instrument. To support his work, he developed a new system of nosology that provided the structure of the modern International Classification of Disease. Farr subscribed to the prevailing miasma theory that diseases were spread through atmospheric vapours and gathered statistical evidence that cholera was caused by miasma. Later on, however, he adopted John Snow's theory that cholera was caused by contaminated water. Farr fashioned statistical methods to describe the health of populations, such as age specific death rates, that are still in use today.[10-12]

In the 1900s, national systems for tracking certain diseases were established, and the term 'surveillance' evolved to describe a population-wide approach to monitor disease.[3] The Malaria Eradication Program, undertaken in 1946, demonstrated the importance of case definitions in surveillance. Concern that cases of malaria were being over-reported in the southern United States led to a requirement that case reports be documented. This change revealed that malaria had disappeared as an endemic disease.[9] The usefulness of an active outreach to identify cases (active surveillance) was clearly demonstrated during the poliomyelitis and smallpox vaccination programs. Cases of vaccine-associated poliomyelitis were shown to be limited to recipients of vaccine from one manufacturer, allowing continuation of the national immunization program. During the smallpox elimination program, surveillance was used to redirect vaccination efforts away from mass vaccinations to highly targeted vaccination programs, leading to global eradication.[3] In 1967, surveillance of disease based on general practice networks started in theNetherland (Continuous Morbidity Registration Nijmegen)[13] and in England and Wales (Weekly Returns Service),[14] both of which still operate today. The use of surveillance to provide clues for further investigation was evident when surveillance of illnesses after influenza vaccination revealed Guillan-Barré syndrome in vaccine recipients, which resulted in the cessation of the vaccine program for that year.[9] In addition, surveillance initiated after the detection of the epidemic of toxic shock syndrome provided a causal clue to tampon use.[3]

In the 1980s, microcomputers transformed surveillance practice enabling more efficient methods for data collection, management and analysis.[3] The role of surveillance in the development of prevention measures before the etiology of a disease is defined was demonstrated in the early 1980s, when surveillance of AIDS guided prevention activities before the human immunodeficiency virus (HIV) was discovered.[3] In the 1990s, surveillance activities started to focus on emerging and re-emerging diseases. The threat of re-emerging diseases became obvious when cholera reappeared in South America in 1991, followed by an outbreak of plague in India in 1994.[15] The 1990s also saw advances in the science of informatics and growth in the use of internet, which accelerated the automation of existing activities and enabled new approaches to surveillance, such as the use of sophisticated statistical methods.[9;16] In 2000, the new discipline of public health informatics evolved, defined as 'the systematic application of information and computer science and technology to public health practice, research, and learning'.[16] In 2001, the intentional dissemination of anthrax spores in the US led to the growth of 'syndromic surveillance systems' that emphasized the earliest possible detection of epidemics. These systems involve automation of nearly the entire process of surveillance.[3] In 2002, the Centers for Disease Control and Prevention released a surveillance system for capturing data that is already available electronically using internet-based transmission of data.[8]

The advances in informatics, as well as the concerns regarding potential terrorist events, had a substantial effect on the current concept of surveillance. Various surveillance paradigms began shifting in the 1990s and continue to evolve today (Table 1.2).

Table 1.2 Evolving paradigms of surveillance [17]

	Traditional	Modern
Focus	Diseases	Syndromes, indicators
Data sources	Clinicians/laboratories	Multiple
Data collection method	Manual	Electronic/existing
Approach	Categorical (disease-specific)	Cross-cutting
Cause of problem	Natural occurrence	Deliberately caused

The focus of surveillance broadened to encompass syndromes, new and multiple sources of data, and deliberately-caused diseases. Furthermore, the methodology moved from a fragmented and largely disease-specific approach to a cross-cutting and integrated one, which focuses on electronic capture of existing data instead of labor-intensive manual methods.[17]

Surveillance based on EMRs in Dutch general practice

Surveillance based on EMRs

Continuous monitoring of changes in morbidity based on data derived from EMRs is an important aspect of the modern concept of surveillance. Such analysis is computationally intensive and has not been possible until recently because the data were not available in electronic form, and the number of possible symptom/sign patterns was too large to manage.[16] Data from EMRs in general practice is particularly interesting for public health agencies, because morbidity in general practice reflects, to a considerable extent, the state of public health. In this thesis, we focus on surveillance of health problems based on EMRs of general practices in the Netherlands.

Dutch general practice

In the Netherlands nearly all citizens are registered by name in a general practice over a long period of time. The few exceptions are the homeless, illegal and uninsured people, and people staying in nursing homes; altogether these groups represent a tiny minority.[18] Anyone with health problems first contacts the GP, who functions as a gatekeeper of the Dutch healthcare system. The GP may refer patients to medical specialists, who report their findings to the GP. Almost all contacts (96%) are handled in general practice, while the remaining part (4%) results in referrals.[19] During out-of-hours, patients are referred to GP out-of-hours cooperatives, who report back to the patients' own GP. Until 2005, patients were either privately or publicly insured, and access to general practice was free for the latter. In 2006, these insurances merged into a new, privately administered, basic health insurance system.[18]

The Netherlands has a leading position in the use of EMRs in general practice. Practically all general practices were computerized in 2001.[20;21] Computerization of general practices in the Netherlands is regulated by the automation group within the Dutch College of General Practitioners (successor of the previous Working Group on Coordination of Information Automation (WCIA)), which develops software requirements for manufacturers of general practice information systems, evaluates those systems, and urges GPs to purchase only those systems that meet these requirements. The early and active role of this group in coordinating automation in general practice has contributed notably to the current leading position of Dutch GPs in using EMRs.[13;22]

National surveillance in Dutch general practice

The national public health agency (i.e. the National Institute for Public Health and the Environment (RIVM)) utilizes (among other sources) data derived from EMRs of five different continuous GP registries to describe the state of public health in the Netherlands. One registry is countrywide: the National Information Network of General Practice (LINH), the remaining four are regional.[18] In this thesis, we used the countrywide registry to establish a national, continuous surveillance system of health problems in general practice.

LINH has its roots in the first Dutch National Survey of General Practice (DNSGP-1), an extensive study among a random sample of 161 GPs who collected data using paper and pencil during a 3-month period in 1987/1988. Additional data was gathered using health interview surveys, GP questionnaires, and video-recordings. Diagnoses recorded by the GPs were coded using the International Classification of Primary Care (ICPC)[23] and grouped into episodes to calculate morbidity rates (i.e. prevalence and incidence rates). Box 1.2 provides definitions of the terms 'diagnosis' and 'episode' and explains why diagnoses need to be grouped into episodes in order to calculate morbidity rates. For a number of years, the DNSGP-1 provided enough information for policy and research.[24;25]

Box 1.2 Diagnoses, episodes, and morbidity rates

A **diagnosis** in general practice can refer to a symptom or complaint (symptom diagnosis), a syndrome (nosological diagnosis) or a disease (pathological/pathophysiological diagnosis).[26] In this thesis, we use the umbrella term diagnosis to refer to any of these categories. An **episode** of care refers to all encounters for the management of a specific health problem.[26] Diagnoses recorded in general practice need to be **grouped** into episodes in order to calculate morbidity rates. Consider, for example, a patient who consults the GP for a cough (diagnosis a), which develops into a pneumonia (diagnosis b) several days later. This health problem should be counted only once when estimating morbidity rates, namely as a case of pneumonia. To avoid double counting, the two diagnoses need to be grouped into one episode of care.

In the early 1990s, however, the Dutch Ministry of Health needed to monitor referrals from general practice to specialist medical care. By that time, personal computers had entered general practice, and LINH, a national network of computerized general practices, was established under the auspices of the Ministry of Health in 1991. This network is a joint project of two research organizations and two professional associations of GPs, i.e. the Netherlands Institute for Health Services Research (Nivel), the Scientific Institute for Quality of Healthcare (IQ healthcare), the Dutch College of General Practitioners (NHG), and the Dutch National Association of General Practitioners (LHV). Initially only referrals were collected (1992-1995); later on, prescriptions and contact counts were added (1996-1999).[24]

About ten years later, the Ministry of Health required not only valid information on the health of the population but also on the performance of general practice in the Netherlands, and decided to repeat the DNSGP-1. This time, data were collected electronically over a one-year period (2001/2002). LINH served as the basis of the DNSGP-2 and was extended to collect a dataset similar to the DNSGP-1.[24] Additional software was installed to record diagnoses of consultations within the EMRs. Afterwards, these diagnoses were manually grouped into episodes by medical coders.[27] An exception were six practices that used episode-oriented EMRs, a new generation of EMRs in which the GP records diagnoses directly into episodes.

After the DNSGP-2, the LINH practices continued to record data on consultations, prescriptions, and referrals, each with the associated diagnosis code. The manual grouping of episodes, however, was too costly and time consuming to continue on a routine basis. Therefore, continuous surveillance of all health problems recorded in EMRs in general practice was not possible. In this thesis, we explored the possibility to develop an automated method for constructing episodes, which would enable us to estimate national morbidity rates of all health problems in general practice. Furthermore, we focused on the development of methods to analyze and interpret these morbidity rates and to further investigate causal clues to detected trends. All in all, we aimed to develop an overall strategy for continuous surveillance of health problems on the basis of data collected within LINH.

AIM, QUESTIONS AND OUTLINE OF THE THESIS

Aim

The main aim of this thesis was to develop a useful strategy for continuous surveillance of health problems based on data from EMRs collected in a national general practice database. This overall strategy entails four subsequent steps:
a) constructing episodes
b) analyzing the data
c) interpreting the results
d) follow-up of detected trends.
In this thesis, we describe the development and the application of this strategy to monitor health problems in Dutch general practice over subsequent years.

Research questions

We formulated the following research questions, all of which refer to the use of data derived from EMRs in general practice. The term 'useful' in these questions denotes effective, efficient, and applicable.

1. What is a useful design of an automated method for constructing episodes?
2. What is the validity of this method for the purpose of estimating morbidity rates?
3. What is a useful general strategy for constructing episodes, analyzing, and interpreting the dataset at large?
4. What is a useful follow-up strategy for investigating causal clues to a detected trend in detail?
5. Which striking trends in the incidence of health problems are detected by application of the developed overall strategy over subsequent years?

Outline

The first two research questions of this thesis focus on the development (chapter 2) and validation (chapters 3 and 4) of an automated method for constructing episodes from diagnoses recorded in EMRs in general practice (EPICON). The implementation of this method in the national general practice database enabled us to actually monitor health problems based on EMRs in general practice. To this end, we needed a general surveillance strategy for the dataset at large. The third research question focuses on this general strategy (chapter 5), while the fourth question addresses a follow-up strategy to further investigate a detected trend (chapter 6). The last research question focuses on the practical value of the developed overall strategy (chapters 5 and 6). Because chapters 2 to 6 were written as separate articles for publication in scientific journals there is some overlap between the chapters, especially in the description of EPICON. This thesis concludes with chapter 7, which discusses the main findings, the methodological issues, and the implications of this thesis.

REFERENCES

1. Merriam-Webster's Collegiate Dictionary. 11th ed. Springfield MA: Merriam-Webster Inc; 2003.
2. Eylenbosch WJ, Noah ND. Surveillance in Health and Disease. Oxford: Oxford University Press; 1988.
3. Beuhler JW. Surveillance. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3th ed. Philadelphia: Lippincott Williams & Wilkins; 2008. 459-480.
4. MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston: Little, Brown and Company; 1970.
5. Thacker SB, Gregg MB. Implementing the concepts of William Farr: the contributions of Alexander D. Langmuir to public health surveillance and communications. Am J Epidemiol. 1996;144:S23-S27.
6. Stoto MA. Public health surveillance in the twenty-first century: achieving population health goals while protecting individuals' privacy and confidentiality. Georgetown Law J. 2008;96:703-719.
7. Greenland S. Rothman KJ. Measures of occurrence. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3th ed. Philadelphia: Lippincott Williams & Wilkins; 2008. 32-50.
8. Centers for Disease Control and Prevention. National Electronic Disease Surveillance System [document on the Internet]. Atlanta: CDC; 2008 [updated 2008 Jun 16; cited 2008 Jun 17]. Available from: http://www.cdc.gov/nedss.
9. Thacker SB, Berkelman RL. History of public health surveillance. In: Halperin W, Baker EL, Monson RR, editors. Public Health Surveillance. New York: John Wiley & Sons Inc; 1992. 1-15.
10. Lilienfeld DE. Celebration: William Farr (1807-1883) - an appreciation on the 200th anniversary of his birth. Int J Epidemiol. 2007;36:985-987.
11. Langmuir AD. William Farr: Founder of modern concepts of surveillance. Int J Epidemiol. 1976;5:13-18.
12. Eyler JM. The conceptual origins of William Farr's Epidemiology: Numerical methods and social thoughts in the 1830s. In: Lilienfeld AM, editor. Times, places, and persons. Aspects of the history of epidemiology. Baltimore: John Hopkins University Press; 1980. 1-21.
13. Höppener P. Automatisering en wetenschappelijk onderzoek in de huisartspraktijk. De mogelijkheden en merites van een nieuw instrument (Automation and scientific research in general practice. The possibilities and merits of a new instrument). PhD [dissertation]. Maastricht: Maastricht University; 1990.
14. Health Protection Agency. Primary Care Surveillance - RCGP Weekly Returns Service [document on the Internet]. London: HPA; 2008 [updated 2007 Nov; cited 2008 Jun 17]. Available from: http://www.hpa.org.uk/webw/HPAweb&Page&HPAwebAutoListName/Page/1203582641405?p=1203582641405.

15. Calain P. Exploring the international arena of global public health surveillance. Health Policy Plan. 2007;22:2-12.
16. Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, Kilbourne EM. Public health informatics: improving and transforming public health in the information age. J Public Health Manag Pract. 2000;6:67-75.
17. Koo D. Leveraging syndromic surveillance. J Public Health Manag Pract. 2005;11:181-183.
18. Gijsen R, Poos MJJC. Using registries in general practice to estimate countrywide morbidity in The Netherlands. Public Health. 2006;120:923-936.
19. Cardol M, De Bakker D, Westert G. The activities of general practitioners: are they still gatekeepers? In: Westert GP, Jabaaij L, Schellevis FG, editors. Morbidity, performance and quality in primary care. Dutch general practice on stage. Oxford: Radcliffe Publishing; 2006. 107-114.
20. European physicians especially in Sweden, Netherlands and Denmark, lead U.S. in use of electronic medical records. Health Care Research 2002;2:1-3.
21. Burt CW, Hing E, Woodwell D. Electronic medical record use by office-based physicians: United States, 2005 [document on the internet]. Hyattsville: National Center for Health Statistics; 2008 [updated 2007 Jan 11; cited 2008 Jun 22]. Available from: http://www.cdc.gov/nchs/products/pubs/pubd/hestats/electronic/electronic.htm.
22. Van der Lei J, Duisterhout JS, Westerhof HP, Van der Does E, Cromme PVM, Boon WM, et al. The introduction of computer-based patient records in the Netherlands. Ann Intern Med. 1993;119:1036-1041.
23. Lamberts H, Woods M. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.
24. Verheij RV, Van der Zee J. Collecting information in general practice: 'just by pressing a single button'? In: Westert GP, Jabaaij L, Schellevis FG, editors. Morbidity, performance, and quality in primary care: Dutch general practice on stage. Oxon: Radcliffe Publishing; 2006. 265-272.
25. Van der Velden K. General practice at work. PhD [dissertation]. Rotterdam: Erasmus University Rotterdam; 1999.
26. WONCA Classification Committee. An international glossary for general/family practice. Fam Pract. 1995;12:341-369.
27. Van der Linden MW, Westert GP, De Bakker DH, Schellevis FG. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Klachten en aandoeningen in de bevolking en in de huisartspraktijk (The second Dutch national survey of general practice. Health and problems in the Dutch population and in general practice). Utrecht, Bilthoven: NIVEL, RIVM; 2004.

Chapter 2

Development of a case-based system for grouping diagnoses in general practice

Marion CJ Biermans
Dinny H de Bakker
Robert A Verheij
Jan V Gravestein
Michiel W van der Linden
Pieter F de Vries Robbé

International Journal of Medical Informatics 2008;77:431-439

ABSTRACT

Introduction: This article describes the development of EPICON; an application to group ICPC-coded diagnoses from electronic medical records in general practice into episodes of care. These episodes can be used to estimate prevalence and incidence rates.
Methods: We used data from 89 practices that participated in the Dutch National Survey of General Practice. Additionally, we held interviews with seven experts, and studied documentation to establish the requirements of the application and to develop the design. We then performed a formative evaluation by assessing incorrectly grouped diagnoses.
Results: EPICON is based on a combination of logical expressions, a decision table, and information extracted from individual cases by case-based reasoning. EPICON is able to group all diagnoses in the selected 89 practices, and groups 95% correctly.
Conclusion: The results cautiously indicate that EPICON's performance will probably be adequate for the purpose of estimating morbidity rates in general practice.

INTRODUCTION

Application domain

General practitioners increasingly register patient data in electronic medical records (EMRs). These data could be a valuable source for epidemiologic research. Accurate recording and coding of diagnoses during consultations can be an especially useful source for estimating prevalence- and incidence rates of diseases encountered in general practice. These rates are important for making probability diagnoses, monitoring diseases in the population, conducting scientific research, and evaluating health care policy.

A diagnosis in general practice can refer to a symptom or a complaint (symptom diagnosis), a syndrome (nosological diagnosis) or a disease (pathological/pathophysiological diagnosis).[1] In this article, we use the umbrella term diagnosis to refer to any of these categories.

Diagnoses in general practice are not directly suitable for estimating prevalence- and incidence rates. This would require that all diagnoses of a patient which refer to the same health problem are grouped. For instance, a patient visits the general practitioner for a cough (diagnosis a) which develops into a pneumonia (diagnosis b) several days later. This health problem should be counted only once when estimating occurrences of diseases, namely as a case of pneumonia. To avoid double counting, diagnosis a and b have to be grouped. Diagnoses referring to the same health problem can be grouped into an episode of care, i.e.; “all encounters for the management of a specific health problem”.[1] An episode of care is usually named after the last diagnosis, which can be used to estimate the numerator of the epidemiological fraction.

Generally, two approaches for constructing episodes can be used. In the first approach, the general practitioner groups diagnoses directly into a problem-oriented or episode-oriented medical record. Lawrence Weed introduced the problem-oriented medical record (POMR) in 1968. The POMR is centered around problems in a problem list.[2] Diagnoses that refer to the same health problem receive the same problem number, which can be used to estimate morbidity rates.[3] A disadvantage of using this method for epidemiologic research is that problem lists are frequently not kept up to date.[4] The new generation of Dutch primary care information systems is episode-oriented: all patient information is actually recorded into episodes of care.[5] Data from these episode-oriented systems are probably very well suited for epidemiologic research. However, these systems are still in an implementation phase. In the second approach, diagnoses are grouped afterwards, through manual review or a computerised method. This approach is useful if episodes are not or inadequately constructed by the general practitioner. In this article, we will describe the development of EPIsode CONstructor (EPICON), an application for grouping diagnoses afterwards into episodes. EPICON makes it possible to use data from EMRs in general practice for estimating prevalence- and incidence rates.

Previous research

Our project builds upon the second Dutch National Survey of General Practice (DNSGP-2), which has been described elsewhere.[6;7] In the DNSGP-2, diagnoses were grouped afterwards into episodes for 89 general practices. A semi-computerised method was used in which ‘easy to group’ diagnoses were grouped automatically (80% of all diagnoses), and ‘difficult to group’ diagnoses were grouped manually (20% of all diagnoses). An example of a difficult case is a patient who is diagnosed with tiredness, and who has also been diagnosed with hypothyroidism and general deterioration. This case is complicated, because tiredness is a very non-specific symptom that is observed in many medical conditions. In other words, there are no explicit rules to decide whether tiredness should be grouped with hypothyroidism, with general deterioration, or as a separate episode. In general, complicated cases involve a multi-class classification task and an absence of clear-cut classification rules. The DNSGP-2 dataset, in particular the manually grouped diagnoses, contains implicit knowledge of the signs, symptoms, and the course of diseases, which could be used to solve the problem of grouping diagnoses.

Case-based reasoning

Basic problem-solving approaches in the field of artificial intelligence are rule-based reasoning (based on if… then… rules), model-based reasoning (based on a causal or functional model), and case-based reasoning (based on examples).[8] We selected case-based reasoning (CBR), because the domain knowledge needed to group diagnoses into episodes, is implicit knowledge, which lends itself more for reasoning based on analogy than for formulating domain rules or for constructing a model. Also, ample cases were available, for the DNSGP-2 dataset provided an extensive case library.

CBR is a problem-solving paradigm based on psychological theories of human cognition which provides a method for constructing intelligent systems. It focuses on analogy as a strategy for solving real-world problems. Human experts differ from novices in their ability to relate problems to previous ones, to reason based on analogies between current and old problems, and to use solutions from earlier experiences. A case-based reasoner solves a new problem by remembering a previous similar situation and reuses information and knowledge from that situation. The following four processes describe a general CBR cycle:

1. Retrieve: Given a new problem, former similar cases are retrieved.
2. Reuse: Information and knowledge in the retrieved cases are used to solve the problem.
3. Revise: The solution is tested for success, and repaired if it fails.
4. Retain: A successful solution is incorporated into the case base for future use. Many case based systems are so-called retrieval-only systems or act primarily as retrieval and reuse systems. They merely perform the retrieval or the retrieval-and-reuse task.[9-14]

Research questions

The objective of our research was to develop a fully-computerised method for the construction of care episodes. This project is divided into a development and evaluation phase. In the development phase, we assessed the requirements, designed, and built the system. In the evaluation phase, we performed a formative evaluation. We formulated the following questions:

Development phase

1. How were diagnoses grouped in the semi-computerised method?
2. a) What are the requirements, and b) what is the case-based design of the fully-computerised method?

Evaluation phase

3. How many, and which diagnoses are misclassified by EPICON?

The aim of this project is to determine whether the development of a computerised grouping method can disclose data from EMRs in general practice for epidemiologic research.

METHODS

Dataset

The dataset used in this research is a longitudinal set of patient records provided by a Dutch network of computerised general practices (LINH).[15;16] The general practitioners within this network record longitudinal data on consultations, including diagnoses, prescriptions, and referrals.

Within the framework of the DNSGP-2, episodes were constructed for LINH-data of one year (2001), which were used to estimate prevalence- and incidence rates of diseases in general practice.[17]

We selected all 89 general practices (comprising 166 individual general practitioners) for which diagnoses were grouped afterwards into episodes from the total number of 96 practices that were included in the DNSGP-2. The patient population, which can be used to estimate the denominator of the epidemiologic fraction, includes all listed patients (n = 343 853). Compared to population figures from Statistics Netherlands, the patient population comprises a representative sample of 2% of the Dutch population regarding age, gender, and type of insurance.

Diagnoses were coded according to the International Classification of Primary Care (ICPC).[18] This classification system has a biaxial structure. The primary axis represents 17 chapters referring primarily to a body system. The other axis represents seven components (C1: symptoms and complaints, C2-C6: process, and C7: syndromes and diseases). For coding diagnoses, both C1 (codes 1-29) and C7 (codes 70-99) can be used. Diagnosis codes were assigned to consultations, prescriptions, and referrals.

For consultations only, the general practitioners, in accordance with strict guidelines, characterised each diagnosis as either belonging to a new or an ongoing 'type of episode'. A new episode refers either to a newly presented health problem or to a recurrent health problem, and contributes to both the incidence and the prevalence. An ongoing episode refers to a continuing health problem, and contributes to the prevalence only.

Development phase

The semi-computerised method that was used in the DNSGP-2 has not been described previously in any detail. Therefore, the development of EPICON started with an inventory of this method (question 1). Information was gathered through semi-structured interviews with four experts who were responsible for different aspects of this method. In addition, we studied documentation on decisions and procedures. Requirements (question 2a) for the fully-computerised method were assessed through semi-structured interviews with three other experts who are currently responsible for scientific research based on LINH. The design of the fully-computerised method (question 2b) is partly based on the computerised part of the semi-computerised method. We started by designing EPI-0, a simple non-case-based variant. EPI-0 groups diagnoses only on the basis of the computerised part of the semi-computerised method; all remaining diagnoses are grouped as separate episodes. Next, in order to group these remaining diagnoses, we developed the case-based design for EPICON by considering various case-based or exemplar-based classification techniques.[19;20] We decided to design EPICON as a retrieval-and-reuse system.

Detailed information on the semi-computerised method, the requirements and the design of the fully computerised method was documented, verified, discussed, and adjusted accordingly. EPICON is constructed according to the developed case-based design, and it is written in Transact-SQL.

Evaluation phase

We compared EPICON to the original, semi-computerised method, which is considered the 'gold standard'. We also compared EPI-0 to the gold standard in order to examine what the case-based part in EPICON adds to EPI-0 (the computerised part of the semi-computerised method).

All diagnoses in the original dataset were regrouped by EPI-0, and by EPICON. To gain insight into the process in which diagnoses are grouped by each variant, we calculated the number of diagnoses grouped in each step in both EPI-0 and EPICON.

For each diagnosis, we assessed whether or not, compared to the 'gold standard' it was grouped correctly by EPI-0 and EPICON. This comparison yielded three types of misclassifications. Table 2.1 shows examples of these three types.

Table 2.1 Examples of misclassified diagnoses

Patient	Diagnosis* (ICPC)	Episode numbers			Misclassifications	
		Semi-comp. method	EPI-0	EPICON	EPI-0	EPICON
Jones	L03	1	1	1	no	no
Jones	L02	1	2	1	link failure	no
Parker	R78	1	1	1	no	no
Parker	R75	2	2	1	no	false link
Adams	L74	1	1	1	no	no
Adams	K74	2	2	2	no	no
Adams	K07	1	3	2	link failure	wrong comb.

* L03: Low back symptoms/complaints without radiation
L02: Back symptoms/problems
R78: Acute bronchitis/bronchiolitis
R75: Sinusitis acute/chronic
L74: Fracture hand/foot bones
K74: Angina pectoris
K07: Swollen ankles/edema

Type 1: link failure. A diagnosis, which is linked to another diagnosis by the semi-computerised method, is not linked by EPI-0 or EPICON. Table 2.1 shows the example of patient Jones with two diagnoses relating to back symptoms (L03 and L02). The semi-computerised method linked the two diagnoses by assigning the same episode number to both diagnoses, whereas EPI-0 failed to link the two diagnoses (they received different episode numbers by EPI-0).

Type 2: false link. A diagnosis not linked to another diagnosis by the semi-computerised method, is linked by EPI-0 or EPICON. In the case of patient Parker in table 2.1, EPICON makes a false link between the diagnoses 'acute bronchitis/bronchiolitis' (R78) and 'sinusitis acute/chronic' (R75).

Type 3: wrong combination. A diagnosis linked to another diagnosis by the semi-computerised method, is also linked by EPI-0 or EPICON, but to the wrong diagnosis. Table 2.1 shows that the semi-computerised method linked the diagnosis 'swollen ankles/edema' (K07) for patient Adams to the diagnosis 'fracture hand/foot bones' (L74). EPICON, however, linked the diagnosis 'swollen ankles/edema' (K07) to the diagnosis 'angina pectoris' (K74).

RESULTS

Development phase

Semi-computerised method

Results of our inventory of the semi-computerised method that was used in the DNSGP-2 are presented in the form of a flowchart. Figure 2.1 shows this method, which consisted of 5 steps (shown in between parentheses).

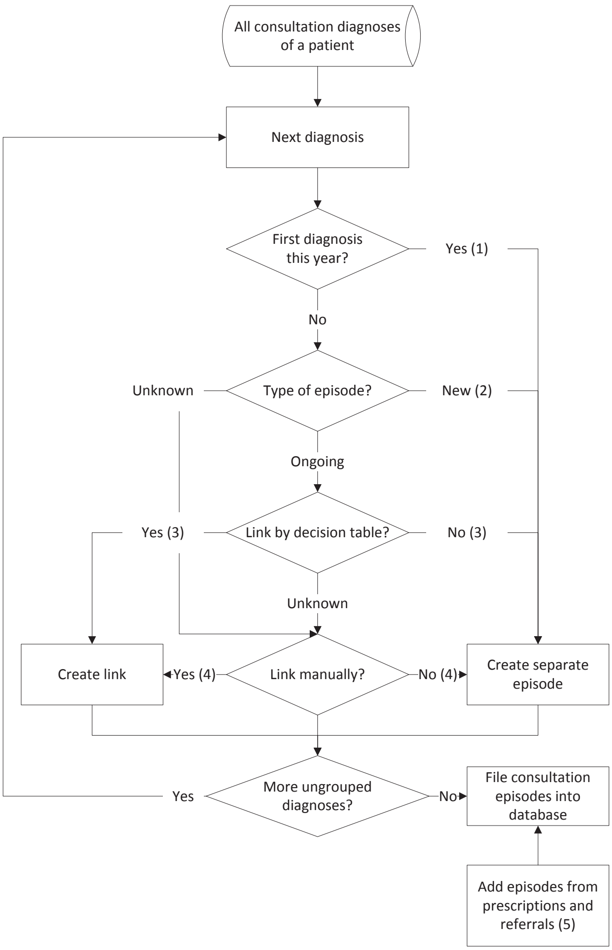


Fig. 2.1 Semi-computerised grouping method

Step 1. The first consultation diagnosis of a patient in a one-year registration period was grouped into a separate episode. ‘Create separate episode’ means that a separate episode number was assigned to a diagnosis. Operationally, an episode is a row of diagnoses with the same episode number. The first diagnosis of this row could be either new or ongoing (as explained in the last paragraph of the section ‘Dataset’), whereas subsequent diagnoses were always ongoing. The first diagnosis in the row determined whether the episode was new or ongoing, the last syndrome/disease diagnosis in the row determined the name of the episode. If no syndrome/disease diagnosis existed, the episode was named after the last symptom diagnosis in the row.

Step 2. The remaining diagnoses were grouped by ‘type of episode’ (new or ongoing). A new diagnosis was grouped as a separate episode. Diagnoses for which the ‘type of episode’ was missing (9% of the total number of diagnoses) were grouped manually. Note the difference between a separate and a new episode. A separate episode can be either new or ongoing.

Step 3. If possible, ongoing diagnoses were grouped based on a decision table, which consists of a combination of an ongoing diagnosis and a previous diagnosis, and a decision whether or not they should be grouped together (see table 2.2). These decisions were made by two physicians (a medical doctor experienced in rheumatology research and a general practitioner experienced in Dutch primary care) who relied on their clinical experience. The table includes only the most frequently occurring combinations (27 092 in total, which is 11% of the total number of 237 016 possible combinations of two diagnoses).

Table 2.2 Decision table for linking an ongoing diagnosis to a previous* diagnosis

Ongoing diagnosis (ICPC-code)	**Previous (new or ongoing) diagnosis (ICPC-code)**	**Decision**	
URI (head cold) (R74)	Knee symptoms/complaints (L15)	Make no link	
URI (head cold) (R74)	Muscle pain/fibrositis (L18)	Unknown	
URI (head cold) (R74)	Influenza (R80)	Make link	
. . .	. . .	. . .	(n = 27 092)

* The program first checked if the diagnosis could be linked to the previous diagnosis. If the result was ‘make no link’ or ‘unknown’, the diagnosis before the previous diagnosis was checked, and so on backwards in time. The program takes thus the sequence of diagnoses (but not the interval between diagnoses) into account.

Step 4. The manual grouping was carried out by 29 medical coders (mostly medical students). The coders had access to all the available data of a patient, including age, gender, and type, date, and the ICPC-code of all former diagnoses. Free text added to an ICPC-code by a general practitioner was also available to the coders. To increase agreement among grouping diagnoses, the coders were personally supervised, worked on the basis of an extensive manual with grouping rules, and discussed difficult cases during meetings.[17]

Step 5. Diagnoses from prescriptions and referrals (i.e., the indications for the prescription or referral) did not pass through the foregoing steps, because they had no 'type of episode'. Diagnoses from prescriptions and referrals that were different from any of the consultation diagnoses of a patient, were added to the database as ongoing episodes afterwards.

Requirements

The main outcome of the interviews was that the fully-computerised method should be in line with the basic principles and methods used in the semi-computerised method. The three experts emphasised that the primary aim of constructing episodes was to estimate the prevalence and incidence rates; a secondary aim was to describe health care use in general practice. A constraint in our study was posed by privacy legislation,[21] which rendered the free text added to diagnosis codes by general practitioners unavailable for the fully-computerised method. Therefore, we could not investigate whether or not the free text contained additional, medical information that would be useful for grouping diagnoses into episodes.

Design

In line with the requirements, we designed the fully computerised-procedure as an extension of the semi-computerised method. Figure 2.2 shows the design for EPICON. The grey parts are derived from the computerised part of the semi-computerised method. EPI-0 consists of these grey parts, and step 4d. Step 2 differs somewhat from the semi-computerised method: all diagnoses for which 'type of episode' is missing, are considered as ongoing by EPI-0 and EPICON.

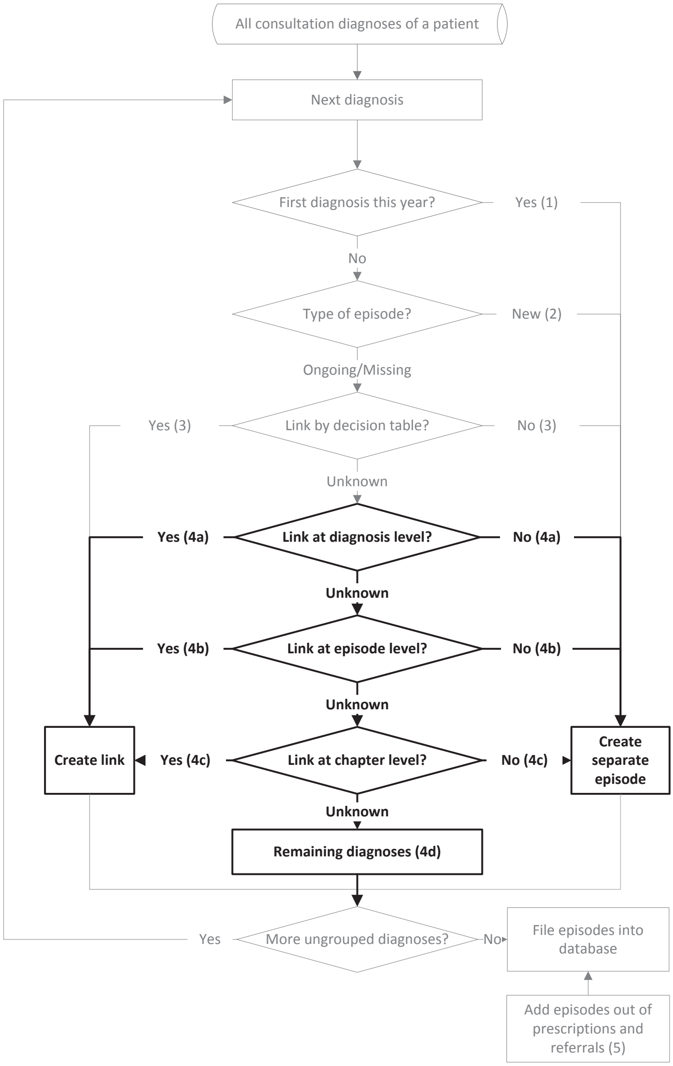


Fig. 2.2 Fully-computerised grouping method

The black parts in figure 2.2 are new. They consist of three case-based steps (4a - 4c), and one 'rest' step (4d).

First, we will explain the case-based design in general and then we will describe the various steps in more detail. Important design decisions in developing a case-based system concern the similarity and granularity of cases. If we would use a very fine grain size (like defining similar cases as patients with the same sex, the same birth date, and the same diagnoses in the same week of the year) we would not find any similar cases. Using a very coarse grain size (such as defining similar cases as patients who also have a symptom diagnosis) would lead to many misclassifications. During the design process, we experimented with various definitions of cases, with the sequence of steps, with the cut-off points used, and evaluated the outcome in order to improve the design. In the final design, we used a definition of cases at the level of diagnoses (step 4a), because this definition provides cases that are both sufficiently similar and numerous (n = 5 772 908 cases). However, not all diagnoses could be grouped in this way. For these remaining diagnoses, we increased the grain size of cases to the level of episodes (step 4b), which resulted in a sufficiently similar, but smaller number of cases (n = 2 095 536 cases). For any remaining diagnoses, we increased the grain size of cases once more to the level of chapters (step 4c) resulting in a large number of broadly similar cases (n = 5 855 741 cases). The various steps are explained in more detail in the section below.

Step 4a. In this step, we used case-based reasoning at the level of diagnoses. In the retrieval process, we defined a problem or an unsolved case as follows: Should diagnosis X be linked to one of the other diagnoses of the same patient? For instance, patient Smith visited the general practitioner because he fainted that day. The general practitioner assigned the diagnosis fainting (A06), and characterised it as ongoing (i.e., it refers to a continuing health problem). Earlier that year, patient Smith visited the general practitioner for other health problems, namely dermatophytosis (S74), general tiredness (A04), and hypoglycemia (T87). In this example, the unsolved case is; should the diagnosis fainting for patient Smith be linked to one of the other diagnoses of patient Smith, i.e. dermatophytosis, tiredness or hypoglycemia?

We defined a former similar or solved case as a manually grouped diagnosis X that belongs to a patient who also has one of the other diagnoses. In the example above, we selected all the patients for whom the diagnosis fainting was manually grouped by one of the medical coders. From this group, we then retrieved those patients who were also diagnosed with dermatophytosis, or tiredness or hypoglycemia.

We were unable to retrieve these solved cases directly from the DNSGP-2 dataset. Therefore, the structure of the original dataset had to be changed into a case base that represents cases. Table 2.3 shows this case base. Each row shows a number of similar cases that belong to the same category because they have the same combination of diagnoses in common.

Table 2.3 Case base 1: cases at diagnosis-level*

Manually grouped diagnosis (ICPC-code)	Other diagnosis of the same patient (ICPC)	Total no of similar cases	Grouped into same episode (n)	Grouped into same episode (%)
URI (head cold) (R74)	Pain: generalised/ unspecified (A01)	38	5	13.2%
URI (head cold) (R74)	Chills (A02)	13	12	92.3%
URI (head cold) (R74)	Fever (A03)	362	202	55.8%
. . .	. . .	. . .	. . .	(n =106 908)

* Combinations were only included in this table if (the total number of combinations was at least 5) or (if the total number of combinations was 3 or 4 *and* either 0% or 100% was grouped into the same episode).

In the reuse process, we carried out the following activities. First, we counted the number of cases that was solved by either grouping or separating for each category, and calculated the percentage of cases that were 'grouped together'. This percentage was then used to solve the case. For each unsolved case, the percentages of all eligible categories were compared to a previously defined cut-off point. If one of the percentages exceeded the cut-off point, diagnosis X was linked to this other diagnosis. If all percentages were below the cut-off point, diagnosis X was grouped separately.

In the patient Smith case, the following cases are retrieved from the case base:
A06-S74, a total of 93 cases, 0 times grouped into the same episode (0.0%);
A06-A04, a total of 167 cases, 68 times grouped into the same episode (40.7%);
A06-T87, a total of 15 cases, 8 times grouped into the same episode (53.0%).
The percentage of the category A06-T87 exceeded the cut-off point for EPICON (set at 48% as explained below). Therefore, EPICON linked the diagnosis fainting for patient Smith to the diagnosis hypoglycemia.

Step 4b. In this step, we defined cases at the level of episodes. Here, an unsolved case is defined as: Should diagnosis X be linked to one of the other episodes of the same patient? We constructed a second case base (n = 84 059 different cases) and used the same procedure as the one in step 4a.

Step 4c. We used case-based reasoning at the level of chapters to handle rare cases that cannot be grouped by any of the preceding steps. In this step, an unsolved case is described as whether diagnosis X should be linked to one of the chapters to which other diagnoses of the same patient belong. We constructed a third case base (n = 10 068 different cases), based on the chapter to which another diagnosis of the same patient belongs, and used the same procedure as in step 4a.

Step 4d. Diagnoses that cannot be grouped by any of the aforementioned steps were set aside. The computerised method continuously tried to group these remaining diagnoses because it would still be possible to group some of these diagnoses when more information from following diagnoses of the same patient became available. In the end, all remaining diagnoses were grouped as separate episodes.

Assessing the cut-off points

The optimal cut-off point for each case-base was assessed by calculating, for various cut-off points, the total number of correctly and incorrectly grouped combinations. For example, using a cut-off point of ≥ 70% in case base 1 (see table 2.3) means that all 38 head cold-pain (R74-A01) combinations are not grouped together, because 13.2% < 70%. This results in 5 incorrectly grouped (link failures) and 33 correctly grouped combinations.

We used two criteria to assess the optimal cut-off points. First, the total number of incorrectly grouped combinations should be as low as possible. Second, to avoid systematic bias, the number of link failures should equal the number of false links. Subsequently, the optimal cut-off points are ≥ 48% for case base 1, ≥42% for case base 2, and ≥ 71% for case base 3.

Evaluation phase

Table 2.4 shows the percentage of diagnoses assigned to each step, in EPI-0 and EPICON. The results show that the large majority of the diagnoses can be grouped by EPI-0; only 13.1% of all diagnoses are remaining diagnoses. Practically all of these remaining diagnoses were grouped by the additional case-based steps in EPICON.

Table 2.4 Percentage of diagnoses grouped in each step

Steps		EPI-0	EPICON
1.	First diagnosis	20.4%	20.4%
2.	New episode	27.4%	27.4%
3.	Decision table	29.4%	29.4%
4a.	Diagnosis level	-	11.3%
4b.	Episode level	-	0.2%
4c.	Chapter level	-	1.4%
4d.	Remaining	13.1%	0.1%
5.	Diagnoses from prescriptions/referrals	9.8%	9.8%
Total no of diagnoses		1 201 234	1 201 234

Figure 2.3 shows the number and type of incorrectly grouped diagnoses. EPI-0 misclassifies 7.1% of all diagnoses. This percentage drops to 4.7% in EPICON, which is a small, but appreciable improvement. Practically all errors in EPI-0 are link failures, whereas EPICON shows a good balance between link failures and false links. For both variants, wrong combinations (81 by EPI-0 and 344 by EPICON) are rare.

Both numerous link failures and numerous false links will produce biased morbidity rates, but in opposite directions. In particular, prevalence rates will be overestimated by link failures and underestimated by false links. Numerous wrong combinations will produce bias in both directions.

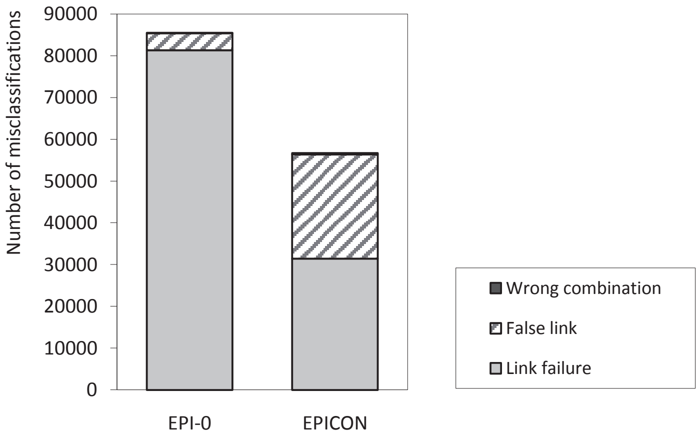


Fig. 2.3 Type of misclassification in EPI-0 and EPICON

We also examined the number of misclassified diagnoses per ICPC-chapter. Figure 2.4 displays the distribution of misclassifications (link failures and false links) across chapters in EPICON. In general, this figure shows a balanced distribution of misclassifications across chapters, i.e., the number of misclassifications is proportional to the total number of diagnoses in a chapter. With the exception of chapter A, the percentage of misclassifications per chapter ranges from 1.4% in chapter W (pregnancy and family planning) to 6.1% in chapter Z (social problems). Additionally, the number of link failures equals the number of false links. An exception is chapter A (general and unspecified), which shows a relatively large number of link failures.

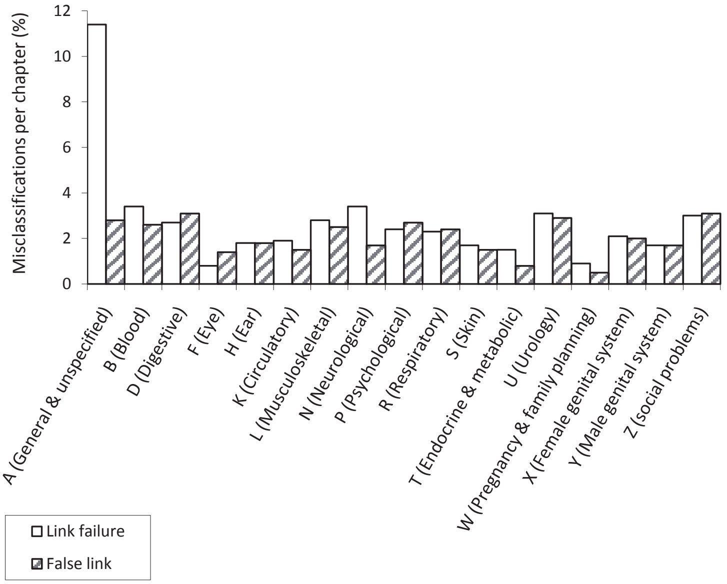


Fig. 2.4 Distribution of misclassifications across ICPC-chapters in EPICON

CONCLUSIONS AND DISCUSSION

To our knowledge, this is the first study into the development of a case-based system for grouping diagnoses in general practice. Previous research in constructing episodes of care focused primarily on grouping insurance claims records and did not use a case-based approach.[22-24]

EPICON groups diagnoses into episodes, based on a combination of logical expressions, a decision table, and information extracted from individual cases by CBR. This application is able to group all diagnoses in our dataset, and groups 95% correctly. These results cautiously indicate that EPICON's performance will probably be adequate for the purpose of estimating morbidity rates in general practice.

It is important to note that not all misclassifications will have an effect on the prevalence and incidence rates. Only misclassifications that alter episode names can affect morbidity rates. Furthermore, the number of altered episode names has to be sufficiently large in proportion to the population 'at risk' in the denominator, to have an effect at all. Therefore, the analyses of misclassifications should be considered a formative evaluation, with the purpose of studying the process of grouping diagnoses. In the next phase of this project, we will perform a summative evaluation in which we will examine if EPICON serves its purpose by comparing morbidity rates. Dependent on the results of the summative evaluation, EPICON will be applied to the LINH database to estimate yearly prevalence and incidence rates of diseases in general practice.

We developed two variants of the fully-computerised method. The most simple variant, EPI-0, is based on the computerised part of the semi-computerised method that was used in the DNSGP-2. EPI-0 consists of logical expressions and a decision table. This variant performs fairly well; it only misclassifies 7.1% of all diagnoses. In the other variant, EPICON, we added three case-based steps, based on the manual grouping that was used in the DNSGP-2. Comparing EPI-0 to EPICON reveals what CBR adds to EPI-0. This procedure lowers the number of misclassified diagnoses only by 2.4%. These results indicate that the large majority of diagnoses made in general practice can be grouped using simple methods, whereas additional information and more complicated methods are required to group the remaining diagnoses correctly.

EPICON shows a balanced distribution of misclassifications across chapters with the exception of chapter A. This chapter includes general and unspecified diagnoses such as pain and tiredness, which are symptoms of many different diseases. Such diagnoses exemplify the problem of grouping diagnoses in general practice. Compared to other medical doctors, general practitioners encounter many patients with a broad range of symptoms, syndromes, and diseases, and will often have to make an educated guess about the relations between them. EPICON is specifically developed for this setting, and is not applicable to other disciplines. The followed procedure, however, could be adopted to develop a system for grouping codes from other classifications systems.

We used the original grouping of the dataset as a ‘gold standard’, which is both the strength and the weakness of this study. The strength lies in the fact that we had access to a large database of carefully constructed episodes to develop, and evaluate the new episode constructions. The weakness is that our comparison may overestimate the performance of EPICON, because the new grouping method is partly based on the same rules as the original grouping method. Furthermore, our study is based on the assumption that the original grouping is the gold standard. We did not test this assumption, because an evaluation of the gold standard was not a focus of our project. Consequently, this study cannot reveal possible flaws in the gold standard, for instance in the decision table, that passed on to EPICON.

This project was carried out in the Netherlands. Dutch general practitioners are in the lead in when it comes to using EMRs; practically all practices were computerized in 2001.[25] Therefore, this project can be seen as representative for all countries with a high degree of Information Technology in general practice. The followed method is essentially applicable to all large databases of routinely recorded diagnoses in general practice.

In this study, we developed a first, workable application that lends itself for further improvement. A further reduction of misclassifications might be achieved by adding time intervals between diagnoses to the case bases. In addition, EPICON could be extended to a learning, and interactive decision support system by presenting grouped diagnoses to a general practitioner in order to test its solutions. The decision support system would assist the general practitioner in grouping diagnoses directly into episodes.

REFERENCES

1. WONCA Classification Committee. An international glossary for general/family practice. Fam. Pract. 1995;12:341-369.
2. Weed LL. Medical records that guide and teach. N Engl J Med. 1968;278:593-600.
3. Carey IM, Cook DG, De Wilde S, Bremner SA, Richards N, Caine S, et al. Developing a large electronic primary care database (Doctors' Independent Network) for research. Int J Med Inform. 2004;73:443-453.
4. Lauteslager M, Brouwer HJ, Mohrs J, Bindels PJE, Grundmeijer HGLM. The patient as a source to improve the medical record. Fam Pract. 2002;19:167-171.
5. Okkes IM, Groen A, Oskam SK, Lamberts H. Advantages of long observation in episode-oriented electronic patient records in family practice. Methods Inf Med. 2001; 40:229-235.
6. Westert GP, Jabaaij L, Schellevis FG, editors. Morbidity, performance and quality in primary care: Dutch general practice on stage. Oxon: Radcliffe Publishing; 2006.
7. Westert GP, Schellevis FG, De Bakker DH, Groenewegen PP, Bensing JM, Van der Zee J. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15:59-65.
8. Luger GF, Stubblefield WA. Knowledge-intensive problem solving. In: Artificial Intelligence: Structures and strategies for complex problem solving. Reading: Addison Wesley Longman Inc; 1998;207-246.
9. Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications. 1994;7:39-59.
10. Bichindaritz I. Case-based reasoning in the health sciences. Art Intell Med. 2005;36:121-125.
11. Kolodner JL. Case-based reasoning. San Mateo: Morgan Kaufmann; 1993.
12. Macura RT, Macura K. Case-based reasoning: opportunities and applications in health care. Artif Intell Med. 1997;9:1-4.
13. Watson I, Marir F. Case-Based Reasoning: A Review. The Knowledge Engineering Review. 1994;9:355-381.
14. Schmidt R, Montani S, Bellazzi R, Portinale L, Gierl L. Case-based reasoning for medical knowledge based systems. Int J Med Inform. 2001;64:355-367.
15. Tacken MAJB. Quality of preventive performance in general practice: the use of routinely collected data [dissertation]. Radboud University Nijmegen; 2005.

16. Verheij RA, Jabaaij L, Abrahamse H, Van den Hoogen H, Braspenning J, Van Althuis T. Netherlands Information Network of General Practice: Facts and figures on Dutch GP care [website on the internet]. Available from: http://www.linh.nl.
17. Van der Linden MW, Westert GP, De Bakker DH, Schellevis FG. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Klachten en aandoeningen in de bevolking en in de huisartspraktijk (The second Dutch national survey of general practice. Health and problems in the Dutch population and in general practice). Utrecht, Bilthoven: NIVEL, RIVM; 2004.
18. Lamberts H, Woods M, editors. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.
19. Nakamura GV, Medin DL, Taraban R, editors. Categorization by humans and machines. San Diego: Academic Press; 1993.
20. Bareis R. Exemplar-based knowledge acquisition: A unified approach to concept representation, classification, and learning. Boston: Academic Press; 1989.
21. Personal Data Protection Act (Wet bescherming persoonsgegevens). Staatsblad 2000; 302.
22. Rosen AK, Mayer Oakes A. Episodes of care: theoretical framework versus current operational realities. Jt Comm J Qual Improv. 1999;25:111-128.
23. Son RY, Taira RK, Bui AAT, Kangarloo H, Cárdenas AF. A context-sensitive methodology for automatic episode creation. Proc AMIA Symp. 2002; 707-711.
24. Wingert TD, Kralewski JE, Lindquist TJ, Knutson DJ. Constructing episodes of care from encounter and claims data: some methodological issues. Inquiry. 1995;32:430-443.
25. European physicians especially in Sweden, Netherlands and Denmark, lead U.S. in use of electronic medical records. Health Care Res. 2002; 2:1-3.

Chapter 3

Estimating morbidity rates from electronic medical records in general practice: Evaluation of a grouping system

Marion CJ Biermans
Robert A Verheij
Dinny H de Bakker
Gerhard A Zielhuis
Pieter F de Vries Robbé

Methods of Information in Medicine 2008;47:98-106

ABSTRACT

Objectives: In this study, we evaluated the internal validity of EPICON; an application for grouping ICPC-coded diagnoses from electronic medical records into episodes of care. These episodes are used to estimate morbidity rates in general practice.
Methods: Morbidity rates based on EPICON were compared to a gold standard; i.e. the rates from the second Dutch National Survey of General Practice. We calculated the deviation from the gold standard for 677 prevalence and 681 incidence rates, based on the full dataset. Additionally, we examined the effect of case-based reasoning within EPICON using a comparison to a simple, not case-based method (EPI-0). Finally, we used a split sample procedure to evaluate the performance of EPICON.
Results: Morbidity rates that are based on EPICON deviate only slightly from the gold standard and show no systematic bias. The effect of case-based reasoning within EPICON is evident. The addition of case-based reasoning to the grouping system reduced both systematic and random error. Although the morbidity rates that are based on the split sample procedure show no systematic bias, they do deviate more from the gold standard than morbidity rates for the full dataset.
Conclusions: Results from this study indicate that the internal validity of EPICON is adequate. Assuming that the standard is gold, EPICON provides valid outcomes for this study population. EPICON seems useful for registries in general practice for the purpose of estimating morbidity rates.

INTRODUCTION

Prevalence and incidence rates in general practice are used by scientists to monitor symptoms and diseases in the population, by policymakers to formulate and to evaluate health care policy, and by general practitioners to make probability diagnoses. Under certain conditions, listed below, the data from electronic medical records (EMRs) in general practice are a valuable source for estimating these rates.

A necessary condition for estimating morbidity rates from EMRs is that patient lists are available. These lists are needed to determine the size of the underlying practice population, i.e., the denominator of the morbidity rates. Countries with a list system, such as the Netherlands, Denmark, the United Kingdom, Italy, Spain, Portugal, and Slovenia, are also countries with a gatekeeping system.[1] The presence of such a system is a very favorable condition, because within this system, the main pathway to medical care is through general practice. Hence, morbidity rates in general practice within these systems provide a good indication of the health status of the general population.[2;3]

Another requirement is that general practitioners record coded diagnoses in their EMRs. To estimate incidence rates, knowing whether diagnoses refer to a new or a continuing health problem is an additional prerequisite. Furthermore, all diagnoses need to be grouped into episodes to estimate the numerator of the morbidity rates. An episode of care includes “all encounters for the management of a specific health problem”.[4] Consider, for example, a patient who consults the general practitioner for a tension headache (diagnosis a), which, the next day, develops into migraine (diagnosis b). Most likely, both diagnoses refer to the same health problem which, when estimating morbidity rates, should be counted only once, namely as a case of migraine. To avoid double counting, the *two* diagnoses need to be grouped in *one* episode of care named ‘migraine’.

Diagnoses can be grouped into episodes, either directly by the general practitioner, or afterwards, through manual review or use of a computerized method. In an earlier study,[5] we described the development of an application called EPICON, which can be used for the grouping of coded diagnoses from EMRs in general practice into episodes of care. EPICON renders it possible to estimate morbidity rates from EMRs in general practice without having to resort to expensive and time-consuming manual review.

EPICON is based on a combination of logical expressions, a decision table, and information extracted from manually grouped diagnoses by case-based reasoning. Previous research on the process of grouping diagnoses shows that EPICON groups 95% of all diagnoses correctly.[5] The present study focuses on how the grouped diagnoses translate into morbidity rates. In this study, we will examine if EPICON serves it purpose by comparing morbidity rates based on EPICON to a gold standard.

The prevalence and incidence rates from the second Dutch National Survey of General Practice (DNSGP-2) are considered as the gold standard.[6-8] In the DNSGP-2, diagnoses from 89 general practices were grouped afterwards into episodes by a semi-computerized method. This method consists of a computerized component, in which diagnoses are grouped automatically, and a manual component, in which diagnoses are grouped by medical coders.

The development of EPICON started with a simple, fully computerized method, called EPI-0. This variant is identical to the computerized component of the semi-computerized method, but instead of referring 'difficult diagnoses' to the manual component, it groups all remaining diagnoses automatically as separate episodes. EPICON groups these remaining diagnoses by case-based reasoning. In this process, EPICON imitates the manual grouping by retrieving and reusing information and knowledge from the DNSGP-2 dataset. For instance, EPICON solves the grouping problem in the example above by counting how often, in the DNSGP-2 dataset, manually grouped combinations of migraine and tension headache were placed together into one episode. The appendix provides a brief description of EPICON and EPI-0.

The main aim of this study is to assess the internal validity of EPICON. Internal validation refers to the performance of a system in a sample used to develop the system. Internal validation is a requirement for external validation, which refers to the performance of a system in a new sample of patients.[9]

In this study, we will address the following research questions:

1. What is the deviation from the gold standard for morbidity rates based on EPICON? In particular, what is the effect of case-based reasoning on the deviation from the gold standard?
2. What is the performance of EPICON among patients not included in the development of EPICON but from the same population?

METHODS

Dataset

We used data from EMRs of practices associated with the Netherlands Information Network of General Practice (LINH).[10-11] The general practitioners within this network assign diagnosis codes to consultations, prescriptions, and referrals. The diagnoses are coded according to the International Classification of Primary Care (ICPC, first edition).[12] This classification system includes 685 different diagnosis codes, classified into 17 chapters each of which refers to a specific body system or problem area. Within each chapter, codes 1-29 refer to symptoms and complaints, and codes 70-99 refer to diseases.

Each consultation diagnosis is characterized as either belonging to a new or an ongoing 'type of episode'. A new episode refers either to a newly presented health problem or to a recurrent health problem, while an ongoing episode describes a continuing health problem.

Representativeness

The 89 practices are representative of all Dutch general practices with respect to region of residence (i.e. the province in which the practice is located), dispensing status (i.e. whether the practice is permitted to dispense its own prescriptions), and degree of urbanization.[13] Solo practices, however, are slightly underrepresented in this sample (53% versus 64%). The patient population comprises a representative sample of 2% of the Dutch population regarding age, gender, and type of medical insurance.[14]

Episode construction

Within the framework of the DNSGP-2, episodes were constructed afterwards for 89 LINH practices for a period of one year (2001). Episodes were named after the last disease code within an episode. If no disease code existed, the episode was named after the last symptom code within an episode. In this study, we regrouped all diagnoses in these 89 practices into episodes using both EPI-0 and EPICON. Consequently, we had one dataset with three different episode constructions: 1) the existing DNSGP-2 episodes (the gold standard), 2) the EPI-0 episodes, and 3) the EPICON episodes.[5]

Split-sample

We used a split-sample approach to examine the performance of EPICON. This approach allows for overestimation, a well known statistical problem where a model is fitted and evaluated using the same dataset.[15] We randomly split the group of patients with diagnoses (n = 256 227) in half. We used half of the dataset (the training set, n = 128 113 patients) to construct the case bases used by EPICON to group diagnoses into episodes. We then used the other half of the dataset (the test set, n = 128 114 patients) to test EPICON, i.e., the diagnoses of these patients were grouped into episodes based on the split sample case bases. This split sample procedure adds a fourth episode construction.

The obvious disadvantage of the split sample technique is its inefficiency, i.e., the training set only makes up half of the total dataset. To quantify this inefficiency, we made a fifth episode construction, in which we applied the split sample case bases (based on the training set) to the training set. In this way, we could 'separate' the intertwining effects of overestimation and inefficiency in the testset. We would know that overestimation (and not inefficiency of the split sample approach) was a problem if the trainingset would show less deviation than the testset.

Prevalence and incidence rates

The prevalence rate is defined as the proportion of the population with a disease during the one-year study period. The incidence rate refers to the occurrence of new episodes of disease during the observed person-years at risk.

We used the episode names that resulted from the five episode constructions for the numerator of the morbidity rates. Eight practices did not record exactly 365 days, because of vacation, sick leave, etc. For that reason, the episodes were weighted for the number of days recorded (i.e. 365/number of days recorded). For the numerator of the prevalence rates, we counted, per episode name, the number of patients with at least one new or ongoing episode. For the numerator of the incidence rates, we counted, per episode name, the number of new episodes.

The mid-year population (i.e., the average of the population at the beginning and the end of the one-year study period, n = 343 853) of the 89 general practices was used as the denominator. Half of the mid-year population was used to calculate the split-sample morbidity rates. Morbidity rates of symptoms and diseases that occur only in the female (W, X) or the male chapters (Y) are based only on the female or the male mid-year population.

A total of 677 different prevalence rates and 681 different incidence rates were calculated for each of the five episode constructions. Some of the 685 possible rates were excluded, because they do not refer to a symptom or disease ('no disease' (A97)), they refer to incident events only ('perinatal mortality' (A95), 'death' (A96), '(un)complicated labour/delivery of live/stillbirth (W90 -W93)), or no cases ('trachoma' (F86)) or no new cases occurred ('neoplasm cardiovascular' (K72), 'syphilis female' (X70)).

Comparison of morbidity rates

Figure 3.1 shows the design of the study. We calculated the differences between the rates based on EPI-0, EPICON, the test set or the training set on the one hand, and the gold standard on the other hand. We compared the deviation of rates based on EPI-0 to the deviation of rates based on EPICON to examine the effect of case-based reasoning. The comparison between the deviation in the testset to the deviation in the trainingset provides insight into the separate effects of overestimation and inefficiency.

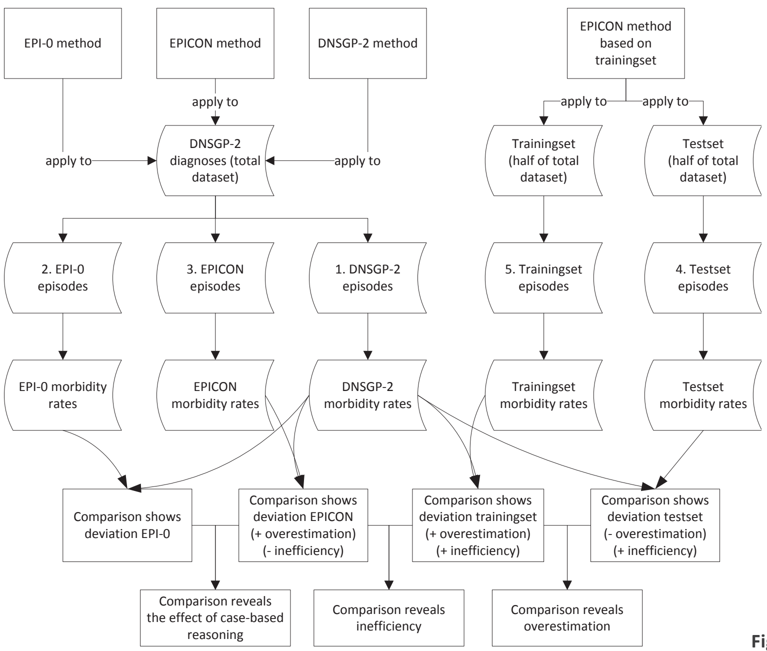


Fig. 3.1 Design of the study

We distinguished between infrequently (rate < 1 per 1000 per year) and frequently (rate ≥ 1 per 1000 per year) occurring symptoms and diseases (rates are based on the gold standard). This divided the large number of rates into two approximately equal parts. For infrequently occurring symptoms and diseases, we calculated the absolute differences between the morbidity rates (absolute deviation). For symptoms and diseases that occurred frequently, we calculated the percentage to which the rates differed from the gold standard (relative deviation).

RESULTS

Morbidity rates based on EPICON

Figures 3.2-3.5 show the morbidity rates based on EPICON in comparison to the gold standard. In general, these morbidity rates deviate slightly from the gold standard and show no systematic bias.

Infrequent prevalence rates (figure 3.2) deviate from –0.04 ('hypoglycemia' (T87)) to 0.06 ('fear of other respiratory disease' (R27)). An outlier is the diagnosis 'investigation with abnormal results' (A91)). The prevalence of this diagnosis is 0.94 per 1000 patients based on EPICON (read from the x-axis of figure 3.2) and 0.85 per 1000 patients according to the gold standard, which is a difference of 0.09 (read from the y-axis of figure 3.2). EPICON probably failed to group this diagnosis correctly in a number of cases, because it is a very non-specific diagnosis that can be linked to many different diagnoses. Frequent prevalence rates (figure 3.3) deviate from -6% ('vomiting/nausea of pregnancy' (W05)) to 12% ('abnormal unexplained blood test' (B85)).

The deviation for infrequent incidence rates (figure 3.4) ranges between –0.05 ('cervicitis/other cervical disease' (X85)) and 0.07 ('hypertension with involvement target organs' (K87)). Outliers are 'generalized/unspecified pain' (A01), which is difficult to group because of its non-specificity, and 'osteoporosis' (L95). EPICON grouped the latter incorrectly in a number of cases because 'osteoporosis' (L95) can be linked to many other diagnoses in the musculoskeletal chapter (L), and occurs mainly in elderly patients who have a great deal of comorbidity. The deviation for frequent incidence rates (figure 3.5) varies from –5% ('acute myocard infarction' (K75)) to 9% ('asthma' (R96)).

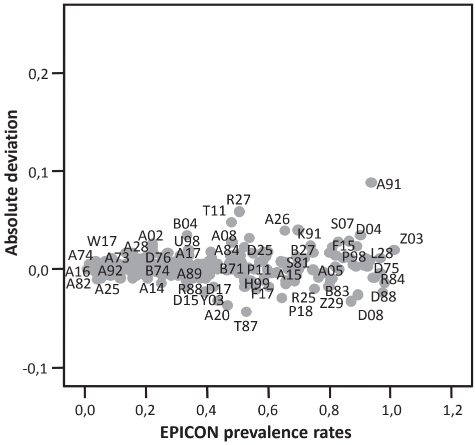


Fig. 3.2 Deviation from the gold standard for infrequent EPICON prevalence rates

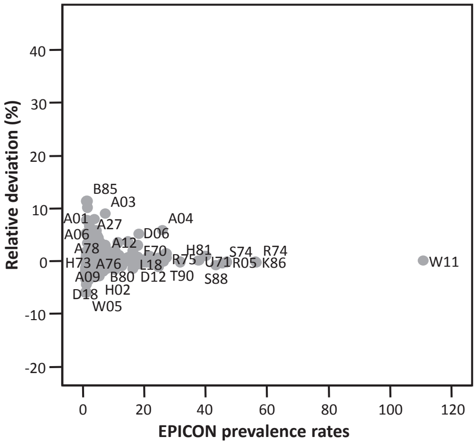


Fig. 3.3 Deviation from the gold standard for frequent EPICON prevalence rates

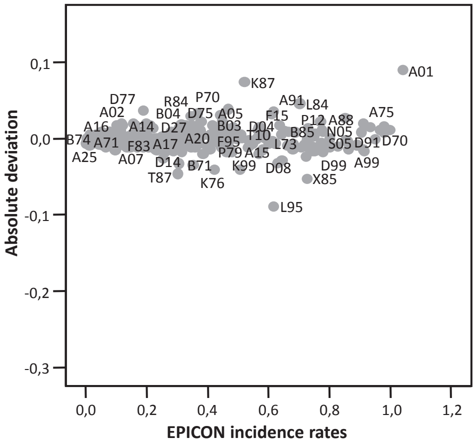


Fig. 3.4 Deviation from the gold standard for infrequent EPICON incidence rates

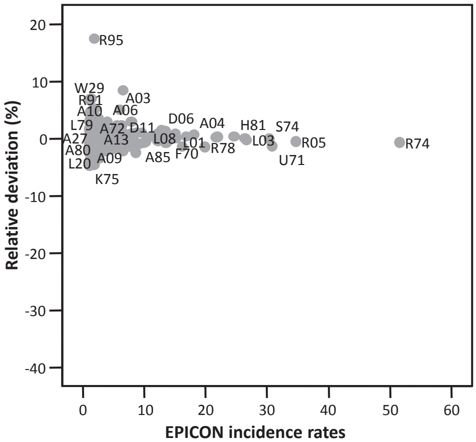


Fig. 3.5 Deviation from the gold standard for frequent EPICON incidence rates

An exception is 'emphysema/copd' (R95)). The incidence rate of this diagnosis based on EPICON is 17% higher than the gold standard (1.78 versus 1.52 per 1000 person-years). The reason for this difference could be that 'emphysema/copd' (R95) occurs frequently in elderly patients with concurrent diseases such as 'decompensatio cordis' (K77), 'pneumonia' (R81), 'chronic bronchitis/bronchietasis' (R91), and 'asthma' (R96). The intricate distinction between exacerbations (usually grouped within an ongoing episode) and complications (usually grouped as a separate episode) within this group of diseases may have caused variation in the manual (the gold standard) and subsequently also in the automatic grouping procedure (EPICON).

The effect of case-based reasoning

The comparison between morbidity rates based on EPI-0 (figures 3.6-3.9) and morbidity rates based on EPICON (figures 3.2-3.5) reveals the effect of case-based reasoning. Morbidity rates based on EPI-0 display a systematic bias, which manifests itself differently for prevalence and incidence rates.

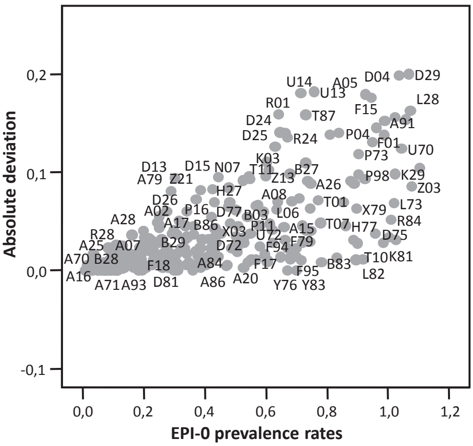


Fig. 3.6 **Deviation from the gold standard for infrequent EPI-0 prevalence rates**

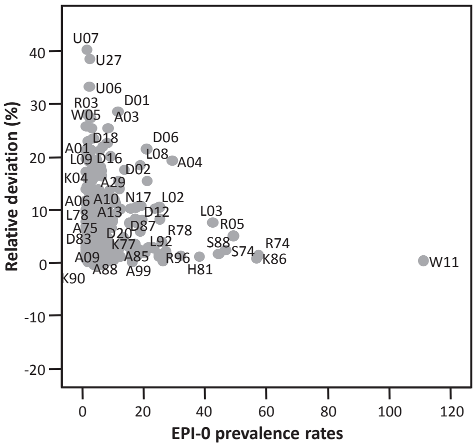


Fig. 3.7 **Deviation from the gold standard for frequent EPI-0 prevalence rates**

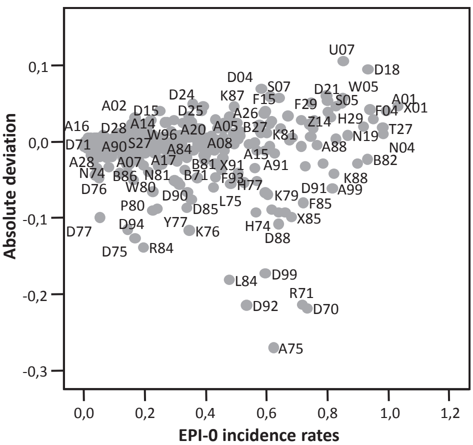


Fig. 3.8 **Deviation from the gold standard for infrequent EPI-0 incidence rates**

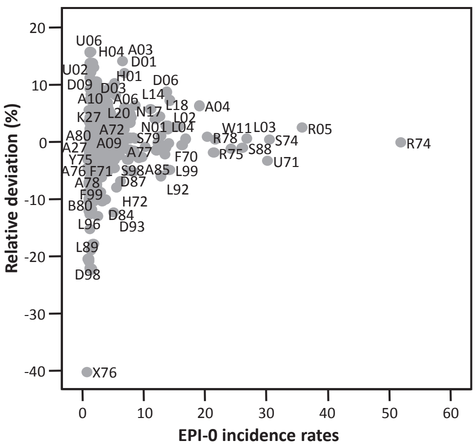


Fig. 3.9 **Deviation from the gold standard for frequent EPI-0 incidence rates**

Prevalence rates based on EPI-0 (figures 3.6-3.7) show a systematic bias toward higher prevalence rates. EPI-0 produces obviously more episodes than the gold standard, which was to be expected since it processes all remaining diagnoses as separate episodes. For instance, EPI-0 always groups the remaining diagnosis ‘other complaints of urine’ (U07) as a separate episode, whereas medical coders, i.e., the gold standard, frequently grouped this diagnosis into an episode called ‘cystitis/other urinary infection’ (U71). The resulting prevalence rates for ‘other complaints of urine’ (U07) are 1.43 per 1000 patients based on EPI-0 (492 patients/343853), and 1.02 per 1000 patients (350 patients/343853) according to the gold standard, which is a difference of 40% (figure 3.7).

Incidence rates based on EPI-0 (figures 3.8-3.9) show a systematic bias toward higher incidence rates for symptoms and complaints (codes 1-29), and lower incidence rates for diseases (codes 70-99). The total number of new episodes in EPI-0 is identical to the total number of new episodes in the gold standard; in both episode constructions, a new diagnosis is grouped as a separate episode. The episode name, however, can be different. For example, all new diagnoses ‘other complaints of urine’ (U07) were grouped as new episodes in EPI-0, whereas the gold standard grouped part of these diagnoses into new episodes of ‘cystitis/other urinary infection’ (U71).

When case-based reasoning is added to the grouping system (EPICON), this systematic bias is removed and the precision of the estimates is improved.

Performance of EPICON in split sample procedure

Figures 3.10-3.13 show the results of the split sample method. Morbidity rates of the test set, that are based on the training set, show no systematic bias, but deviate more from the gold standard than the morbidity rates that are based on the full dataset. Figures of morbidity rates of the training set that are based on that same training set show a similar pattern (not shown). We therefore believe that a major part of the imprecision of our testset is due to inefficiency (for the training set is only half of the total dataset) and a minor part is due to overestimation.

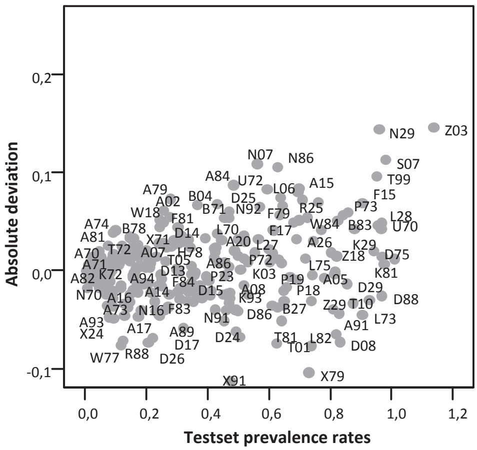


Fig. 3.10 Deviation from the gold standard for infrequent test set prevalence rates

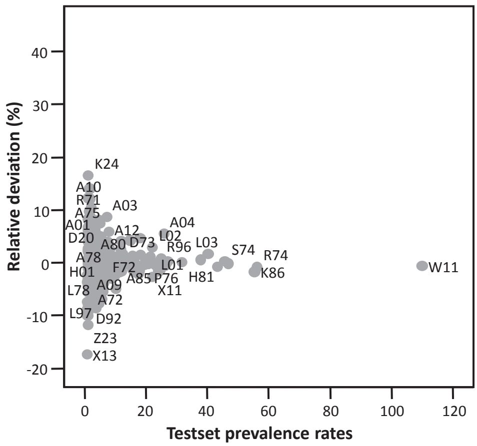


Fig. 3.11 Deviation from the gold standard for frequent test set prevalence rates

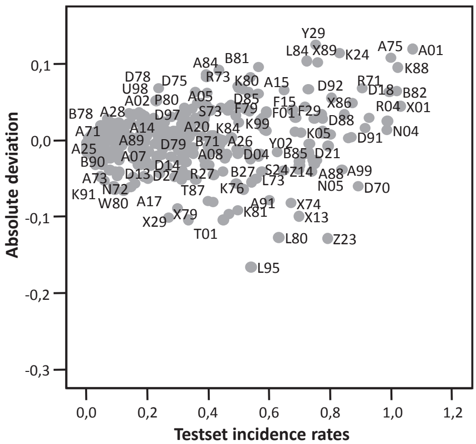


Fig. 3.12 Deviation from the gold standard for infrequent test set incidence rates

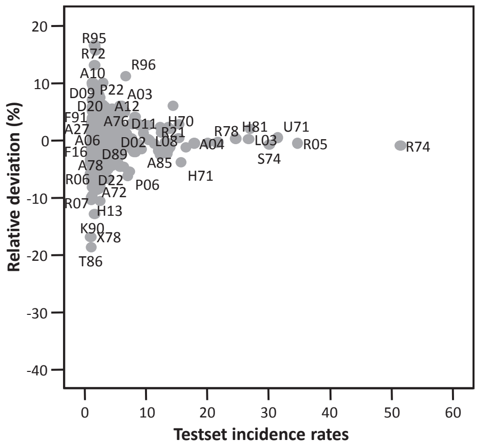


Fig. 3.13 Deviation from the gold standard for frequent test set incidence rates

DISCUSSION AND CONCLUSIONS

Results from this evaluation study indicate that EPICON is a useful tool to disclose data from EMRs in general practice for estimating morbidity rates. In general, morbidity rates based on EPICON deviate only slightly from the gold standard and show no systematic bias. EPICON performs less well, however, when it comes to diagnoses that can be linked to many other diagnoses, such as 'generalized/unspecified pain' (A01).

This judgement is related to the purpose of EPICON, which is to estimate morbidity rates in general practice. These rates are used at population level. For instance, to describe the extent of health problems, to forecast public health or to study health differences in time or between regions. Considering what morbidity rates are used for, some deviation from the gold standard, such as an increase from 13 to 14 per 1000 patients per year, is regarded as acceptable. Our judgement might be completely different with another purpose in mind. For example, we would not recommend to use EPICON for a decision about the individual treatment of a patient.

This study revealed that the role of case-based reasoning in EPICON is very important for acquiring precise and unbiased morbidity rates. We started with a simple, not case-based method (EPI-0) to automatically group diagnoses into episodes. In the next step, we added case-based reasoning, which led to the creation of EPICON. The evaluation showed that the performance of the grouping system improved considerably by adding case-based reasoning. Furthermore, the results indicate that the current set of cases is large enough for the purpose of estimating morbidity rates. Nevertheless, the performance of EPICON could be improved by increasing the number of cases, in particular by adding cases of patients with rare symptoms and diseases.

A limitation of our study is that both EPICON and the evaluation are based on the assumption that the standard is gold. This assumption seems plausible since data from the DNSGP-2 are widely used in the Netherlands, but there is no evidence to support this assumption. However, an evaluation of the gold standard was not a focus of our project. Therefore, this study cannot reveal possible flaws in the gold standard that were subsequently passed on to EPICON.

An evaluation of the gold standard would include a comparison between morbidity rates from the DNSGP-2 and rates from other registries in general practice. Such a comparison is challenging, because the existing registries differ in factors as region of residence, patient population, classification system used to code diagnoses, measures taken to ensure the quality of the registration, practice software, and in the methods used to calculate morbidity rates.[3] More research is needed to quantify the impact of these factors on morbidity rates.

Findings from the split sample procedure indicate that EPICON performs reasonably well on a simulated, independent dataset. Therefore, EPICON seems transportable to other similar datasets of diagnoses in general practice. The split sample procedure is a tough test, because only half of the total dataset is used to construct the case bases. Less demanding tests[16] would probably show that overestimation is only a limited issue in this case.

We conclude that the internal validity of EPICON is satisfactory. EPICON seems useful for LINH and similar registries in general practice for the purpose of estimating morbidity rates. Further research should aim at the external validation of EPICON.

REFERENCES

1. Boerma WG. Profiles of general practice in Europe [dissertation]. Nivel; 2003.
2. Linden M, Gothe H, Ormel J. Pathways to care and psychological problems of general practice patients in a “gate keeper” and an “open access” health care system. A comparison of Germany and the Netherlands. Soc Psychiatry Psychiatr Epidemiol. 2003;38:690-7.
3. Gijsen R, Poos MJJC. Using registries in general practice to estimate countrywide morbidity in the Netherlands. Public Health. 2006;120:923-36.
4. WONCA Classification Committee. An international glossary for general/family practice. Fam Pract. 1995;12:341-69.
5. Biermans MCJ, De Bakker DH, Verheij RA, Gravestein JV, Van der Linden MW, De Vries Robbé PF. Development of a case-based system for grouping diagnoses in general practice. Int J Med Inform. 2008;77:431-439.
6. Westert GP, Jabaaij L, Schellevis FG, editors. Morbidity, performance and quality in primary care: Dutch general practice on stage. Oxon: Radcliffe; 2006.
7. Van der Linden MW, Westert GP, De Bakker DH, Schellevis FG. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Klachten en aandoeningen in de bevolking en in de huisartspraktijk (The second Dutch national survey of general practice. Health and problems in the Dutch population and in general practice). Utrecht, Bilthoven: NIVEL, RIVM; 2004.
8. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, Van der Zee J. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15:59-65.
9. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130:515-24.
10. Tacken MAJB. Quality of preventive performance in general practice: the use of routinely collected data [dissertation]. Radboud University Nijmegen; 2005.
11. Verheij RA, Jabaaij L, Abrahamse H, Van den Hoogen H, Braspenning J, Van Althuis T. Netherlands Information Network of General Practice. Facts and figures on Dutch GP care [website on the internet]. [updated 2006 Aug 1; cited 2006 Nov 30]. Available from: http://www.linh.nl
12. Lamberts H, Woods M, editors. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.

13. Kenens R, Hingstman L. Cijfers uit de registratie van huisartsen (Figures from the registration of general practitioners). Utrecht: Nivel; 2001.
14. Statistics Netherlands [database on the internet]. Voorburg/Heerlen: Statistics Netherlands. [updated daily; cited 2005 Oct 14]. Available from: http://www.cbs.nl
15. Altman DG, Royston P. What do we mean by validating a prognostic model? Statist Med. 2000;19:453-73.
16. Steyerberg EW, Harrell Jr FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54:774-81.

APPENDIX

Figure 3.14 shows EPICON in the form of a flowchart. EPICON consists of 5 steps (shown in between parentheses).

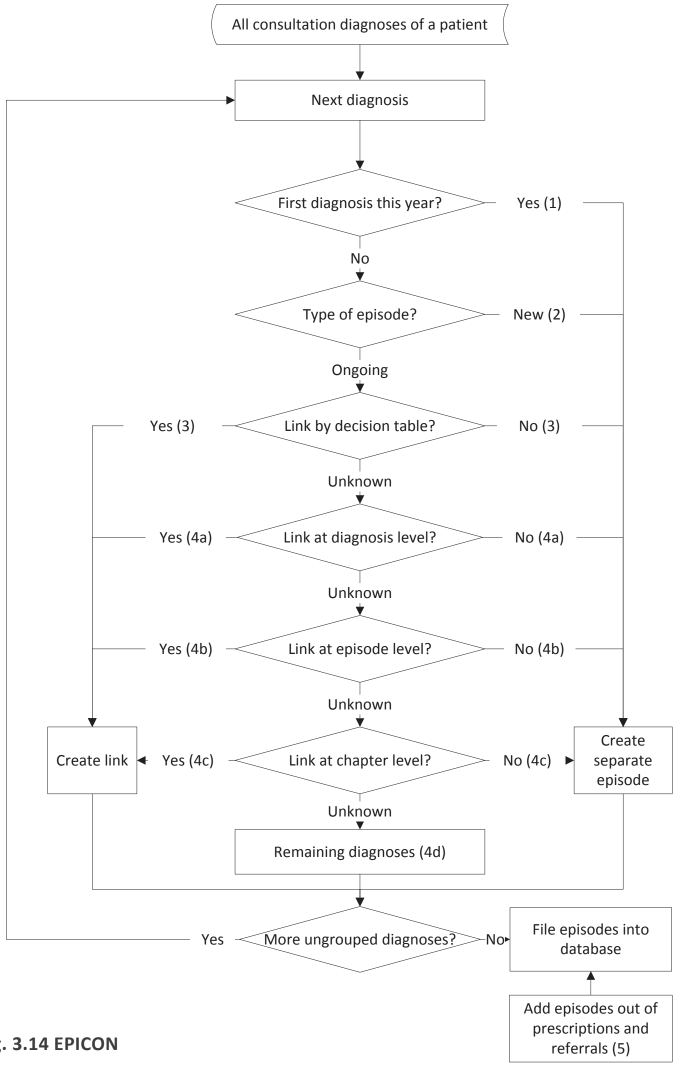


Fig. 3.14 EPICON

Step 1. The first consultation diagnosis of a patient in a one-year registration period is grouped as a separate episode (i.e. the diagnosis receives a separate episode number). Operationally, an episode is a row of diagnoses with the same episode number. The first **diagnosis** of this row can be either new or ongoing, whereas subsequent diagnoses are always ongoing. The first diagnosis in the row determines whether the **episode** is new or ongoing, the last disease diagnosis in the row determines the name of the episode. If no disease diagnosis exists, the episode is named after the last symptom diagnosis in the row.

Step 2. The remaining diagnoses are grouped by 'type of episode' (new or ongoing). A new diagnosis is always grouped as a separate episode.

Step 3. If possible, ongoing diagnoses are grouped by a decision table, which consists of a combination of an ongoing diagnosis and a previous diagnosis, and a decision whether or not they should be grouped together.

Step 4. The remaining diagnoses are grouped by case-based reasoning, which is a problem-solving approach in the field of artificial intelligence. A case-based reasoner solves a problem by remembering a previous similar situation and reuses information and knowledge from that situation. We used the manual grouping in the DNSGP-2 as a case library from which previous cases could be retrieved to solve the problem of grouping diagnoses. EPICON uses three kinds of cases, one at the level of diagnoses (4a), one at the level of episodes (4b), and one at the level of chapters (4c).

In step 4a, an unsolved case is defined as: Should diagnosis X be linked to one of the other **diagnoses** of the same patient? A solved case is defined as a manually grouped diagnosis X that belongs to a patient who also has one of the other diagnoses. EPICON reuses these solved cases to group diagnosis X.

In step 4b, EPICON uses the same procedure as in step 4a, but it uses a more coarse definition of cases in order to group any remaining diagnoses. Here, an unsolved case is defined as: Should diagnosis X be linked to one of the other **episodes** of the same patient?

In step 4c, the grain size of cases is increased once more, to the level of chapters, to handle rare cases that cannot be grouped by any of the preceding steps. In this step, an unsolved case is described as whether diagnosis X should be linked to one of the **chapters** to which other diagnoses of the same patient belong.

Diagnoses that cannot be grouped by any of the aforementioned steps are set aside in step 4d. EPICON continuously tries to group these remaining diagnoses, because it might still be possible to group some of these diagnoses when more information from following diagnoses of the same patient becomes available. In the end, EPICON groups all remaining diagnoses as separate episodes.

Step 5. Diagnoses from prescriptions and referrals (i.e., the indications for the prescription or referral) do not pass through the foregoing steps, because they have no 'type of episode'. Diagnoses from prescriptions and referrals that are different from any of the consultation diagnoses of a patient are added to the database as ongoing episodes.

EPI-0 consists of the same steps, except for the case-based part (4a - 4c), which does not exist in EPI-0.

Chapter 4

External validation of EPICON: A grouping system for estimating morbidity rates using electronic medical records

Marion CJ Biermans
Geert H Elbers
Robert A Verheij
Willem Jan van der Veen
Gerhard A Zielhuis
Pieter F de Vries Robbé

Journal of the American Medical Informatics Association 2008;15:770-775

ABSTRACT

Objective: To externally validate EPICON, a computerized system for grouping diagnoses from EMRs in general practice into episodes of care. These episodes can be used for estimating morbidity rates.

Design: Comparative observational study.

Measurements: Morbidity rates from an independent dataset, based on episode-oriented EMRs, were used as the gold standard. The EMRs in this dataset contained diagnoses which were manually grouped by GPs. The authors ungrouped these diagnoses and regrouped them automatically into episodes using EPICON. The authors then used these episodes to estimate morbidity rates that were compared to the gold standard. The differences between the two sets of morbidity rates were calculated and the authors analyzed large as well as structural differences to establish possible causes.

Results: In general, the morbidity rates based on EPICON deviate only slightly from the gold standard. Out of 675 diagnoses, 36 (5%) were considered to be deviating diagnoses. The deviating diagnoses showed differences for two main reasons: "differences in rules between the two methods of episode construction" and "inadequate performance of EPICON."

Conclusion: EPICON performs well for the large majority of the morbidity rates. We can therefore conclude that EPICON is useful for grouping episodes to estimate morbidity rates using EMRs from general practices. Morbidity rates of diseases with a broad range of symptoms should, however, be interpreted cautiously.

INTRODUCTION

General Practitioners (GPs) increasingly use Electronic Medical Records (EMRs) to record information regarding the treatment of patients. North Western Europe, especially Sweden, Denmark, and the Netherlands, has a leading position in the use of EMRs in general practice.[1;2] Primarily used for patient care, EMRs can also be utilized for scientific research, particularly the estimation of prevalence and incidence rates (i.e., morbidity rates) in general practice. Morbidity rates in general practice may provide a good indication of the health status of the general population. This is especially true for countries where the main pathway to medical care runs through general practice. These rates are useful for monitoring health in the population as well as for developing and evaluating health care policy.

In order to estimate morbidity rates, diagnoses recorded in EMRs in general practice need to be grouped into episodes of care. We developed EPIsode CONstructor (EPICON), a tool for the computerized grouping of diagnoses in general practice.[3] The EPICON tool renders it possible to estimate morbidity rates automatically from EMRs in general practice. Results from a previous evaluation study indicate that the internal validity of EPICON is adequate.[4] The present study is aimed at examining the external validity of EPICON through the use of an independent dataset.

Background

A diagnosis in general practice can refer to a symptom or a complaint (symptom diagnosis), a syndrome (nosological diagnosis), or a disease (pathological/pathophysiological diagnosis).[5] In this article, we will use the umbrella term “diagnosis” to refer to any of these categories. In order to estimate the numerator of morbidity rates, the diagnoses recorded in EMRs need to be coded and grouped into episodes of care (all encounters for the management of a specific health problem),[5] since double counting may occur if episodes are not grouped. For instance, if a patient visits the GP initially with abdominal pain, and during a second visit a few days later the pain appears to be based on appendicitis, only the appendicitis should be counted in the morbidity rates. The abdominal pain has to be linked to the episode appendicitis.

Generally, two approaches for constructing episodes can be distinguished. In one approach, diagnoses are recorded in contact-oriented EMRs. To be able to distinguish between incidence and prevalence, an indication is usually given whether a diagnosis represents the start of a new episode (new) or is part of an episode that started in the past (ongoing). In this approach, diagnoses are grouped into episodes afterwards, either through manual review or by using EPICON. In the second approach, diagnoses are directly recorded into episodes by the GP in episode-oriented EMRs.[6] As far as we know, these episode-oriented EMRs are used routinely only in the Netherlands and in Malta.[7] In this respect, the Netherlands is a leading country in the Western world. In the Netherlands, all GPs rely on EMRs. Most practices use contact-oriented EMRs, while some of them use the new generation of episode-oriented EMRs.

We developed EPICON, a case-based application for grouping into episodes the diagnoses from contact-oriented EMRs. The EPICON application is based on a combination of logical expressions, a decision table, and former cases in which diagnoses from patients in general practice were grouped manually. These former cases were derived from 89 practices that participated in the second Dutch national survey of general practice (DNSGP-2).[8;9] The development of this application has been described in detail elsewhere.[3]

The evaluation of EPICON, which can be qualified as a classification system, falls within the field of validating prognostic models. Different hierarchical levels of validation can be distinguished in evaluating prognostic models, starting with the internal validation (i.e., the performance of a system in the sample used to develop the system) as level 0.[10] In a previous study, we used a split sample approach to examine the internal validity of EPICON.[4] The dataset was split into two sets; one was used to develop EPICON and the other was used to test the developed system. Findings from the split-sample procedure showed that the internal validity of EPICON is adequate. Based on these results, EPICON has been brought into use to generate the yearly morbidity rates of practices with contact-oriented EMRs of the Netherlands Information Network of General Practice (LINH).[11;12]

The next levels of validation refer to the transportability of the system, i.e., whether the system has the ability to generate accurate results in a sample taken from a population other than the one that was used to developed the system. In an independent validation, several components of transportability, including historical, geographic, and methodologic transportability, can be distinguished.[10]

The goal of this study is to provide an independent validation of EPICON. We will use data from episode-oriented EMRs, in which GPs actually record diagnoses into episodes of care, as an independent dataset. Morbidity rates based on these GP-grouped episodes are considered the gold standard and will be compared to morbidity rates based on EPICON-grouped episodes. In this study, we will address the following research questions: 1)What is the deviation from the gold standard for morbidity rates that are based on EPICON-grouped episodes? 2)What are the causes of these deviations?

METHODS

Dataset

We used data from six general practices that participate in a network of general practices in the northern part of the Netherlands (RNG) and that record all patient data using episode-oriented EMRs.[13] We consider this dataset to be independent because the data from these practices were not used in the development of EPICON. We used data from 2002 through 2005, although in the case of one practice, the data from 2005 were not available at the time of analysis. The dataset consists of a total number of 473,350 diagnoses which were coded according to the International Classification of Primary Care (ICPC, first edition).[14] The six practices participate in LINH, and this network requires that all practices record whether a consultation diagnosis is new or ongoing, a necessity to distinguish between incidence and prevalence. For the six included practices, this field was derived from the GP-grouped episodes.*

Episode Constructions

The GPs manually recorded and grouped the diagnoses for all their patients into episodes at the moment of the consultation. A second episode construction that regrouped the same diagnoses automatically was created by EPICON. It is possible for EPICON to group episodes differently from the GPs, which is called a misclassification. There are three types of possible misclassifications: Link failure (type 1) occurs when a diagnosis, which is linked to another diagnosis by the GP, is not linked by EPICON. A

* Operationally, an episode is a row of diagnoses that carry the same episode number. The first diagnosis of a new episode number was characterized as new; all other diagnoses were marked as ongoing.

false link (type 2) originates when two diagnoses that were not linked by the GP are linked by EPICON. A wrong combination (type 3) occurs when both the GP and EPICON linked a diagnosis to another diagnosis, but the second diagnoses were different.[3] Misclassifications do not always cause differences in morbidity rates; they can only cause differences when they change a sufficient number of episode names (some misclassifications do change the sequence of links within an episode, but ultimately do not change the episode name).

Morbidity Rates

The next step in the study was to calculate two sets of prevalence and incidence rates: one set based on the GP-grouped episodes and the other set based on the EPICON-grouped episodes. The prevalence rate is defined as the proportion of the population with a particular disease during the period of one year, and the incidence rate refers to the occurrence of new episodes of a certain disease during the observed person-years at risk.

To establish the numerator of the prevalence rates, we counted, per episode name, the number of patients with at least one (new or ongoing) episode. When calculating the numerator of the incidence rates, we counted, per episode name, the number of new episodes. We used the mid-year population (i.e., the average of the population at the beginning and the end of each year) as the denominator, which varied from 24,067 in 2005 to 32,053 in 2003. Morbidity rates of diagnoses that occur only in the female (W, X) or male chapters (Y) are based only on the female or the male mid-year population. In total, we calculated morbidity rates for 675 different diagnoses.

Next, we compared the morbidity rates based on EPICON with the gold standard. In order to make this comparison, we distinguished between infrequent rates (less than one per 1000 patients per year according to the gold standard) and frequent rates (at least one per 1000 patients per year according to the gold standard), which divided the large number of rates into two approximately equal parts. The absolute differences were calculated for infrequent rates (absolute deviation), the relative differences, i.e., the percentages, for frequent rates (relative deviation).

Selection

We used the calculated differences to select a number of deviating diagnoses for further qualitative analysis. The criteria for selecting deviating diagnoses are based on sole outliers or structural differences:

- Criteria for deviating infrequent diagnoses:
  a. Absolute difference in at least one year < -0.2 or > 0.2 (sole outliers); or
  b. Absolute difference in all four years < -0.05 or > 0.05 (structural differences).
- Criteria for deviating frequent diagnoses:
  a. Relative difference in at least one year < -15% or > 15% (sole outliers); or
  b. Relative difference in all four years < -5% or > 5% (structural differences).

Qualitative Analysis

To clarify the causes of the observed differences, we conducted a qualitative analysis of these deviating diagnoses. The analysis included a detailed examination of both the documented grouping rules used by the GPs and the grouping methods used by EPICON (i.e., the algorithm, the decision table, and the cases). In addition, we interviewed one of the GPs and the coordinator of the registration network.

RESULTS

Morbidity Rates

Figure 4.1 shows the 2002 morbidity rates based on EPICON compared with the gold standard. Each scatter plot shows some hundreds of morbidity rates, named according to the ICPC.[14] (Because of the large number of rates, the plots show only part of all code names). For example, the prevalence of "other cardiovascular symptoms/complaints" (K29) in Figure 4.1a is 0.45 per 1000 patients per year in the gold standard (shown at the x-axis). The prevalence based on EPICON is 0.62 per 1000 patients per year which is a difference of 0.17 with the gold standard (shown at the y-axis).

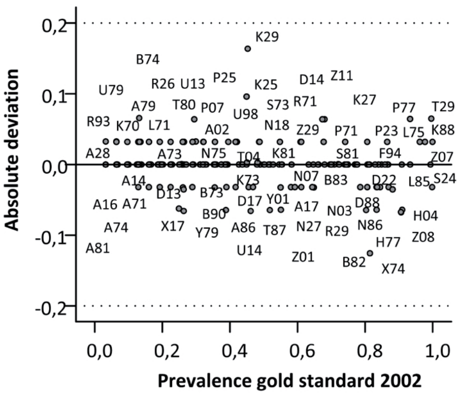


Fig. 4.1.a Deviation from the gold standard for infrequent prevalence rates.

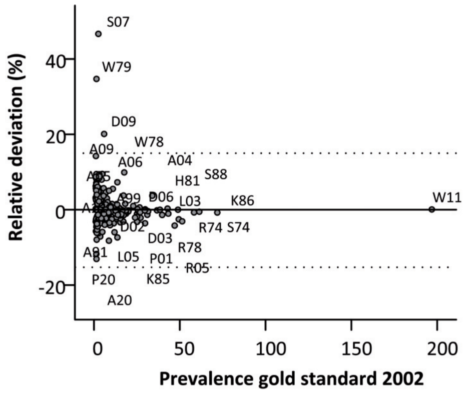


Fig. 4.1.b Deviation from the gold standard for frequent prevalence rates.

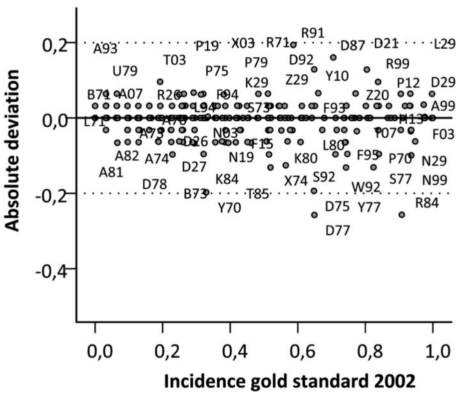


Fig. 4.1.c Deviation from the gold standard for infrequent incidence rates.

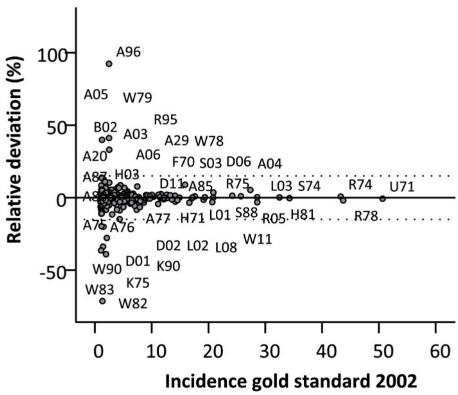


Fig. 4.1.d Deviation from the gold standard for frequent incidence rates.

- **A morbidity rate, per 1000 patients per year, with corresponding code name from the International Classification of Primary Care14**

- - Limit for sole outliers (a morbidity rate beyond this line is considered as deviating)

Fig. 4.1 **Deviation of morbidity rates (based on EPICON-grouped episodes) from the gold standard (i.e., morbidity rates based on GP-grouped episodes), 2002.**

Figures 4.2-4.5 (Appendix) show the results for the morbidity rates of 2003 through 2005. The figures show a similar pattern in each category. In general, the infrequent morbidity rates based on EPICON (figures 4.1a, 4.1c, 4.2, and 4.4) deviate only slightly from the gold standard; the deviation of most rates is either 0 or close to 0. There are just a few deviating diagnoses, i.e., sole outliers (rates beyond the dotted lines) or structural differences (not clearly visible in the figures, but listed in the last column of Table 4.1). Although the frequent morbidity rates (Figures 4.1b, 4.1d, 4.3, and 4.5) show more deviating diagnoses, with most of them occurring among the frequent incident rates (Figures 4.1d and 4.5), the large majority of the frequent morbidity rates deviates not at all or just a few percent. With the exception of the diagnoses that comply with the selection criteria, there is no systematic bias, because the rates are proportionally distributed above and below the x-axis.

Out of all 675 analyzed diagnoses, 36 diagnoses (5%) were deviant. Five of these diagnoses were not further scrutinized because they included fewer than five misclassifications in a three year period; i.e., “haematemesis/vomiting blood” (D14), “fear of pregnancy” (W02), “unwanted pregnancy” (W79), “benign neoplasm female genital” (X80), and “genital herpes female” (X90). The remaining 31 diagnoses were analyzed in detail.

Causes of Deviations

Table 4.1 shows all deviating diagnoses. Similar diagnoses that deviate for the same reason are grouped together into one category. We found two main causes for deviations: differences in rules for grouping, and inadequate performance of EPICON.

Differences in Rules

Differences in the rules for grouping used by GPs and EPICON was the reason for the deviations found in the categories “diagnoses related to death,” “diagnoses related to pregnancy,” and some of the “unrelated diagnoses” (i.e., “elevated blood pressure” (K85), “hypertension complicated” (K87), and “chronic obstructive pulmonary disease” (R95)).

For instance, when looking at the incidence rates of “death” (A96) and “pregnancy” (W78), it becomes clear that the rates based on EPICON are higher than the gold standard, whereas the incidences of some causes of death (cardiovascular diseases and

Table 4.1 Deviating diagnoses per category

Deviating diagnoses (ICPC)	Selection criteria*				
	Sole outliers				In all
	2002	2003	2004	2005	4 years
Diagnoses related to death					
Euthanasia request/discussion (A20)			I>15%	I>0.2	I>5%
Death (A96)	I>15%	I>15%	I>15%	I>15%	I>5%
Suicide/suicide attempt (P77)					I<0.05
Cardiovascular diseases					
Acute myocardial infarction (K75)	I<-15%	I<-15%	I<-15%	I<-15%	I<-5%
Heart failure (K77)					I<-5%
Stroke/cerebrovascular accident (K90)	I<-15%	I<-15%	I<-15%	I<-15%	I<-5%
Neoplasms					
Malignant neoplasm colon/rectum (D75)			I<-0.2		I<-0.05
Malignant neoplasm digestive other (D77)	I<-0.2				I<-0.05
Malignant neoplasm bronchus/lung (R84)	I<-0.2		I<-0.2	I<-0.2	I<-0.05
Malignant neoplasm breast female (X76)		I<-15%			
Malignant neoplasm prostate (Y77)					I<-0.05
Diagnoses related to pregnancy					
Pregnancy (W78)				PI>15%	PI>5%
End points of pregnancy					
Abortion spontaneous (W82)	I<-15%	I<-15%	I<-15%	I<-15%	I<-5%
Abortion induced (W83)	I<-15%	I<-15%	I<-15%	I<-0.2	I<-5%
Uncomplicated labour/delivery livebirth (W90)	I<-15%	I<-15%	I<-15%	I<-15%	I<-5%
Complicated labour/delivery livebirth (W92)				I<-0.2	I<-0.05
Adverse effects					
Adverse effect medical agent (A85)					P>5%
Possible symptoms					
Sweating problem (A09)					I>5%
Nausea (D09)	P>15%	P>15%			P>5%
Rash generalized (S07)	P>15%	P>15%	P>15%	P>15%	P>5%
General diagnoses					
General deterioration (A05)	I>15%	I>15%	I>15%	I>15%	I>5%
General symptom/complaint other (A29)					I>5%
Complication of medical treatment (A87)					I>5%
Disease digestive system, other (D99)	I<-15%				
Other arterial obstruction/PVD (K92)		I<-15%			I<-5%
Cardiovascular disease other (K99)			I<-15%		I<-0.05
Unrelated diagnoses					
Elevated blood pressure (K85)					P<5%
Hypertension complicated (K87)		I>15%			
Epilepsy (N88)		I>0.2			
Chronic obstructive pulmonary disease (R95)	I>15%				I>5%
Syphilis male (Y70)					I<-0.05

*** P=Prevalence, I = Incidence**

neoplasms) and end points of pregnancy are lower than the gold standard. This difference can be attributed to the fact that GPs use the rule that "death" (A96) or "pregnancy" (W78) should be linked to the cause of death or the end point, whereas EPICON uses the opposite rule, classifying "death" (A96) or "pregnancy" (W78) as a separate, new episode.

The main cause for differences in the morbidity rates of "adverse effect medical agent" (A85) is a difference in the rule for naming the episodes. The EPICON rule is that the last disease code (ICPC codes 70–99) in time is used as the episode name for that episode. When there is no disease code in the episode, the last symptom code (ICPC codes 1–29) is used as the episode name. GPs can choose every diagnosis as the episode name.

Furthermore, the interviews revealed that there are differences in handling the rules. The GPs, who group manually, vary in their use of the grouping rules, whereas EPICON groups automatically, and consequently, EPICON has no inter-doctor variation.

Inadequate Performance of EPICON

Differences in the category "general diagnoses," "possible symptoms of adverse effects," and some of the "unrelated diagnoses" (i.e., "epilepsy" (N88) and "syphilis male" (Y70)) are mainly due to problems in EPICON's performance. In some cases, EPICON displays linking problems, for instance, EPICON frequently failed to link possible symptoms of adverse effects, such as "nausea" (D09) to "adverse effect medical agent" (A85). In other cases, EPICON both displays link failures and creates false links. The main reason for this problem is that many different diagnoses can be linked to unspecified and general diagnoses. Moreover, the interviews showed that within the category of "general diagnoses," a large inter-doctor variation exists in recording diagnoses as they are drawn from a reservoir of unknown or infrequently occurring diagnoses.

DISCUSSION

In this study, we examined the external validity of EPICON, a system for grouping diagnoses into episodes whose purpose is to estimate morbidity rates in general practice. We used an independent dataset derived from GPs who record diagnoses using episode-oriented EMRs. Morbidity rates based on the GP-grouped episodes were considered the gold standard and compared to morbidity rates based on

EPICON-grouped episodes. The results indicate that EPICON performs well for the large majority of diagnoses.

Only 5% (n = 36) of all analyzed diagnoses (n = 675) shows a substantial deviation in at least one year or a structural small deviation over four years. We found more deviations in rates for frequent than for infrequent diagnoses, because differences in episode construction have more effect on frequent than on infrequent rates. Except for these deviating diagnoses, we did not observe any systematic bias.

An explanation for part of these deviating diagnoses is a difference in grouping rules. Some of the rules used by EPICON for grouping diagnoses differ from those utilized by GPs. For instance, the GPs linked "death" (A96) to the cause of death, whereas EPICON is based upon a dataset in which it was decided to link "death" not to the cause of death. Both decisions have their advantages and disadvantages and it is not possible to claim that one rule is more valid than the other. In addition, we found differences in handling the rules and in naming the episodes.

Other deviating diagnoses are explained by inadequate performance of EPICON. This accounts, in particular, for deviations in unspecified and general diagnoses since many symptoms can be grouped within these diagnoses. The EPICON application is based on the probability that the cases in the training set were grouped. It does not use the same information in the grouping process as the GP does, such as age, gender, and the duration of the disease, which may cause misclassifications. Furthermore, it is possible that some cases did not occur in the training set. In all likelihood, however, EPICON will not be used for these unspecified and general diagnoses, because they are considered less important for epidemiological research. Should EPICON be used for these diagnoses, the resulting morbidity rates should be interpreted with caution. The EPICON application might be adjusted for these insufficiencies by: a) adding cases, and b) including variables such as gender, age, and duration of the disease into EPICON.

The strength of this study is that we used an independent dataset to examine the external validity of a prognostic model. There is not much research describing an external validation of a prognostic system.[15] Another strength is that we performed a quantitative as well as a qualitative analysis, providing insight into both the number of deviating rates and the main causes of these deviations.

A limitation of this study is that the criteria for defining deviating diagnoses are to some extent arbitrary, so should the criteria be altered, we will find less or more deviating codes. In addition, our judgement that the large majority of the morbidity rates deviates only slightly, is also subjective, although with the figures we provided, it is possible to judge deviations for oneself. Furthermore, our conclusions about the external validity of EPICON are based upon one test that was carried out in one setting.

This one test does provide some insight into the historical, geographical, and methodological transportability of EPICON. Regarding the historical transportability, EPICON was originally developed using data from 2001, and in this study, was applied to data from 2002 through 2005. The historical transportability of EPICON seems adequate as we did not observe any problem in the application to another time period. In the long run, however, changing medical insights that affect grouping rules may reduce the historical transportability.

Regarding the geographical transportability, EPICON is based on data from a nationally representative patient population, derived from 89 general practices throughout the Netherlands.[3] The six practices that were used for the external validation are located in the Northern part of the Netherlands. The patient population of these six practices is quite similar to the nationally representative population and we encountered no differences in episode-construction that could be attributed to the difference in geographical location. This cautiously indicates that EPICON is transportable to similar datasets in other regions or countries. We need more evidence, however, before we can draw any firm conclusions. If EPICON is applied to data from another geographical area, this ungrouped dataset should be compared first to the dataset upon which EPICON is based. For instance, the similarity of the age and gender distributions of both populations, and the frequencies of ungrouped diagnoses could be examined. The EPICON application should not be used if this comparison reveals large differences, such as many diagnoses of diseases that do not occur in the Netherlands.

Additionally, this study provides some insight into methodological transportability, because two different methods of collecting data were used. In this study we applied EPICON to data from episode-oriented EMRs whereas it was originally developed using data from contact-oriented EMRs. The EPICON tool is thus applicable to both contact-oriented and episode-oriented EMRs. A possible application for EPICON in episode-oriented EMRs is to discover differences in (handling) the rules used by different GPs. However, a limitation of the application of EPICON to contact-oriented EMRs is that the characterization of a diagnosis as either “new” or “ongoing” might be lacking. Although this field can be easily added to an existing EMR system, this solution does require extra recording for the GP.

These results regarding the generalizibility of EPICON need to be confirmed in cumulative tests across diverse settings. Furthermore, an important topic for future research is the extent to which differences in (the design of) EMRs affect the morbidity rates derived from these EMRs.

CONCLUSION

This study shows that the external validity of EPICON is sufficient for the purpose of estimating morbidity rates in general practice. Only a limited number of diagnoses (5%) deviates from the gold standard. There are two main causes for these deviations: inadequate performance of EPICON, and differences in rules. The latter cause seems particularly to apply to deviations in diagnoses related to death and pregnancy. The EPICON application performs less well when it comes to unspecified and general diagnoses, and hence caution is required when EPICON is used for these rates.

REFERENCES

1. European physicians especially in Sweden, Netherlands and Denmark, lead U.S. in use of electronic-medical records. Health Care Res. 2002;2:1-3.
2. Burt CW, Hing E, Woodwell D. Electronic medical record use by office-based physicians: United States, 2005 [document on the internet]. Hyattsville: National Center for Health Statistics; 2008 [updated 2007 Jan 11; cited 2008 Jun 22]. Available from: http://www.cdc.gov/nchs/products/pubs/pubd/hestats/electronic/electronic.htm.
3. Biermans MCJ, De Bakker DH, Verheij RA, Gravestein JV, Van der Linden MW, De Vries Robbé PF. Development of a case-based system for grouping diagnoses in general practice. Int J Med Inform. 2008;77:431-439.
4. Biermans MCJ, Verheij RA, De Bakker DH, Zielhuis GA, De Vries Robbé PF. Estimating morbidity rates from electronic medical records in general practice. Evaluation of a grouping system. Methods Inf Med. 2008;47:98-106.
5. WONCA Classification Committee. An international glossary for general/family practice. Fam Pract. 1995;12:341-69.
6. Okkes IM, Groen A, Oskam SK, Lamberts H. Advantages of long observation in episode-oriented electronic patient records in family practice. Methods Inf Med. 2001;40:229-35.
7. Lagasse R, Desmet M, Jamoulle M, Correa G, Roland M, Hoyois P, et al. European situation of the routine medical data collection and their utilisation for health monitoring. EURO-MED-DATA. Final Report. Bruxelles: Université Libre de Bruxelles; 2001.
8. Westert GP, Schellevis FG, De Bakker DH, Groenewegen PP, Bensing JM, Van der Zee J. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15:59-65.
9. Westert GP, Jabaaij L, Schellevis FG, editors. Morbidity, performance and quality in primary care: Dutch general practice on stage. Oxon: Radcliffe Publishing; 2006.
10. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130:515-24.
11. Tacken MAJB. Quality of preventive performance in general practice: the use of routinely collected data [dissertation]. Radboud University Nijmegen; 2005.
12. Verheij RA, Jabaaij L, Abrahamse H, Van den Hoogen H, Braspenning J, Van Althuis T. Netherlands Information Network of General Practice: Facts and figures on Dutch GP care [website on the internet]. Available from: http://www.linh.nl.

13. Van der Veen WJ, Meyboom-de Jong B. Age and Gender. In: Jones R, Britten N, Culpepper L, Gass DA, Grol R, Mant D, et al, editors. Oxford textbook of primary medical care. Oxford: Oxford University Press; 2004. 153-161.
14. Lamberts H, Woods M, editors. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.
15. Stolwijk AM, Straatman H, Zielhuis GA, Jansen CAM, Braat DDM, Van Dop PA, Verbeek ALM. External validation of prognostic models for ongoing pregnancy after in-vitro fertilization. Hum Reprod. 1998;13:3542-49.

APPENDIX CHAPTER 4

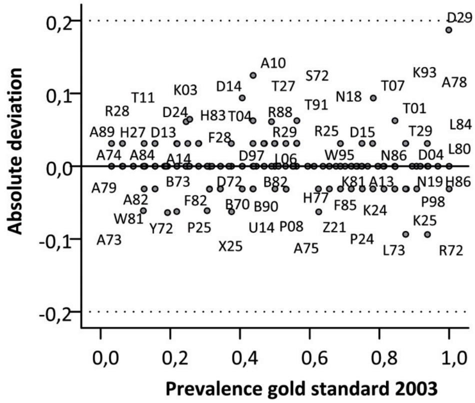


Fig. 4.2 **Deviation from the gold standard for infrequent prevalence rates.**

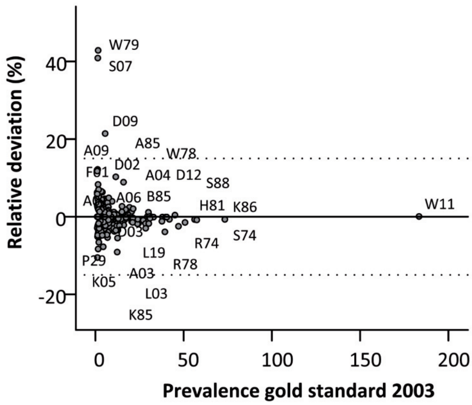


Fig. 4.3 **Deviation from the gold standard for frequent prevalence rates.**

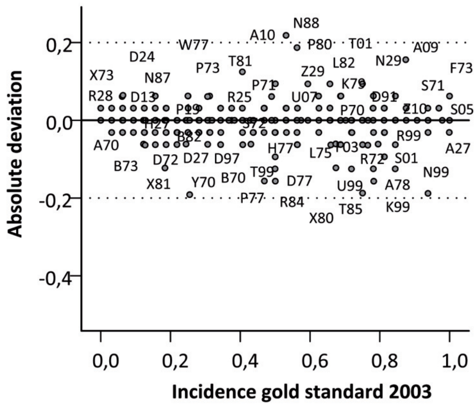
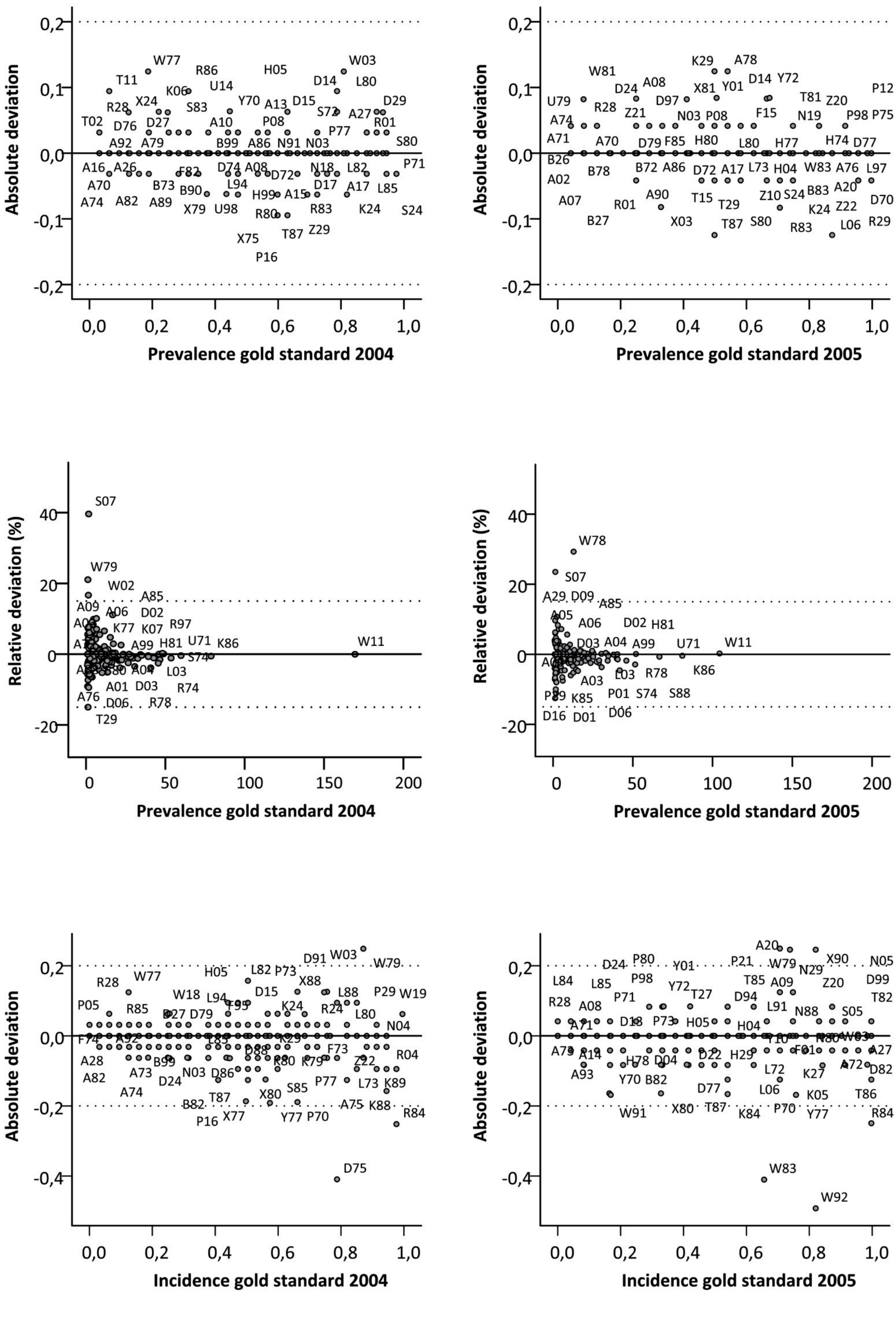


Fig. 4.4 **Deviation from the gold standard for infrequent incidence rates.**


Absolute deviation
Prevalence gold standard 2004
Prevalence gold standard 2005
Relative deviation (%)
Prevalence gold standard 2004
Prevalence gold standard 2005
Absolute deviation
Incidence gold standard 2004
Incidence gold standard 2005

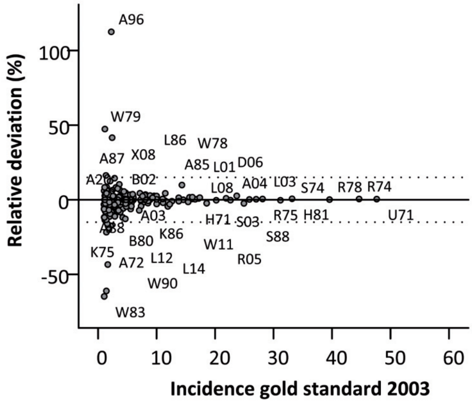
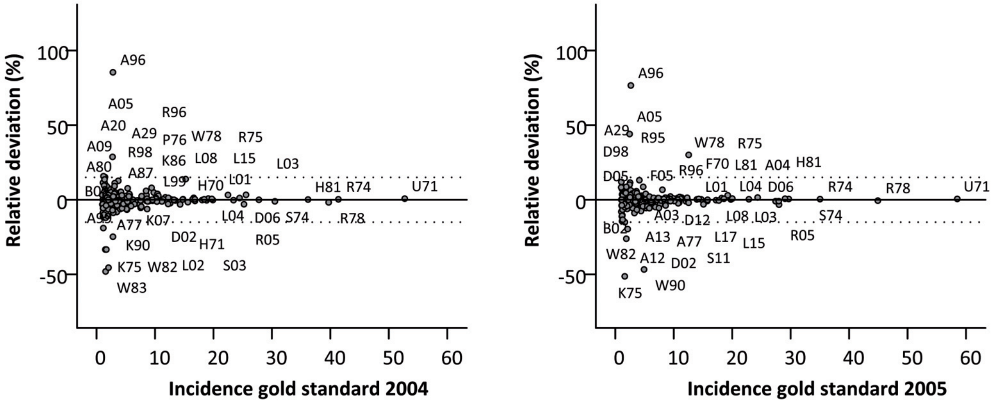


Fig. 4.5 **Deviation from the gold standard for frequent incidence rates.**


Relative deviation (%)
100
50
0
-50
A96
A05
R96
A20
A29
P76
W78
R75
A09
R98
K86
L08
L15
L03
A80
A87
L99
H70
L01
H81
R74
U71
L04
D06
S74
R78
A77
K07
K90
D02
H71
R05
K75
W82
L02
S03
W83
0
10
20
30
40
50
60
Incidence gold standard 2004
Relative deviation (%)
100
50
0
-50
A96
A05
A29
R95
W78
R75
D98
F70
L81
A04
H81
D05
F05
R96
L01
L04
D06
R74
R78
U71
A03
D12
L08
L03
S74
B02
A13
A77
L17
L15
R05
W82
A12
D02
S11
K75
W90
0
10
20
30
40
50
60
Incidence gold standard 2005

Chapter 5

Striking trends in the incidence of health problems in the Netherlands (2002-05): Findings from a new strategy for surveillance in general practice

Marion CJ Biermans
Peter Spreeuwenberg
Robert A Verheij
Dinny H de Bakker
Pieter F de Vries Robbé
Gerhard A Zielhuis

Submitted for publication

ABSTRACT

Background: This study aimed to detect striking trends based on a new strategy for monitoring public health.

Methods: We used data over four years from electronic medical records of a large, nationally representative network of general practices. Episodes were either directly recorded by general practitioners or were constructed using a new record linkage method (EPICON). The episodes were used to estimate raw morbidity rates for all codes of the International Classification of Primary Care. Multilevel Poisson regression models were used to analyse the trend over time for 15 health problems that showed an obvious change over time. Based on these models, we calculated adjusted incidence rates corrected for clustering, sex, and age.

Results: During 2002-05, both men and women increasingly consulted the general practitioner because of concern about a drug reaction, a change in faeces/bowel movements, and urination problems. Men showed an increase in consultations for prostate problems and venereal diseases. The incidence of chronic internal knee derangement decreased for both sexes. Women consulted their general practitioner less frequently about sterilization and fear of being pregnant.

Conclusion: The strategy developed proved to be useful to detect trends across a short period of time. Changes in the health care market, such as the increasing availability of over-the-counter drugs and various large advertising campaigns for medications may explain some of the findings. The increasing incidence of health problems in the urogenital area deserves attention as it could reflect increases in the incidence of sexually transmitted diseases and urinary tract infections.

INTRODUCTION

Morbidity trends in general practice provide important information for public health agencies, because they reflect, to some extent, morbidity trends in the general population. In particular, this holds for countries that use a gatekeeping system, because these systems utilize general practices as the main pathway to medical care. In countries where patients have free access to specialist care, trends in general practice could still provide an indication of morbidity changes in the population, although there is a higher chance of morbidity 'leaking' to other health care providers.

Electronic medical records (EMRs) in general practice can be a valuable source for estimating these trends, because they can provide the large and longitudinal datasets that are needed to observe morbidity over time. In the past, this routine data has been widely used to estimate morbidity trends of specific diseases. However, using this data to signal striking and unusual changes in a large range of health problems, is only just emerging.[1] This can probably be attributed to the complexity of this type of research, which has various requirements.

For starters, it calls for a sufficiently large registration network of general practitioners (GPs) who accurately code diagnoses of their patients using a coding system. Next, morbidity rates (i.e. incidence and prevalence rates) over specific time periods have to be computed. In order to estimate the numerators of these rates, contact diagnoses have to be grouped into episodes of care. An episode of care includes "all encounters for the management of a specific health problem".[2] An episode could be either one diagnosis or a sequence of diagnoses that reflects the course of a disease over time. To be able to distinguish between incidence and prevalence, we also need an indication whether the presented health problem represents the start of a new episode or is part of an episode that started in the past.

Generally, two approaches for constructing episodes can be distinguished. The first approach involves the use of episode-oriented EMRs. In this new generation of EMRs, the GP records diagnoses directly into episodes.[3] In the second approach, diagnoses from contact-oriented EMRs are grouped afterwards, through manual review or use of a computerised method. In a previous paper, we presented EPICON, an application for automatically grouping diagnoses from contact-oriented EMRs into episodes of care. Both the development and the evaluation of EPICON have been described in detail elsewhere.[4-6]

In addition, the size of the population at risk in the denominator has to be determined. In countries where every patient is registered with a GP (so called list system), the denominator can easily be determined, while more complicated approaches are available for countries without a list system.[7] Finally, a method has to be developed to detect trends in morbidity rates over time that takes into account that count data are used which are clustered within patients and within practices.

In this study, we developed a strategy that met all these requirements, and involved the application of EPICON to yearly data from a very large, nationally representative computerised network of Dutch general practices,[8;9] and the use of multilevel poisson regression models. The Netherlands has favourable conditions for monitoring public health based on data from EMRs in general practice, because it utilises both a gatekeeping and a list system. The aim of our research is to detect striking trends in the incidence of health problems in the Netherlands during 2002-2005.

METHODS

Dataset

We used data from practices participating in the Netherlands Information Network of General Practice (LINH).[8;9] This network consists of a large and dynamic pool of practices, which differs from year to year as some practices leave the network and others join up. The GPs within this network assign codes from the International Classification of Primary Care (ICPC) to consultations, prescriptions, and referrals.[10] Medical information about out of hours care is transferred to the GPs' offices, coded, and included in the database as well. The majority of the practices uses contact-oriented EMRs and a minority uses episode-oriented EMRs. In a previous study, we compared morbidity rates between EPICON-grouped and GP-grouped diagnoses. The results of that study indicate that morbidity rates based on EPICON-grouped diagnoses from contact-oriented EMRs can be used in conjunction with morbidity rates based on GP-grouped diagnoses from episode-oriented EMRs provided that the same grouping rules are used.[6] The present study includes only health problems (15 in total) for which the rules used by EPICON for grouping diagnoses in the contact-oriented EMRs are similar to those utilised by the GPs in the episode-oriented EMRs.

In this study, we included data from 2002 (the first year in which EPICON could be applied) through 2005 (the last year for which data were available at the time of analyses). The total number of LINH-practices was 83 in 2002, 80 in 2003, 61 in 2004 and 71 in 2005, but only practices were included that met certain criteria for accuracy and completeness. This resulted in the following number of practices in the analyses: 69 in 2002, 66 in 2003, 43 in 2004, and 42 in 2005 (220 practice-years in total).

The included practices are representative of all Dutch practices in terms of urbanisation level and practice type (i.e. solo or group practice), but practices in the northern part of the country are slightly overrepresented at the expense of practices in the western part.[11] For each year, the total patient population of these practices provides a representative sample of at least 1% of the Dutch population regarding age and gender.[12]

Raw morbidity rates

Several steps were taken to compute raw morbidity rates. For contact-oriented practices, episodes were constructed using EPICON. For episode-oriented practices, (6 in 2002, 6 in 2003, 6 in 2004, and 5 in 2005), we used the episodes that were recorded by the GP. The episodes were weighted for the length of recording of a practice within one year, to account for vacation, sick leave, etc. (198 of the 220 included practice-years covered a complete year and 22 covered slightly less than a year).

Next, we estimated the yearly morbidity rates. For the numerators, we counted, per episode name, the number of patients with at least one episode (prevalence) and the number of new episodes (incidence). For the denominators, we used the mid-year population (i.e. the average of the population at the beginning and the end of a year). The total mid-year population varied from 174138.5 in 2005 to 296716.0 in 2002.

We then computed raw incidence rates for the total population, for men and women separately, and for various age groups (0, 1-4, 5-14, 15-24, 25-44, 45-64, 65-74, 75+). These raw morbidity rates are comparable to the rates in the second Dutch National Study of General Practice.[13;14]

Out of the raw incidence rates for all ICPC codes (N=683), we selected 15 health problems that showed an obvious increase or decrease over time for further investigation.

Crude and adjusted incidence rates

We used multilevel Poisson regression models to analyse the trend over time for the selected health problems.[15;16] Because of the large number of dependent variables (15), we developed general models that could be applied to each outcome. For all models, the dependent variable is the number of new episodes of a health problem in a year for one patient.

These multilevel analyses were used, because the data is clustered within levels: repeated observations of a health problem are clustered within patients (level one) who are clustered within general practices (level two). We modelled (i.e., corrected for) the variance due to clustering within levels in the random part of all models. The fixed part of all models includes the four years which is necessary to test for a trend over time. For the total population, we used a crude and an adjusted model. The crude model, which corrects just for clustering, includes only the four years. The adjusted model includes the four years controlling for sex and age.

For each health problem, we stratified by sex, and used the adjusted model (without controlling for sex) to test for a linear trend over time. The adjusted model was also used for the total population to test whether the age effect differed between two years. Statistical significance was set at $p<0.05$.

Based upon the crude and the adjusted model, we calculated incidence rates for the total population and for men and women separately.

RESULTS

Looking at the raw incidence rates, eight health problems increased and seven health problems decreased considerably over the 2002-2005 period. Table 5.1 shows the results of the analyses over time. Nine of the 15 selected health problems show a significant increase (6) or decrease (3) over time.

Table 5.1 Results of tests for a linear trend over time 2002-2005 (p-values)

Health problems (ICPC code)[a]	Males	Females
Increasing based on raw incidence rates		
Concern about drug reaction (A13)	0.0008*	0.0039*
Change in faeces/bowel movements (D18)	0.0015*	0.0010*
Hypertension with involvement target organs (K87)	0.1797	0.1049
Frequent/urgent urination (U02)	0.0261*	0.0013*
Other urination problems (U05)	0.0004*	0.0005*
Family planning/other (W14)	n/a	0.1573
Symptoms/complaints prostate (Y06)	0.0003*	n/a
Fear of venereal disease (Y25)	0.0001*	n/a
Decreasing based on raw incidence rates		
Sprains & strains of other joints (L79)	0.1821	0.1229
Chronic internal knee derangement (L97)	0.0028*	0.0547
Other head injury without skull fracture (N80)	0.1738	0.3482
Hypertrophy/chronic infection tonsils/adenoids (R90)	0.0719	0.4237
Diaper rash (S89)	0.6315	0.4708
Fear of being pregnant (W02)	n/a	0.0113*
Family planning/sterilization (W13)	n/a	0.0000*

a: International Classification of Primary Care
*** Statistically significant at $p < 0.05$**

Figure 5.1 shows the adjusted incidence rates for all significantly changing health problems and for ‘hypertrophy/chronic infection tonsils/adenoids (R90)’.

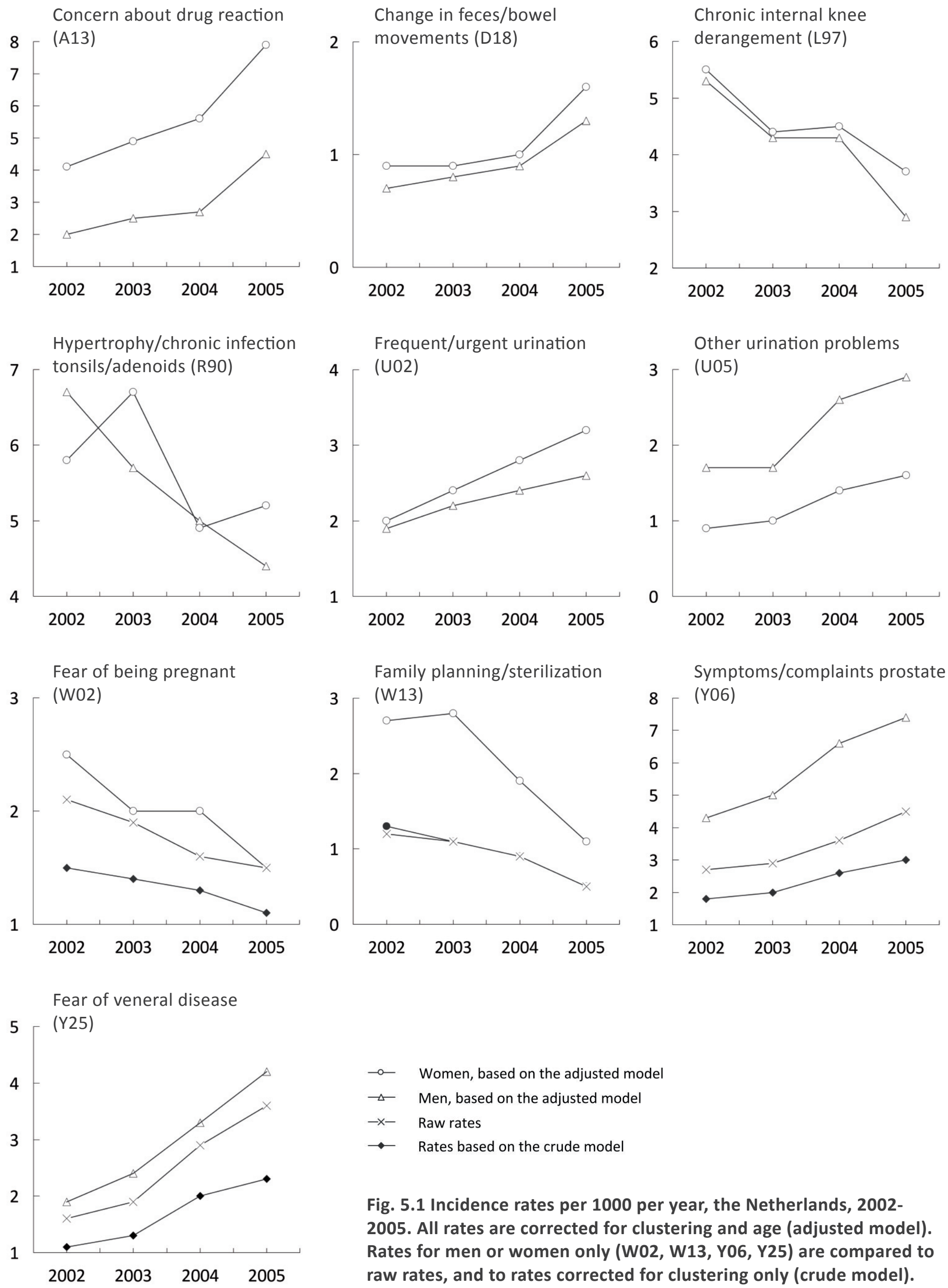


Fig. 5.1 Incidence rates per 1000 per year, the Netherlands, 2002-2005. All rates are corrected for clustering and age (adjusted model). Rates for men or women only (W02, W13, Y06, Y25) are compared to raw rates, and to rates corrected for clustering only (crude model).

'Concern about drug reaction (A13)' is a common problem in general practice, especially among women and the elderly. Within the ICPC coding system, the symptom-diagnosis 'concern about drug reaction (A13)' is distinguished from the disease-diagnosis 'adverse effect medical agent proper dose (A85)'. Both men and women increasingly consulted the GP because of concern about a reaction to drugs, especially in 2005. The increase occurs mainly among patients of 45 years of age and older, which results in a highly significant change of the effect of age over time ($p = 0.0000$).

GPs use the uncommon diagnosis 'change in faeces/bowel movements (D18)' for a change in the pattern of defecation. Separate ICPC-codes exist for 'diarrhoea (D11)' and 'constipation (D12)'. In particular, infants and elderly patients are bothered by changes in faeces/bowel movements. The increase, however, did not occur among the youngest and oldest, but among people in the age range of 15 through 74.

'Chronic internal knee derangement (L97)' starts to occur at the age of five and is a frequent problem in the age group of 15 through 64. Chronic internal knee derangement decreased for men and women in the examined time period. The effect of age over time changed significantly ($p = 0.0002$) for this health problem due to a remarkable decrease among young people (15 through 24 years of age).

'Hypertrophy/chronic infection tonsils/adenoids (R90)' shows borderline significance for men, but not for women. This health problem, which occurs mainly in infants and children, decreased especially in male infants and young boys (the raw incidence decreased from 18.5 to 1.6 for male infants and from 27.0 to 21.1 per 1000 per year for boys from one through four years of age). For female infants and young girls, the trend also decreases, but it fluctuates more over time.

'Frequent/urgent (U02) and other urination problems (U05)' are common problems among the elderly. 'Other urination problems (U05)' refers to problems other than 'painful urination (U01)' and 'incontinence (U04)' for which separate ICPC-codes exist. Frequent, urgent, and other urination problems show an increase in various age groups, but in particular in people aged 75 and over.

Women who consult the GP for 'fear of being pregnant (W02)' are usually between 15 and 44 years of age. In the 2002-2005 period, the number of consultations for fear of pregnancy decreased, especially among young women (15 through 24 years of age).

Consultations for 'family planning/sterilization (W13)' among women start at the age of 20, increase until the age of 40, and then decline to zero at 50 years of age. Starting in the year 2003, women decreasingly consulted their GP about sterilization, a decline which occurred especially among women in their thirties.

We found a considerable increase for 'symptoms/complaints prostate (Y06)', a health problem which commences among men at the age of 25 and is a common problem among elderly males. Prostate symptoms/complaints increased for both middle-aged and older men.

Starting at age 15, young men frequently consult their GP about 'fear of venereal disease (Y25)'. These consultations peak among men in their twenties, and then decline slowly to zero among the elderly. This health problem shows a considerable increase among men between 15 and 44 years of age.

To illustrate the effect of clustering on the incidence rates, we added both the raw rates (no corrections) and the crude rates (corrected for clustering only), to four health problems in figure 5.1. The differences between the raw rates and the crude rates clearly show that clustering may have considerable impact on incidence rates in general practice. The crude rates are generally lower than the raw rates, because part of the variance of the raw incidence is attributable to the fact that these data are clustered within patients and within practices. The adjusted rates (corrected for clustering and age) are usually higher than the crude rates because they show the incidence of a patient with a mean age (i.e. the age variables were centered).

DISCUSSION

In this study, we monitored trends in the incidence of health problems in general practice through the application of a new method for record linkage and the use of multilevel Poisson regression models for analysing trends in clustered data. The results show that the incidence of nine health problems changed significantly over the 2002-2005 period in the Netherlands.

Dataset

Our results are based on data obtained from a large network of computerised general practices in the Netherlands, a dataset which has several advantages. First, the patient population of the included practices forms a representative sample of the Dutch population regarding age and gender. Second, the health of this patient population reflects the health status of the general population because the Netherlands uses a gate keeping system. Third, the dataset is large enough to find trends across time, and finally, the use of routine data is efficient and reduces the risk of information bias.

A limitation of the dataset is that some practices had to be excluded from the analyses because their record keeping was not sufficiently, which resulted in a considerable loss of practices for 2004 and 2005. Furthermore, the selected practices record not always 100% completely, which could mean that the calculated rates underestimate the true rates to a limited extent. Recently, the network developed an application to improve the quality of recording. The application uses a number of indicators to measure the quality of recording of a practice. These measurements are used to provide the practice with automatic feedback on its quality of recording.

METHODS

We used EPICON, a new application to disclose data from EMRs in general practice for estimating morbidity rates. EPICON constructs episodes from ICPC-codes, which can be used to estimate the numerator of morbidity rates. Results from studies assessing the internal and external validity show that EPICON performs adequate for the purpose of estimating morbidity rates.[5;6]

In our analyses, we correct for clustering of observations within patients and within practices through the use of multilevel Poisson regression models. The results show that this clustering may have considerable impact on the incidence rates. It is therefore advisable to correct incidence rates for clustering when data is clustered within levels, for instance in incidence rates based on samples stratified by general practice, hospital, or neighborhood.

Interpretation of findings

To interpret the findings, the full range of factors that could affect the registration of patient data in EMRs in general practice has to be taken into account. In box 5.1, we distinguish between four categories of possible causes of changing incidence rates. The first category falls within the domain of medical informatics, which comprises aspects of the validity of using EMRs in general practice for the purpose of estimating morbidity rates. In the previous paragraphs, we discussed our findings from this perspective. The second and third categories are related to the area of health policy; the fourth category is the area of epidemiologic research. Before considering a real change in the incidence of disease, possible causes that fall into the first three categories should be excluded.

Box 5.1 Possible causes of changing incidence rates

Category 1: Changes in recording as a result of

a) change in GPs' perspective of disease;
b) change in rules for coding diagnoses;
c) change in rules for grouping diagnoses into episodes;
d) change in quality of recording;
e) change in general practice software;
f) unknown change in recording.

Category 2: Changes in supply of care as a result of

a) change in organization of care;
b) change in availability of drugs including
 - introduction of a new drug;
 - change in reimbursement of medication;
 - change in availability of prescription versus over-the-counter drugs;
c) unknown change in supply of care.

Category 3: Changes in demand for care as a result of

a) media attention including advertising;
b) change in patients' perspective of disease;
c) change in patients' expectations of general practice;
d) unknown change in demand for care.

Category 4: Changes in incidence of disease as a result of

a) natural course of disease;
b) medical intervention;
c) change in lifestyle behaviour;
d) change in compilation of the patient population;
e) change in environment;
f) unknown cause that changes incidence of disease.

We discussed the causes listed in box 5.1 with three experts in the field of general practice registration networks (see acknowledgements) in order to interpret the findings of our study. The most plausible explanations are described below. Note that we did not study these explanations; they are merely hypotheses. We also compared our findings to the annual incidence rates of a similar general practice network in England and Wales (Weekly Returns Service).[17] This comparison is limited to ICPC-codes that could be mapped to corresponding ICD-9 related Read codes from the Weekly Returns Service.

The increasing 'concern about drug reaction (A13)' might be due to a change in the availability of over-the-counter drugs. In 1999, a large number of drugs that was previously available on prescription only became available over-the-counter, although chronic users still received a reimbursement on presentation of a prescription. Subsequent legislation (starting from January 1, 2004) limited the reimbursement for chronic users considerably.[18] The use of over-the-counter drugs shows a steady increase in 2002-2005 period.[19] The observed increase of 'concern about drug reaction (A13)' could thus be caused by patients who consulted their GP about possible adverse effects of over-the-counter drugs.

An explanation for the increase of 'change in faeces/bowel movements (D18)' might be a very successful, broad campaign on colon cancer in 2004 by a large non-profit organization. Among other things, the campaign explained that a change in faeces/bowel movements is one of the possible signs of colon cancer.[20;21]

We have no explanation for the observed decrease of 'chronic internal knee derangement (L97)'. In English and Welsh practices, the incidence of 'internal derangement of knee' (ICD-9 code 717) decreased also in the study period.[17]

The decrease of 'hypertrophy/chronic infection tonsils/adenoids (R90)' might be part of an ongoing decline that started decades ago. Another Dutch registry in general practice noticed a decrease in the incidence of this disease starting in the 1970s. This decline is possibly related to the simultaneous decreasing trend in tonsillectomy.[22] Furthermore, in 1993 the haemophilus influenza type b vaccine was added to the routine childhood immunization schedule, which could have caused a further decrease in this disease. In England and Wales, the incidence of 'chronic disease of tonsils and adenoids' (ICD-9 code 474) remained more or less the same between 2002-2005.[17]

We found a decrease in consultations for 'fear of being pregnant (W02)', especially among young women, a finding which could be partly explained by the increased use of contraceptives among young people. Results of a 2005 study on sexual health among young people show an increased use of contraceptives among sexually active school going youth compared to 1995.[23]

The decrease in consultations for sterilization among women (W13) is possibly related to the increase in childbearing age. Results of a study comparing different birth cohorts indicate that sterilizations are postponed (delay effect), whereas no catch-up effect is observed yet.[24] Furthermore, in 2004 new legislation excluded sterilization from the basic health insurance plan[25] and since then, only supplemental plans cover sterilization. This health policy measure may have attributed to a decrease in consultations for sterilization in the period after June 1, 2004.

The increase in 'symptoms/complaints prostate (Y06)' might be due to the fact that new drugs for benign prostate hyperplasia became available in the period under study. This medical condition causes urination problems, and between 2002 and 2005, the pharmaceutical industry launched a broad advertising campaign focusing on urinary problems. The availability of these new drugs accompanied by the advertising campaign may have resulted in an increase in consultations for both prostate and urination problems in general practice.

The increased 'fear of venereal disease (Y25)' is not limited to men. We found an increased 'fear of venereal disease (X23)' among women as well, but this health problem was not selected for further investigation (the raw incidence increased from 10.0 to 19.6 among young women (15 through 24 years of age) and from 2.7 to 6.5 among women aged 25 through 44 years of age). The increased fear is probably explained by a considerable rise in the number of sexually transmitted diseases (STDs) in the examined period. In particular, Chlamydia, gonorrhoea and Lues showed an increase among both men and women. This trend is attributed to the availability of effective therapy for AIDS (Highly Active Antiretroviral Therapy) which resulted in an increase in unprotected sexual behaviour.[26;27]

We found an increase in 'frequent/urgent (U02)' and 'other urination problems (U05)' for both men and women. The availability of, and campaign for, new drugs for benign prostate hyperplasia as well as the increase in STDs provide a partial explanation for the observed increase in urination problems since both urinary tract infections and sexually transmitted diseases may cause urination problems. Furthermore, these diseases share similar risk factors such as recent sexual activity.[28] The increase in urination problems among middle-aged and older women might be due to another, unknown cause of increasing urinary tract infections. Results of our study show an increase in 'urinary tract infections (U71)' among women of 25 years of age and older, but this health problem was not selected for further investigation (the raw incidence for all women increased from 69.0 in 2002 to 77.6 in 2005).

We could provide plausible explanations for most of our findings. These presumed causes, however, were not investigated in this study and no causal relationships were assessed. Therefore, these interpretations should be considered as hypotheses for further investigation.

Implications of findings

This study shows that data from EMRs can be a valuable source to monitor trends in the incidence of health problems in general practice over a relatively short time period. Extracting diagnoses from EMRs, provided that they are recorded accurately, is an efficient method to obtain a dataset that is large enough to pick up both usual and unusual changes over time. In this paper we provide an example of a strategy for processing, analysing and interpreting such data. This strategy is essentially applicable to all databases of routinely collected data in general practice. These new possibilities are relevant for future research, not only in the Netherlands, but in all countries with a high degree of Information Technology. Future research should aim to develop methods that disclose and use data from EMRs for research purposes, in particular methods for ensuring the quality of this data.

Results of this study regarding time trends in morbidity point at three important issues. First, attention should be paid to the increasing concern about reactions to drugs and the possible relation with the availability of over-the-counter drugs. Second, our findings indicate that advertising campaigns may have considerable impact on incidence rates in general practice. More research is needed to investigate the effect of advertising campaigns on the health care market. Third, the increasing incidence of urination problems and fear of venereal diseases is a cause for concern. Both health problems are probably related to an increase in the incidence of STDs and urinary tract infections.

SUPPLEMENTARY MATERIAL

The supplementary material provides further details on the selection procedures, the analyses, and an overview of all adjusted rates with 95% confidence limits.

ACKNOWLEDGEMENTS

The authors would like to thank the following people for their contributions to the interpretations of the findings of this study: **François G. Schellevis**, Netherlands Institute for Health Services Research, Utrecht, and EMGO Institute, Department of General Practice, VU University Medical Centre, Amsterdam, The Netherlands; **Eloy H. van de Lisdonk**, Department of General Practice, University Medical Centre Nijmegen, The Netherlands; and **Toine L.M. Lagro-Janssen**, Department of General Practice, University Medical Centre Nijmegen, The Netherlands. We also thank **Douglas M. Fleming**, Birmingham Research Unit of the Royal College of General Practitioners, Birmingham, England for his valuable assistance with the incidence rates of the Weekly Returns Service.

REFERENCES

1. Smith G, Hippisley-Cox J, Harcourt S, Heaps M, Painter M, Porter A, et al. Developing a national primary care-based early warning system for health protection - a surveillance tool for the future? Analysis of routinely collected data. J Public Health (Oxf). 2007;29:75-82.
2. WONCA Classification Committee. An international glossary for general/family practice. Fam Pract. 1995;12:341-69.
3. Okkes IM, Groen A, Oskam SK, Lamberts H. Advantages of long observation in episode-oriented electronic patient records in family practice. Methods Inf Med. 2001;40:229-35.
4. Biermans MCJ, De Bakker DH, Verheij RA, Gravestein JV, Van der Linden MW, De Vries Robbé PF. Development of a case-based system for grouping diagnoses in general practice. Int J Med Inform. 2008;77:431-39.
5. Biermans MCJ, Verheij RA, De Bakker DH, Zielhuis GA, De Vries Robbé PF. Estimating morbidity rates from electronic medical records in general practice. Evaluation of a grouping system. Methods Inf Med. 2008;47:98-106.
6. Biermans MCJ, Elbers GH, Verheij RA, Van der Veen WJ, Zielhuis GA, De Vries Robbé PF. External validation of EPICON: A grouping system for estimating morbidity rates using Electronic Medical Records. J Am Med Inform Assoc. 2008;15:770-775.
7. Bartholomeeusen S, Kim CY, Mertens R, Faes C, Buntinx F. The denominator in general practice, a new approach from the Intego database. Fam Pract. 2005;22:442-7.
8. Verheij RA, Jabaaij L, Abrahamse H, Van den Hoogen H, Braspenning J, Van Althuis T. Netherlands Information Network of General Practice. Facts and figures on Dutch GP care [website on the internet]. Available from: http://www.linh.nl.
9. Tacken MAJB. Quality of preventive performance in general practice: the use of routinely collected data [dissertation]. Radboud University Nijmegen, 2005.
10. Lamberts H, Woods M, editors. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.
11. Hingstman L, Kenens RJ. Cijfers uit de registratie van huisartsen (Figures from the registration of general practitioners). Utrecht: Nivel; 2005.
12. Statistics Netherlands [database on the internet]. Voorburg/Heerlen: Statistics Netherlands. [updated daily; cited 2008 Jan 14]. Available from: http://www.cbs.nl.
13. Westert GP, Jabaaij L, Schellevis FG, editors. Morbidity, performance and quality in primary care: Dutch general practice on stage. Oxon: Radcliffe; 2006.

14. Westert GP, Schellevis FG, De Bakker DH, Groenewegen PP, Bensing JM, Van der Zee J. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15:59-65.
15. Twisk JWR. Applied multilevel analysis. Cambridge: Cambridge University Press; 2006
16. Hedeker D, Gibbons RD. Longitudinal data analysis. New York: Wiley; 2006.
17. Royal College of General Practitioners. The Birmingham Research Unit [website on the internet]. [cited 2008 Oct 4]. Available from: http://www.rcgp.org.uk/clinical_and_research/bru.aspx.
18. College voor Zorgverzekeringen. Overzicht beleidsmaatregelen farmaceutische zorg 1995-2006 (Review policy measures pharmaceutical care 1995-2006) [document on the internet]. Diemen: CVZ; 2008 [cited 2008 Jan 14]. Available from: http://www.gipdatabank.nl.
19. Statistics Netherlands. Gezondheid en zorg in cijfers 2006 (Health and care in figures 2006). Voorburg/Heerlen: Statistics Netherlands; 2006.
20. Maag Lever Darm Stichting. Darmkanker.info (Coloncancer.info) [document on the internet]. Nieuwegein: MDLS; 2008 [cited 2008 Jan 14]. Available from: http://www.darmkanker.info.
21. Hill & Knowlton. Jaarverslag Maag Lever Darm Sichting 2004 (Annual report Stomach Liver Intestine Foundation 2004). Rotterdam: The happy horseman, 2004.
22. Van de Lisdonk EH, Van den Bosch WJHM, Lagro-Janssen ALM, editors. Ziekten in de huisartspraktijk (Diseases in General Practice). Maarssen: Elsevier gezondheidszorg; 2003.
23. De Graaf H, Meijer S, Poelman J, Vanwesenbeeck I. Seks onder je 25e. Sexuele gezondheid van jongeren in Nederland anno 2005 (Sex under 25. Young people's sexual health in the Netherlands of 2005). Delft: Eburon; 2005.
24. De Graaf A. Geboorteregeling in 2003 (Family planning in 2003). Bevolkingstrends. 2004:52;23-27.
25. Regeling ziekenvervoer Ziekenfondswet (Regulation patient transport Health Insurance Act). Staatscourant 16 April 2004.
26. Van de Laar MJW, Van Veen MG. Soa samengevat (STDs summarised) [document on the internet]. Bilthoven: RIVM; 2008 [updated 2008 Apr 17; cited 2008 Jan 14]. Available from: http://www.nationaalkompas.nl.
27. Van de Laar MJW. SOA nemen opnieuw toe (STDs increasing again). Infectieziekten bulletin. 2005;16:116-17.
28. Huppert JS, Biro F, Lan D, Mortensen JE, Reed J, Slap GB. Urinary symptoms in adolescent females: STI or UTI? J Adolesc Health. 2007;40:418-24.

SUPPLEMENTARY MATERIAL CHAPTER 5

INTRODUCTION

The following sections provide details on the article 'Striking trends in the incidence of health problems in the Netherlands (2002-05). Findings from a new strategy for surveillance in general practice'. In particular, we elaborate on the selection of general practices, the selection of 15 health problems for further investigation, and the multilevel analyses. In addition, we present a tabular overview of the adjusted rates with 95% confidence limits.

SELECTION OF GENERAL PRACTICES

In the study, we included only general practices that met certain criteria for accuracy and completeness. This selection procedure is based on the following criteria:

(a) at least 60% of the ICPC-codes assigned to consultations had to be filled out, and
(b) consultation, and
(c) prescription records had to be recorded adequately throughout the year (at least 10% of all consultation and 10% of all prescription records in each quarter), and
(d) at least 75% of the field 'new' or 'ongoing' in contact-oriented EMRs had to be filled out. Within LINH, this field is added to contact-oriented EMRs. General practitioners have to characterize each consultation diagnosis either as 'new', which refers to a newly presented/recurrent health problem, or 'ongoing', which refers to a continuing health problem.

SELECTION OF 15 HEALTH PROBLEMS

Out of the raw incidence rates for all ICPC codes (N=683), we selected 15 obviously changing health problems for further investigation. Figure 5.2 shows a flowchart of this selection process. We excluded rare health problems (prevalence ≤ 0.5 per 1000 patients in 2002-2005) and rest group health problems (ICPC-codes 29 and 99). We included all health problems that displayed a substantial increase (or decrease) between 2002 and 2005 (absolute difference ≥| 0.3 |and relative difference ≥ | 30%|). Subsequently we checked for outliers. For three health problems, the changes over time could mainly be attributed to one or two practices ('other symptoms/complaints multiple/unspecified muscles (L19)', 'meningitis/encephalitis (N71)', and 'painful urination (U01)'). These health problems were excluded, leaving a selection of 15 health problems.

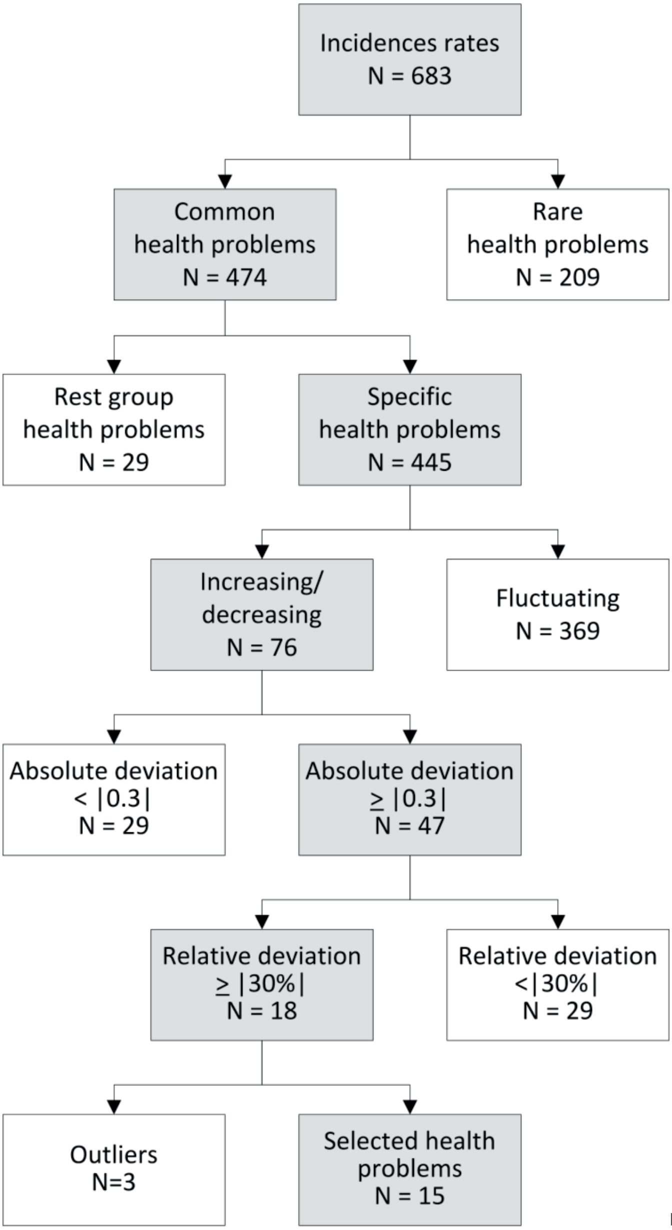


Fig. 5.2 Selection of health problems

MULTILEVEL ANALYSES

To analyse the data, we used multilevel multivariate repeated measurements models for count data (poisson error distribution allowing for overdispersion, with a log link function, MLwiN specifics estimation was penalized quasi-likelihood with first order). For every year we specified the same equation, all four equations were estimated simultaneously. We did not estimate the covariances between the different years, neither at the level of the practice nor at the level of the individual patients. This was due to sparseness of the data: less than 0.5% of the patients have the specific episode and almost none have new specific episodes in two or more years.

The equation for one year is as follows:

$$y_{ij} = \beta_0(\text{intercept}) + \beta_1(\text{sex})_{ij} + \beta_2(\text{age})_{ij} + \beta_3(\text{age}^{\wedge 2})_{ij} + \beta_4(\text{age}^{\wedge 3})_{ij} + \beta_5(\text{length})_{ij} + \mu_j + \varepsilon_{ij}$$

i patient 1..... n

j practice 1....N

μ_j between practice variance (normal distribution)

ε_{ij} between patient variance within practice (extra poisson distribution)

The models are restricted to the ages upon which a health problem occurs (we included only those five-year ages for which at least 5 new episodes of a health problem occurred in the period 2002-2005. Table 5.2 shows the age range used in the analyses as well as the raw prevalence, the mean age, and the male/female proportion for each of the examined health problems.

Table 5.2 Overview of selected health problems

Health problems (ICPC code)[a]	Prev. (raw) 2005	Mean age 02-05	Age range	Males (%) 02-05
Increasing based on raw incidence rates				
Concern about drug reaction (A13)	5.7	54.9	0-99	31.5
Change in feces/bowel movements (D18)	1.7	49.9	0-94	44.6
Hypertension with involv. target organs (K87)	9.4	65.7	30-94	55.8
Frequent/urgent urination (U02)	4.3	47.5	1-99	44.9
Other urination problems (U05)	2.6	53.7	0-99	61.3
Family planning/other (W14)	11.5	30.2	10-59	0.0
Symptoms/complaints prostate (Y06)	7.3	62.9	25-94	100.0
Fear of venereal disease (Y25)	3.9	29.7	15-64	100.0
Decreasing based on raw incidence rates				
Sprains & strains of other joints (L79)	2.1	30.4	0-94	49.9
Chronic internal knee derangement (L97)	3.6	38.8	5-89	48.5
Other head injury without skull fracture (N80)	0.8	26.2	0-94	52.3
Hypertro./chronic infect. tonsils/adenoids (R90)	2.0	8.1	0-64	50.5
Diaper rash (S89)	1.1	2.0	0-9	36.3
Fear of being pregnant (W02)	1.8	28.2	10-54	0.0
Family planning/sterilization (W13)	0.9	36.4	20-49	0.0

a: ICPC, International Classification of Primary Care

The age and sex variables were centered, i.e. the adjusted model estimates the incidence for a patient with an average age and an average male/female proportion. This centering differed between the dependent variables, for an overview of these values see table 5.2 (mean age and males %). The variable length corrects for the length of recording of a patient within a practice within one year.

For testing the trends over the years we used linear contrast tests for the fixed parameters, the intercept, sex and the three age variables. For testing the sex trends a model was fitted in which the intercept was removed and a separate yearly average was estimated for male and female, both 0,1 coded. The data were analysed using MLwiN 2.02.

OVERVIEW OF THE ADJUSTED RATES

Table 5.3 provides an overview of the adjusted rates of the 15 selected health problems with confidence limits.

Table 5.3 Adjusted incidence rates (corrected for clustering and age) per 1000 per year with 95% Confidence Intervals

Health problems (ICPC code)[a]	Men	Women
Concern about drug reaction (A13)		
2002	2.0 (1.5 - 2.7)	4.1 (3.1 - 5.5)
2003	2.5 (1.8 - 3.4)	4.9 (3.6 - 6.7)
2004	2.7 (1.8 - 4.2)	5.6 (3.7 - 8.4)
2005	4.5 (3.2 - 6.5)	7.9 (5.6 - 11.2)
Change in feces/bowel movements (D18)		
2002	0.7 (0.5 - 0.9)	0.9 (0.7 - 1.1)
2003	0.8 (0.6 - 1.1)	0.9 (0.7 - 1.1)
2004	0.9 (0.6 - 1.2)	1.0 (0.7 - 1.3)
2005	1.3 (1.0 -1.7)	1.6 (1.2 - 2.1)
Hypertension with involvement target organs (K87)		
2002	4.1 (3.0 - 5.6)	2.6 (1.9 - 3.6)
2003	4.6 (3.3 - 6.4)	3.6 (2.5 - 5.0)
2004	6.5 (4.8 - 8.9)	3.0 (2.1 - 4.2)
2005	5.1 (3.5 - 7.4)	4.2 (2.9 - 6.1)
Sprains and strains of other joints (L79)		
2002	3.1 (2.4 - 4.0)	3.1 (2.4 - 4.0)
2003	2.9 (2.2 - 3.8)	2.9 (2.2 - 3.8)
2004	2.5 (1.8 - 3.6)	2.8 (2.0 - 4.0)
2005	2.4 (1.6 - 3.4)	2.2 (1.5 - 3.1)
Chronic internal knee derangement (L97)		
2002	5.3 (4.3 - 6.5)	5.5 (4.5 - 6.7)
2003	4.3 (3.4 - 5.3)	4.4 (3.6 - 5.5)
2004	4.3 (3.1 - 5.9)	4.5 (3.3 - 6.2)
2005	2.9 (2.1 - 4.0)	3.7 (2.7 - 5.1)
Other head injury without skull fracture (N80)		
2002	0.7 (0.5 - 0.9)	0.6 (0.5 - 0.8)
2003	0.5 (0.3 - 0.6)	0.4 (0.3 - 0.6)
2004	0.6 (0.4 - 0.9)	0.4 (0.3 - 0.7)
2005	0.5 (0.3 - 0.7)	0.5 (0.3 - 0.7)
Hypertrophy/chronic infection tonsils/adenoids (R90)		
2002	6.7 (5.1 - 8.9)	5.8 (4.4 - 7.6)
2003	5.7 (4.4 - 7.4)	6.7 (5.2 - 8.7)
2004	5.0 (3.8 - 6.6)	4.9 (3.7 - 6.4)
2005	4.4 (2.9 - 6.7)	5.2 (3.4 - 7.9)

Table 5.3 Continued

Health problems (ICPC code)a	Men	Women
Diaper rash (S89)		
2002	7.8 (5.4 - 11.2)	15.7 (11.5 - 21.5)
2003	7.2 (4.9 - 10.5)	12.9 (9.2 - 18.0)
2004	5.5 (3.6 - 8.3)	9.5 (6.5 - 14.0)
2005	6.6 (2.5 - 17.3)	12.6 (5.5 - 28.5)
Frequent/urgent urination (U02)		
2002	1.9 (1.5 - 2.3)	2.0 (1.6 - 2.4)
2003	2.2 (1.7 - 2.7)	2.4 (1.9 - 3.0)
2004	2.4 (1.9 - 3.2)	2.8 (2.1 - 3.6)
2005	2.6 (2.1 - 3.3)	3.2 (2.5 - 3.9)
Other urination problems (U05)		
2002	1.7 (1.3 - 2.2)	0.9 (0.6 - 1.1)
2003	1.7 (1.4 - 2.2)	1.0 (0.7 - 1.3)
2004	2.6 (2.0 - 3.5)	1.4 (1.0 - 1.9)
2005	2.9 (2.3 - 3.9)	1.6 (1.2 - 2.2)
Fear of being pregnant (W02)		
2002	n/a	2.5 (1.9 - 3.2)
2003	n/a	2.0 (1.6 - 2.7)
2004	n/a	2.0 (1.4 - 2.8)
2005	n/a	1.5 (1.1 - 2.0)
Family planning/sterilization (W13)		
2002	n/a	2.7 (2.2 - 3.4)
2003	n/a	2.8 (2.2 - 3.6)
2004	n/a	1.9 (1.4 - 2.7)
2005	n/a	1.1 (0.7 - 1.7)
Family planning other (W14)		
2002	n/a	4.9 (3.7 - 6.4)
2003	n/a	6.4 (4.9 - 8.4)
2004	n/a	6.1 (4.9 - 7.5)
2005	n/a	6.6 (5.0 - 8.7)
Symptoms/complaints prostate (Y06)		
2002	4.3 (3.5 - 5.3)	n/a
2003	5.0 (4.0 - 6.3)	n/a
2004	6.6 (5.2 - 8.3)	n/a
2005	7.4 (5.7 - 9.5)	n/a
Fear of venereal disease (Y25)		
2002	1.9 (1.4 - 2.6)	n/a
2003	2.4 (1.8 - 3.2)	n/a
2004	3.3 (2.4 - 4.4)	n/a
2005	4.2 (3.1 - 5.6)	n/a

a: International Classification of Primary Care

Chapter 6

Decreasing incidence of adenotonsillar problems in Dutch general practice: Real or artefact?

Marion CJ Biermans
Ellen HM Theuns-Lamers
Peter Spreeuwenberg
Robert A Verheij
Johannes C van der Wouden
Pieter F de Vries Robbé
Gerhard A Zielhuis

Submitted for publication

ABSTRACT

Background: The incidence of hypertrophy and recurrent infections of tonsils/adenoid in the Netherlands appears to be decreasing: is this a real decrease in the incidence of disease or an artefact?
Aim: To investigate possible causes of the decreasing incidence of adenotonsillar problems among Dutch children.
Design: Observational study.
Setting: A nationally representative general practice database.
Method: Incidence rates were calculated over 2002-2005 among children aged 0-14 years. Multilevel Poisson regression analyses were used to examine the following possible causes of changing incidence rates: change in recording (i.e. more substitution codes), change in the demand for care (i.e. fewer visits to the general practitioner), and change in the supply of care (i.e. fewer antibiotic prescriptions and referrals). Indications for a real change in the incidence of disease were examined by calculating incidence rates of other clinical manifestations of microbial pathogens that may cause adenotonsillar problems.
Results: The incidence rate decreased significantly (p=0.02) from 3.0 to 1.3 per 1000 children per year. Correcting for demand for and supply of care led to a smaller decline in yearly incidence, from 2.9 to 1.7 per 1000 children per year (p=0.11). No clearly similar trend was found in other clinical manifestations of viruses and bacteria that may cause adenotonsillar problems.
Conclusion: Part of the declining trend can be explained by a change in the demand for and supply of care, but no apparent causal clue emerged for the residual declining trend in the incidence of disease.

INTRODUCTION

Adenotonsillar problems, notably obstructive hypertrophy and infection, are very common in early childhood and an important reason for parents to visit the general practitioner (GP) with their child. Recurrent infections of the tonsils or adenoid can be a source of both upper and lower respiratory tract infections. Hypertrophy can lead to sleep-disordered breathing, eating disorders, and even growth problems.[1-5]

The relation between hypertrophy and recurrent infections of adenotonsillar tissue is unclear. Although hypertrophy is associated with recurrent infections of adenotonsillar tissue, many children present hypertrophy in the absence of infection.[1;6-8] Some suggest that tonsils and adenoid may only appear to be large due to their prominence or relative size in the throat during childhood.[7] Recent MRI studies, however, indicate that tonsils and adenoid grow proportionally to the skeletal structures during normal child development.[9;10]

Treatment options for adenotonsillar problems in general practice include watchful waiting, symptomatic medication, antibiotics, or referral to an otorhinolaryngologist. Recurrent tonsillitis, as well as obstructive sleep apnoea due to adenotonsillar hypertrophy, are important indications for (adeno)tonsillectomy, a common surgical procedure onchildren. There is an ongoing debate, however, about the proper indications for (adeno)tonsillectomy resulting in widely varying surgical rates.[11]

The few studies that examined the incidence of this health problem in general practice indicate, at least for the Netherlands, that the incidence of hypertrophy and recurrent infections of tonsils/adenoid among the total population is decreasing.[12,13] Considering the differing opinions on diagnostics and treatment consequences for this health problem, the question arises whether there is a 'real' decrease in the incidence of this disease, or one originated by 'artefacts'.

In an earlier paper, we distinguished between four categories of possible causes of changing incidence rates in general practice: changes in recording, changes in the demand for care, changes in the supply of care, and 'real' changes in the incidence of a disease.[13] Based on this categorization, we composed the following hypotheses.

Changes in recording

A shift in recording might explain the decrease in hypertrophy and recurrent infections of tonsils/adenoid. GPs might have replaced the diagnostic code for this particular health problem by other codes.

Changes in the demand for care
The overall consultation frequency of Dutch children aged 0-14 years decreased during 1987 to 2001.[14] This trend may have continued and may explain the decrease in adenotonsillar problems.
Changes in the supply of care
Dutch GPs usually work according to guidelines. In 1999 the guideline 'Acute sore throat' was revised, advocating restraint in prescribing antibiotics. Recommendations for referring to an otorhinolaryngologist also changed, now requiring four (instead of three) severe episodes of tonsillitis per year.[15] These policy changes might have reduced the number of consultations for this specific health problem. The less children are treated with antibiotics or (adeno)tonsillectomy, the less parents are inclined to visit the GP with their child for these treatments.
Changes in the incidence of a disease
Viruses and bacteria are important pathogens in the aetiology of infections and hypertrophy of the tonsils and adenoid. Viruses dominate in pre-school children.[16] Viral pathogens are: adenovirus, parainfluenza virus, respiratory syncytial virus (RSV), rhinovirus, Epstein-Barr virus (EBV), and herpes simplex virus (HSV). Bacterial species that have been isolated are: Haemophilus influenzae, Streptococcus pyogenes, Streptococcus pneumonia, and Staphylococcus aureus.[16-28] A decrease in the incidence of these pathogens could explain a decrease in adenotonsillar problems.

The present study examines (causes of) the trend in the incidence of hypertrophy and recurrent infections of tonsils/adenoid among children aged 0-14 years in the Netherlands, based on the available evidence in a national GPs' database.

METHOD

Dataset

We used data from the electronic medical records (EMRs) of Dutch general practices participating in the Netherlands Information Network of General Practice (LINH) during 2002-2005.[29,30] The GPs within this network code consultations, prescriptions, and referrals on a routine basis using the International Classification of Primary Care (ICPC).[31] All ICPC codes were grouped into episodes, either by EPICON, a new record linkage method (for contact-oriented EMRs) or by GPs (for episode-oriented EMRs). [32,33] Only practices that met predefined criteria for accuracy and completeness were included in the analyses; this resulted in 69 practices in 2002, 66 in 2003, 43 in 2004 and 42 in 2005. The included practices are considered representative for all Dutch practices regarding urbanization, practice type (i.e. single-handed or group practice), and region.[13]

Study population

A total of 80836 patients, boys and girls aged 0-14 year, was available for this analysis. Age was measured on the first of July in the year concerned. The population is representative for the total Dutch population regarding age and sex.[34]

Outcome

The main outcome variable was the yearly incidence of hypertrophy and recurrent infections of tonsils/adenoid (ICPC code R90). Indications for a 'real' change in the incidence of disease were examined by estimating incidence rates of clusters of other diseases that share the same pathogens with adenotonsillar problems. These clusters were composed as follows:

a) Based on the literature, we listed the major microbial pathogens of hypertrophy and infections of the tonsils and adenoid.[16-28]

b) We then looked for other frequent clinical manifestations of these pathogens and selected the corresponding ICPC codes. Clinical manifestations involving a general ICPC code, such as coughing (R05), were excluded.

c) For each pathogen a cluster of ICPC codes was composed (see Appendix I).

Explanatory variables

Possible changes in recording were examined by composing a cluster of possible substitution codes for ICPC code R90 (see Appendix II). The overall consultation frequency was used as a measure for the demand for care, whereas changes in the supply of care were measured by referrals to an otorhinolaryngologist and prescriptions of antibiotics.

Table 6.1 gives a detailed description of these explanatory variables.

Table 6.1. Explanatory variables

Variable	Description	Range
Raw incidence of substitution cluster	Total number of new episodes of substitution codes in one year per patient/total mid-year population	288.12 - 364.78 (per 1000 children per year)
Consultation frequency	Total number of consultations[a] per practice per year/ practice mid-year population	0.54 - 3.68 (per child per practice per year)
Referral frequency	Total number of referrals to an otorhinolaryngologist for a new episode of 'hypertrophy/recurrent infections of tonsils/ adenoid' per practice per year/practice mid-year population	0.00 - 21.05 (per 1000 children per practice per year)
Antibiotics frequency	Total number of antibiotic prescriptions for a new episode of 'hypertrophy/recurrent infections of tonsils/ adenoid' per practice per year/practice mid-year population	0.00 - 75.79 (per 1000 children per practice per year)

a The number of consultations was measured as the number of valid ICPC-codes entered during consultations, excluding ICPC-codes derived from referrals and prescriptions.

Incidence rates

Raw and adjusted incidence rates were calculated. The raw rates were computed as the total number of new episodes of a (cluster of) health problem(s) divided by the mid-year population (i.e. the average of the population aged 0-14 years at the beginning and the end of a year). The episodes were weighted for the recording period of each practice within each year to account for holidays, etc.

The adjusted incidence rates were corrected for clustering and for age, sex, and recording period. For this, we designed a basic model that includes the four years and corrects for variance due to clustering within levels (which is necessary because repeated observations of a health problem are clustered within patients, who are clustered within practices) and for age, sex and the length of recording of a practice within one year. The age and sex variables were centred at 8.1 years and 50.5% males (the mean age, and the average male/female proportion of this health problem in the population),[13] i.e. the model estimates the incidence for an eight year old patient.

In order to explain the trend over time for hypertrophy and recurrent infections of tonsils/adenoid, the explanatory variables were then entered stepwise into the basic model. We only included variables that a) increased or decreased over time in the expected direction based on the raw analyses, and b) were significantly associated with hypertrophy and recurrent infections of tonsils/adenoid (according to the 95% confidence intervals (CIs) of the rate ratio (RR)). For each of these models, we tested for a linear trend over time. In the final model, the consultation, referral, and frequency of antibiotic prescriptions were added (in this order) to the basic model. Multilevel Poisson regression models (MLwiN 2.02) were used to estimate the adjusted incidence rates.

RESULTS

Overall trend

Figure 6.1 shows the annual adjusted incidence rates of hypertrophy and recurrent infections of tonsils/adenoid among children aged 0-14 years. Between 2002 and 2005, this rate decreased from 2.96 (95% CI 2.16-4.07) to 1.27 (95% CI 0.67-2.41) per 1000 children per year. The decrease is significant for the total group of children ($p=0.0171$), and for boys ($p=0.021$) and girls ($p=0.022$) separately.

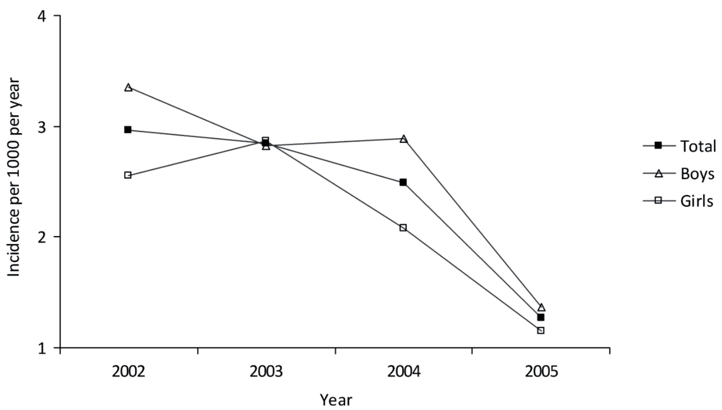


Fig. 6.1 Adjusted incidence rates (basic model) of hypertrophy and recurrent infections of tonsils/adenoid among children aged 0-14 years.

Raw rates

The raw rates of consultations, referrals, and antibiotic prescribing declined over 2002-2005. For instance, the total annual consultation frequency decreased from 1.98 in 2002 to 1.79 in 2005. Instead of the hypothesised increase of substitution codes, however, we found a decrease in the incidence rate of the cluster of substitution codes. Even when we computed a cluster of only upper respiratory codes, we noticed a decrease (Figure 6.2). Based on these raw rates, we could reject the hypothesis that the decrease was caused by a shift in recording. Therefore, this variable was not included in the multi-level Poisson regression analyses.

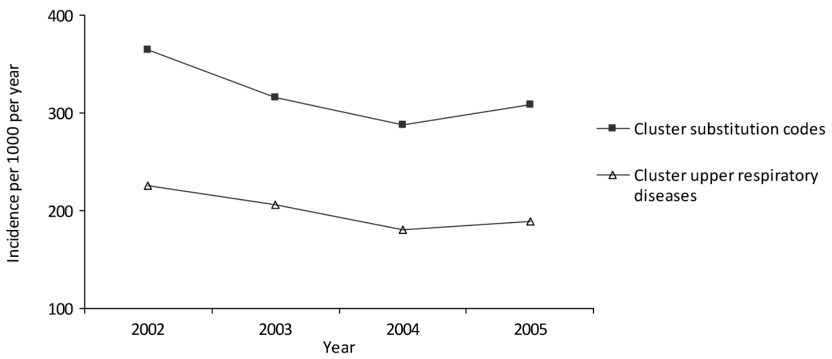


Fig. 6.2 Raw incidence rates of clusters of substitution codes and upper respiratory diseases among children aged 0-14 years.

Adjusted rates

Table 6.2 shows the RRs for the other explanatory variables. These RRs can be interpreted in such a way that a mean decrease of one consultation per child per practice per year (for instance, when the consultations per child decrease from three to two per year) is associated with a 23% decrease in risk of new episodes of hypertrophy and recurrent infections of tonsils/adenoid over the examined four-year period. All three variables are significantly associated with hypertrophy and recurrent infections of tonsils/adenoid (i.e. the 95% CIs exclude unity).

Table 6.2. Associations between the explanatory variables and new episodes of hypertrophy and recurrent infections of tonsils/adenoid

Variable	**RR[a] (95% CI)**
Consultation frequency	1.23 (1.02-1.48)
Referral frequency	1.09 (1.07-1.12)
Antibiotics frequency	1.03 (1.02-1.04)

[a] Based on the final model

Thus, these three variables have the potential to explain the decrease in adenotonsillar problems, because a) they declined, and b) they are associated with these problems. Figure 6.3 shows the impact of each of the variables on the adjusted incidence rate of hypertrophy and recurrent infections of tonsils/adenoid. Each variable explains a small part of the decrease of this incidence rate. The linear trend over time was no longer significant after referral (p=0.075) and antibiotics frequency (p=0.105) were entered into the model.

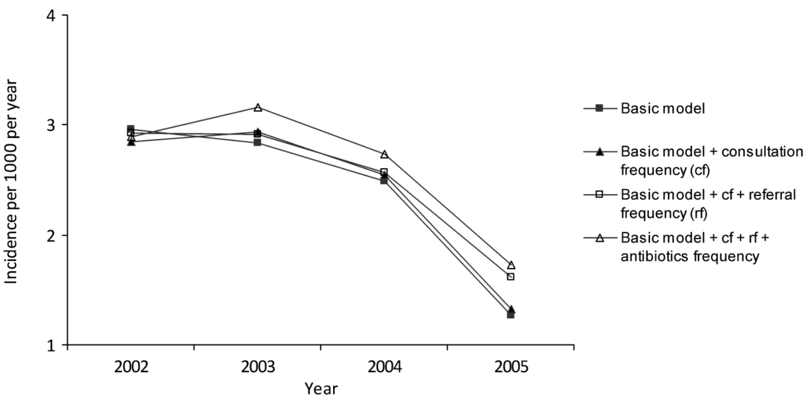


Fig. 6.3 Adjusted incidence rates (various models) of hypertrophy and recurrent infections of tonsils/adenoid among children aged 0-14 years.

Clinical manifestations of pathogens found in hypertrophy/recurrent tonsillitis

Figure 6.4 shows the clusters of other clinical manifestations related to the same microbial pathogens held responsible for adenotonsillar problems, which show no linear decrease over time. The shape of the trend over time corresponds to the trend observed in the cluster of respiratory diseases (Figure 6.2) and to the decline in adenotonsillar problems over 2002-2004, whereas the rise of clinical manifestations in 2005 is opposed to the decrease of adenotonsillar problems in 2005 (Figure 6.3).

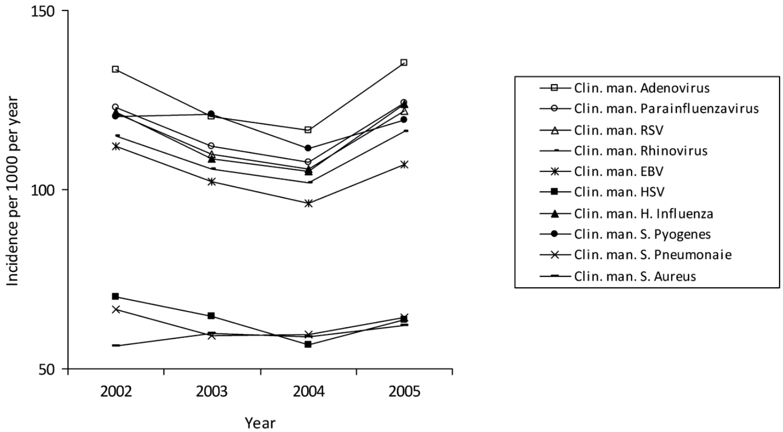


Fig. 6.4 Adjusted incidence rates (basic model) among children aged 0-14 years of clusters of clinical manifestations of pathogens found in hypertrophy and recurrent infections of tonsils/adenoid.

DISCUSSION

Summary of main findings

The incidence of recurrent infections and hypertrophy of tonsils/adenoid among Dutch children decreased over the years 2002-2005. Part of the decline can be explained by changes in the demand for and supply of care during this period, whereas no indications were found for a change in recording. We found no clear indications for a 'real' change in the incidence of disease.

Strengths and limitations of the study

This study illustrates the potential use of data derived from EMRs in general practice. A strength of the study is that a large network of general practices, and therefore a large number of patients, were included in the analyses. The study population was representative for the Dutch population regarding age and sex. Furthermore, the database enabled to study the most obvious hypotheses for the observed decrease in adenotonsillar problems.

The study was limited, however, to the available evidence in the database. Hence, the decrease might be explained by other factors that were not measured in this study, such as changes in day-care attendance. Furthermore, the study was limited to a four-year period, whereas changes that occurred before this time frame may have caused the decrease. In particular, the large decrease in (adeno)tonsillectomy rates during the 1970s and 1980s in the Netherlands[11] may have caused a general decrease in attention for adenotonsillar problems among parents, GPs and otorhinolaryngologists. This historical change in supply may still have a lingering effect on the current incidence rates.

Furthermore, we did not measure the change in actual pathogens that cause the disease. Instead, we used a proxy consisting of clusters of clinical manifestations of pathogens found in hypertrophy and infections of tonsils/adenoid, which has several disadvantages. First, we cannot be certain that the identified viruses and bacteria actually cause adenotonsillar problems, because the presence of an organism in a patient's throat and its culture from a swab does not mean that it is pathogenic.[5] Second, other microbial pathogens may also cause the clinical manifestations that were used in this study. The incidence rates of these clusters of clinical manifestations should therefore be interpreted cautiously; they only give an indication and cannot provide any evidence of a change in the incidence of major pathogens held responsible for adenotonsillar problems.

Finally, not all referrals and prescriptions were linked to an ICPC code by the GPs. Each year, ICPC codes were missing for 3-5% of the referrals to otorhinolaryngologists and for 7-10% of the antibiotic prescriptions. Therefore, we might have underestimated the referral and antibiotics frequency, but we have no reason to believe that higher estimates would have changed our conclusions.

Comparison with existing literature

The findings of this study raise the question whether the decrease of adenotonsillar problems is related to an overall decline in upper respiratory tract diseases. Both adenotonsillar problems and other upper respiratory tract diseases declined over 2002-2004, whereas opposite trends were observed over 2005. Several studies report that the incidence of various respiratory diseases, including acute tonsillitis, declined over the last decade.[35-38] Some suggest that this decline is caused by a reduced inclination of patients to present respiratory illness to their GP[36], whereas others suggest a 'real' decrease in the incidence of respiratory tract infections.[38] The results of our study appear to be in favour of the last hypothesis, because the decrease in the overall consultation frequency had only little impact on the observed decline in adenotonsillar problems (Figure 6.3). The basic model, however, controls for clustering within practices, which may already reduce some of the variance caused by differences in consultation frequency. Hence, the actual impact of consultation frequency might be somewhat larger. Nevertheless, our findings demonstrate that consultation frequency provides no conclusive explanation for the decline in adenotonsillar problems.

In the present study, we observed a decline in both the antibiotics and the referral frequency for hypertrophy and recurrent infections of tonsils/adenoid among children over the years 2002-2005. This finding corresponds with declining antibiotic prescribing trends for acute tonsillitis, and with decreasing referral rates for hypertrophy and recurrent infections of tonsils/adenoid among Dutch children during 1987 to 2001.[35;39;40]

Implications for future research

We observed a substantial decline in adenotonsillar problems among children over 2002-2005, partly explained by factors that are not causally linked to the aetiology of the disease itself. For the residual decline, no causal clue has yet emerged. Future research should focus on the question whether the observed decrease is part of an overall decline in upper respiratory tract infections resulting from a changing pattern in the occurrence of major pathogens in the population.

REFERENCES

1. Kara CO, Ergin H, Koçak G, Kiliç I, Yurdakul M. Prevalence of tonsillar hypertrophy and associated oropharyngeal symptoms in primary school children in Denizli, Turkey. Int J Pediatr Otorhinolaryngol. 2002;66:175-179.
2. Kvaerner KJ, Tambs K, Harris JR, Mair IWS, Magnus P. Otitis media: relationship to tonsillitis, sinusitis and atopic diseases. Int J Pediatr Otorhinolaryngol. 1996;35:127-141.
3. Bonuck K, Parikh S, Bassila M. Growth failure and sleep disordered breathing: A review of literature. Int JPediatr Otorhinolaryngol. 2006;70:769-778.
4. Tal A, Bar A, Leiberman A, Tarasiuka A. Sleep characteristics following adenotonsillectomy in children with obstructive sleep apnea syndrome. Chest. 2003;124:948-953.
5. Cowan DL, Hibbert J. Acute and chronic infection of the pharynx and tonsils. In: Kerr AG, Hibbert J, editors. Scott-Brown's Otolaryngology. Laryngology and head and neck surgery. 6th ed. Oxford: Butterworth-Heinemann; 1997. 5/4/1-5/4/24.
6. Potsic, WP. Assessment and treatment of adenotonsillair hypertrophy in children. Am J Otolaryngol. 1992;13:259-264.
7. Cable HR, Batch AG, Stevens DJ. The relevance of physical signs in recurrent tonsillitis in children. A prospective study. J Laryngol Otol. 1986;100:1047-1051.
8. Barr GS, Crombie IK. Comparison of size of tonsils in children with recurrent tonsillitis and in controls. BMJ. 1989;298:804.
9. Arens R, McDonough JM, Corbin AM, Hernandez ME, Maislin G, Schwab RJ, et al. Linear dimensions of the upper airway structure during development. Am J Respir Crit Care Med. 2002;165:117-122.
10. Vogler RC, Wippold FJ, Pilgram TK. Age-specific size of the normal adenoid pad on magnetic resonance imaging. Clin Otolaryngol. 2000; 25:392-395.
11. Van den Akker EH, Hoes AW, Burton MJ, Schilder AGM. Large international differences in (adeno)tonsillectomy rates. Clin Otolaryngol. 2004;29:161-164.
12. Van de Lisdonk EH, van den Bosch WJHM, Lagro-Janssen ALM, editors. Ziekten in de huisartsenpraktijk (Diseases in General Practice). Maarssen: Elsevier gezondheidszorg; 2003.
13. Biermans MCJ, Spreeuwenberg P, Verheij RA, De Bakker DH, De Vries Robbé, Zielhuis GA. Striking trends in de incidence of health problems in the Netherlands (2002-05). Findings from a new strategy for surveillance in general practice. Submitted.
14. Van den Berg MJ, Kolthof ED, De Bakker DH, Van der Zee J. De werkbelasting van huisartsen (The workload of general practitioners). Utrecht: Nivel; 2004.

15. Dagnelie CF, Zwart S, Balder FA, Romeijnders ACM, Geijer RMM. NHG Standard 'Acute sore throat' from the Dutch College of General Practitioners, first revision, 1999 [document on the internet]. [cited 2008 April 23]. Availablefrom: http://nhg.artsennet.nl/upload/104/guidelines2/E11.htm.
16. Putto A. Febrile exudative tonsillitis: Viral or streptococcal? Pediatrics. 1987;80:6-12.
17. Tosca MA, Riccio AM, Marseglia GL, Caligo G, Pallestrini E, Ameli F, et al. Nasal endoscopy in asthmatic children: assessment of rhinosinusitis and adenoiditis incidence, correlations with cytology and microbiology. Clin Exp Allergy. 2001;31:609-615.
18. Bisno AL. Acute pharyngitis. N Engl J Med. 2001;344:205-211.
19. Suvilehto J, Roivainen M, Seppänen M,Meri S, Hovi T, Carpén O, et al. Rhinovirus/enterovirus RNA in tonsillar tissue of children with tonsillar disease. J Clin Virol. 2006;35:292-297.
20. Gaffney RJ, Timon CI, Freeman DF, Walsh MA, Cafferkey MT. Bacteriology of tonsil and adenoid and sampling techniques of adenoidal bacteriology. Respir Med. 1993;87:303-308.
21. DeDio RM, Tom LWC, McGowan KLM, Wetmore RF, Handler SD, Potsic WP. Microbiology of the tonsils and adenoids in a pediatric population. Arch Otolaryngol Head Neck Surg. 1988;114: 763-765.
22. Van Staaij BK, Van den Akker EH, De Haas van Dorsser EHM, Fleer A, Hoes AW, Schilder AGM. Does the tonsillar surface flora differ in children with and without tonsillar disease? Acta Otolaryngol. 2003;123:873-878.
23. Endo LH, Ferreira D, Montenegro MCS, Pinto GA, Altemani A, BortoletoJr AE, et al. Detection of Epstein-Barr virus in tonsillar tissue of children and the relationship with recurrent tonsillitis. Int J Pediatr Otorhinolaryngol. 2001;58:9-15.
24. Brook I. Aerobic and anaerobic bacteriology of adenoids in children: a comparison between patients with chronic adenotonsillitis and adenoid hypertrophy. Laryngoscope. 1981;91:377-382.
25. Gaffney RJ, Cafferkey MT. Bacteriology of normal and diseased tonsils assessed by fine needle aspiration: Haemophilus influenzae and the pathogenesis of recurrent acute tonsillitis. Clin Otolaryngol. 1998;23:181-185.
26. Brook I, Yocum P, Foote PA. Changes in the core tonsillar bacteriology of recurrent tonsillitis: 1977-1993. Clin Infect Dis. 1995;21:171-176.
27. JeongJH, Lee DW, Ryu RA, Lee YS, Lee SH, Kang JO, et al. Bacteriologic comparison of tonsil core in recurrent tonsillitis and tonsillar hypertrophy. Laryngoscope. 2007;117:1-6
28. Stjernquist-Desatnik A, Prellner K, Schalén C. High recovery of Haemophilus influenzae and group A streptococci in recurrent tonsillar infection or hypertrophy as compared with normal tonsils. J Laryngol Otol. 1991;105:439-441.

29. Verheij RA, Jabaaij L, Abrahamse H, Van den Hoogen H, Braspenning J, Van Althuis T. Netherlands Information Network of General Practice. Facts and figures on Dutch GP care [website on the internet]. [cited 2008 April 23]. Available from: http://www.linh.nl.
30. Tacken MAJB. Quality of preventive performance in general practice: the use of routinely collected data [dissertation]. Radboud University Nijmegen; 2005.
31. Lamberts H, Woods M, editors. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.
32. Biermans MCJ, De Bakker DH, Verheij RA, Gravestein JV, Van der Linden MW, De Vries Robbé PF. Development of a case-based system for grouping diagnoses in general practice. Int J Med Inform. 2008;77:431-439.
33. Biermans MCJ, Verheij RA, De Bakker DH, Zielhuis GA, De Vries Robbé PF. Estimating morbidity rates from electronic medical records in general practice. Evaluation of a grouping system. Methods Inf Med. 2008;47:98-106.
34. Statistics Netherlands [database on the internet]. Voorburg/Heerlen: Statistics Netherlands. [updated daily; cited 2008 March 27]. Available from: http://www.cbs.nl.
35. Diepenhorst HIJ, Otters HBM, Van Suijlekom-Smit LWA, Schellevis FG, Van der Wouden JC. Tonsillitis acuta bij kinderen in de huisartsenpraktijk: verandering van incidentie en beleid? (Acute tonsillitis in children in the GP's practice: changes in incidence and policy?). Huisarts Wet. 2004;47:399-404.
36. Kuyvenhoven M, Van Essen G, Schellevis F, Verheij T. Management of upper respiratory tract infections in Dutch general practice: antibiotic prescribing rates and incidences in 1987 and 2001. Fam Pract. 2006;23:175-179.
37. Gommer AM, Poos MJJC. Neemt het aantal mensen met infecties van de bovenste luchtwegen toe of af? (Is the number of people with upper respiratory tract infections increasing or decreasing?). [cited 2008 Feb 28]. Available from: http://www.rivm.nl.
38. Fleming DM, Ross AM, Cross KW, Kendall H. The reducing incidence of respiratory tract infection and its relation to antibiotic prescribing. Br J Gen Pract. 2003;53:778-783.
39. Otters HBM, Van der Wouden JC, Schellevis FG, Van Suijlekom-Smit LWA, Koes BW. Trends in prescribing antibiotics for children in Dutch general practice. J Antimicrob Chemother. 2004;53:361-366.
40. Otters H, Van der Wouden JC, Schellevis FG, Van Suijlekom-Smit LWA, Koes BW. Dutch general practitioners' referral of children to specialists: a comparison between 1987 and 2001. Br J Gen Pract. 2004;54:848-852.

APPENDIX I

Clusters of other clinical manifestations (and corresponding ICPC codes) of pathogens found in hypertrophy and recurrent infections of tonsils/adenoid:

- Adenovirus: coryza (R74), otitis media (H71-H72), bronchitis (R78), pneumonia (R81), diarrhea (D73), pharyngitis (R74), conjunctivitis (F70), tonsillitis (R22/R76)
- Parainfluenza virus: laryngotracheobronchitis (R77), pneumonia (R81), bronchiolitis (R78), otitis media (H71-H72), sinusitis (R75), URI (R74), tonsillitis (R22/R76)
- RSV: bronchiolitis (R78), pneumonia (R81), coryza (R74), sinusitis (R75), otitis media (H71-H72), tonsillitis (R22/R76)
- Rhinovirus: coryza (R74), acute sinusitis (R75), otitis media (H71-H72), pharyngitis (R74), bronchitis (R78), asthma attack (R96), tonsillitis (R22/R76)
- EBV: otitis media (H71-H72), diarrhea (D73), URI (R74), infectious mononucleosis (A75), pneumonia (R81), tonsillitis (R22/R76)
- HSV: gingivostomatitis (D82), oral leasions (S71), keratitis (F73), conjunctivitis (F70), blepharitis (F72), laryngitis (R77), tonsillitis (R22/R76)
- H. Influenza: acute sinusitis (R75), otitis media (H71-H72), acute conjunctivitis (F70), exacerbation chronic bronchitis (R78), pneumonia (R81), meningitis (N71), epiglottis (R77), septic artritis (L70), osteomyelitis (L70), cellulitis (S10), tonsillitis (R22/R76)
- S. pyogenes: sinusitis (R75), cellulitis (S10), otitis media (H71-H72), acute rheuma (K71), acute glomerulonefritis (U88), impetigo (S84), cellulitis (S10), meningitis (N71), tonsillitis (R22/R72)
- S. pneumoniae: pneumonia (R81), meningitis (N71), sinusitis (R75), otitis media (H71-H72), tonsillitis (R22/R76)
- S. aureus: furunkel (S10), impetigo (S84), pneumonia (R81), folliculitis (S11), tonsillitis (R22/R76)

APPENDIX II

Cluster of possible substitution codes for hypertrophy and recurrent infections of tonsils/adenoid (R90):

R01: pain attributed to respiratory system
R04: other breathing problems
R05: cough
R07: sneezing/nasal congestion
R08: other symptoms of nose
R09: symptoms/complaints sinus
R21: symptoms/complaints throath
R22: symptoms/complaints tonsils
R29: other symptoms respiratory system
R72: strep throath/scarlet fever *
R74: URI (head cold) *
R75: sinusitis acute/chronic *
R76: tonsillitis acute *
R83: other infections respiratory system
R97: hayfever, allergic rhinitis
R99: other disease respiratory system
A01: pain generalized/unspecified
A02: chills
A03: fever
A75: infectious mononucleosis
A77: other viral diseases
A78: other infectious diseases
H71: acute otitis media/myringitis *
H72: serious otitis media, glue ear *
H74: chronic otitis media, other infections of ear*

* Codes included in the upper respiratory cluster in Figure 6.2

Chapter 7

General discussion

In this final chapter, we present the main findings of this thesis, discuss methodological issues, and elaborate on the generalizability of the findings. Finally, we address the implications of this thesis for public health, general practice, and future research.

MAIN FINDINGS

In this thesis, we developed a useful strategy for continuous surveillance of health problems based on data from electronic medical records (EMRs) collected in a national general practice database.[1;2] This overall strategy entails four subsequent steps: a) constructing episodes, b) analyzing the data, c) interpreting the results, and d) follow-up of detected trends. The following research questions were addressed, all of which refer to the use of data derived from EMRs in general practice. The term 'useful' in these questions denotes: effective, efficient, and applicable.

1. What is a useful design of an automated method for constructing episodes?
2. What is the validity of this method for the purpose of estimating morbidity rates?
3. What is a useful general strategy for constructing episodes, analyzing, and interpreting the dataset at large?
4. What is a useful follow-up strategy for investigating causal clues to a detected trend in detail?
5. Which striking trends in the incidence of health problems are detected by application of the developed overall strategy over subsequent years?

1. We designed an automated method to group diagnoses into episodes. This design was used to build EPICON, which appeared useful for estimating morbidity rates in general practice.

Diagnoses in general practice are not directly suitable for estimating morbidity rates. These diagnoses need to be grouped into episodes in order to estimate the numerators of these rates. An episode could be either one diagnosis or a sequence of diagnoses that reflect the course of disease over time, for example a tension headache leading to migraine. In other words, we need to know which diagnoses of a patient refer to the same health problem in order to estimate morbidity rates. Generally, two approaches for constructing episodes can be distinguished. In the first approach, diagnoses from contact-oriented EMRs are grouped afterwards, through manual review or use of a computerized method. The second approach involves the new generation of episode-oriented EMRs, in which the general practitioner (GP) records diagnoses directly into episodes.[3]

Chapter two describes the development of EPICON, an application for automatically grouping diagnoses from contact-oriented EMRs into episodes. Following the requirements, EPICON was designed as an extension of a semi-computerized method that was developed in previous research.[4-6] EPICON is based on a combination of logical expressions, a decision table, and knowledge extracted from manually grouped cases by case-based reasoning, a problem-solving approach in the field of artificial intelligence.[7-10] A case-based reasoner solves a problem by remembering previous similar situations and re-uses the information and knowledge from those situations. We used a dataset of manually grouped episodes, which contains implicit knowledge of the signs, symptoms, and the course of diseases, to solve the problem of grouping diagnoses.[4-6] These manually grouped episodes were used to make case bases from which previous cases could be retrieved in order to group diagnoses from a new dataset automatically into episodes. Preliminary results indicated that EPICON's performance would probably be adequate for estimating morbidity rates in general practice.

2. The validity of EPICON is adequate for the purpose of estimating morbidity rates in general practice. The implementation of EPICON to a national general practice database yielded a nationwide, continuous registry of morbidity rates in general practice.

In chapter three, we examined the internal validity of EPICON by comparing morbidity rates based on EPICON with morbidity rates from the Dutch National Survey of General Practice (DNSGP-2), which we considered to be the gold standard.[4-6] The dataset was divided into a training set, which was used to construct the case bases, and a test set that was used to test EPICON. Morbidity rates were calculated based on these episodes and compared with the morbidity rates found in the DNSGP-2. The results show that the internal validity of EPICON is adequate. We also examined the effect of case-based reasoning within EPICON. This effect is evident, the addition of case-based reasoning reduces both systematic and random error.

Chapter four describes an investigation on the external validity of EPICON. In this study, we used an independent dataset derived from GPs who record diagnoses using episode-oriented EMRs.[11;12] This dataset contained diagnoses which were manually grouped by GPs. We ungrouped these diagnoses and regrouped them automatically into episodes using EPICON. Morbidity rates based on these EPICON-grouped episodes were compared with morbidity rates based on the GP-grouped episodes, which we considered to be the gold standard. The results show that EPICON performs well for the large majority of the morbidity rates. Some of the morbidity rates, however, deviated substantially or structurally from the gold standard. We found two main reasons for these deviations: 'differences in rules between the two methods of episode construction' and 'inadequate performance of EPICON'. The latter applies especially to unspecified and general diagnoses, and therefore caution is required when EPICON is used for these morbidity rates.

We conclude that the internal as well as the external validity of EPICON is adequate for the purpose of estimating morbidity rates in general practice. Furthermore, morbidity rates based on EPICON-grouped diagnoses from contact-oriented EMRs can be used in conjunction with morbidity rates based on GP-grouped diagnoses from episode-oriented EMRs *provided that the same grouping rules are used*.

EPICON has been brought into use to generate episodes from diagnoses of practices with contact-oriented EMRs that provide data for a national general practice database.[1;2] Because of this implementation, we now have a continuous national registry of morbidity rates in general practice. This registry is a substantial addition to the existing GP registries, which are either not continuous or are regional.[13] Furthermore, the large size of this registry allows to calculate general morbidity rates of rare diseases and precise morbidity rates of common diseases.

3. We developed a general strategy, including the construction of episodes, the analysis of data, and the interpretation of findings, for surveillance of health problems based on EMRs in general practice.

Chapter five presents a general strategy for constructing episodes, and for analyzing and interpreting data from a national general practice database.[1;2] This database contains data from both contact-oriented and episode-oriented EMRs. For contact-oriented practices episodes were constructed using EPICON, whereas for episode-oriented practices we used the episodes recorded by GPs.

Analyses of data derived from EMRs in general practice have to take into account that this data is clustered within patients and within practices, which requires the use of multilevel analyses.[14;15] Prevalence rates, which are dichotomous data (i.e. one counts whether or not a patient has a particular health problem) can be estimated with multilevel logistic regression analyses, while incidence rates, which are count data (i.e. one counts the number of new episodes of a particular health problem) require multilevel Poisson regression analyses. Furthermore, analyses of this data for surveillance purposes have to deal with the large range of health problems covered.

Therefore, we developed general multilevel regression models that can be applied to each health problem. These models are useful to detect striking trends in the abundance of available data. The analyses result in adjusted morbidity rates that are corrected for clustering, age, sex, and recording period. We also calculated raw morbidity rates (i.e. common rates) and crude morbidity rates (i.e. rates corrected for clustering only) to acquire insight into the results of the general models. The comparison between raw and crude rates showed that clustering may have considerable impact on morbidity rates in general practice. Therefore, it is advisable to correct morbidity rates for clustering when data is clustered within levels, for instance in morbidity rates based on data from different general practices or from different hospitals.

In order to interpret the results of the analyses, the full range of factors that could affect registration of patient data in EMRs in general practice has to be taken into account. Possible causes of changing incidence rates were listed and grouped into four categories: changes in recording, changes in the demand for care, changes in the supply of care, and 'real' changes in the incidence of disease. Before considering a 'real' change in the incidence of disease, possible causes that fall into the first three categories should be excluded. The developed list can be used to check whether all factors have been taken into consideration, and the list can serve as a starting point for more detailed studies.

4. We developed a follow-up strategy to investigate causal clues to a detected trend. Application of this strategy to the database permits further exploration whether detected trends are due to 'artefacts' or might reflect a 'real' change in the incidence of disease.

In chapter six, we studied a detected trend in detail, i.e. the decreasing incidence of adenotonsillar problems, based on the available evidence in the general practice database. Specific hypotheses were formulated about the cause of the decrease based on the developed list of possible causes of changing incidence rates. The main aim of this study was to examine whether the decline reflects a 'real' decrease in the incidence of disease or is merely originated by 'artefacts', like a change in recording. Indications for a 'real' change in the incidence of disease were examined by estimating incidence rates of clusters of other diseases that share the same pathogens with adenotonsillar problems. The artefact hypotheses were tested by constructing explanatory variables from the available data that is recorded in the EMRs, such as prescriptions and referrals. These explanatory variables were then entered stepwise into the general multilevel regression models. The available evidence enabled to study the most obvious hypotheses for the observed decrease over time.

5. Application of the developed strategy over subsequent years revealed several striking trends in the incidence of health problems, which provided insights that were not available otherwise.

The developed general strategy was used to monitor the incidence of health problems in general practice over subsequent years (chapter five). In the period 2002-2005 several striking trends emerged in the Netherlands. In particular, we found increasing concern about reactions to drugs among middle-aged and elderly patients. This increase might be related to the simultaneous increase in the use of over-the-counter drugs.[16;17] In addition, we found an increase in 'a change in feces/bowel movements', which might be due to a nationwide campaign explaining that a change in feces/bowel movements is one of the possible signs of colon cancer.[18;19] In addition, we detected an increase in the incidence of urogenital problems (including urination problems) among men and women, fear of veneral disease among men aged 15-44 years, and prostate problems among middle-aged and older men. Some of these problems might be explained by the availability of, and campaign for, new drugs for benign prostate hyperplasia in the period under study, whereas others could reflect an increase in the number of sexually transmitted diseases and urinary tract infections.[20-22]

The study also revealed a decrease in the incidence of hypertrophy and recurrent infections of tonsils/adenoid among the total population, which was explored in more detail in chapter six. In this study, we examined (possible causes of) this health problem among children aged 0-14 years. The results show that the incidence rate decreased significantly (p=0.02) from 3.0 to 1.3 per 1000 children per year. Part of this decline can be explained by changes in the demand for and supply of care during this period, whereas no indications were found for a change in recording. Although unmeasured factors in demand or supply may further explain the decrease, a 'real' decline in the incidence of disease, possibly related to an overall decline in upper respiratory tract diseases, cannot be ruled out.

METHODOLOGICAL CONSIDERATIONS

A broad approach was used to develop an overall strategy for continuous surveillance that combines elements from epidemiology, public health, medical informatics, and general practice. This broad approach enabled us to develop a strategy that entails four subsequent steps: the construction of episodes, the analysis of data, the interpretation of data, and the follow-up of detected trends. The main drawback of this broad approach is that we could not elaborate on all elements in great detail.

Development of EPICON

We started with the development of EPICON, an application for grouping diagnoses into episodes. The major strength of EPICON is that it enables the automated construction of episodes from diagnoses recorded in contact-oriented EMRs. This method is very efficient compared to the time-consuming and expensive manual grouping method that was used in the DNSGP-2.[6] This manual grouping method, which was used to group diagnoses from 89 practices for one year only, took about one year and cost approximately 200,000 euro. Because of these high annual costs, the manual grouping method was not performed on a yearly basis. In comparison: EPICON runs for about half a day to construct episodes for an annual dataset, while the costs for developing and maintaining EPICON are relatively low. Thus, EPICON is an efficient method that allows to generate morbidity rates that would not have been available otherwise.

In line with the requirements, EPICON was built on the grouping methods that were developed in the DNSGP-2. This choice was efficient; we could extend the semi-computerized method and use the manually grouped diagnoses as a case library. The

DNSGP-2 was consider to be the gold standard, which is plausible taking into account the careful data collection and processing procedures as well as the widespread use of this data. There is no evidence, however, to support this assumption. Consequently, possible flaws in the DNSGP-2 might have been passed on to EPICON.

We developed a practical, case-based application using a simple and transparent design. We selected case-based reasoning as an approach, because the domain knowledge that is needed to group diagnoses into episodes is implicit knowledge, which lends itself more to reasoning based on analogy (i.e. resemblance to previous cases) than to formulating 'if-then' rules, or for constructing a causal or functional model. Furthermore, ample cases were available because the DNSGP-2 dataset provided an extensive case library. However, because we ceased designing after we had developed the first workable application, EPICON could be improved further. Suggestions for further development are described in the final section under 'Future research'.

Evaluation of EPICON

We then evaluated EPICON extensively, which is one of the strengths of this thesis. During the development phase, misclassified diagnoses were explored which provided feedback about the actual grouping process. This can be considered a formative evaluation. We also performed two summative evaluations in which we examined whether EPICON fulfills its purpose, i.e. whether it is useful for estimating morbidity rates in general practice. Both the internal and external validity were examined. A limitation of both these studies is that we performed the evaluation ourselves, whereas an evaluation by independent researchers would have been more objective. Furthermore, the evaluations are based on subjective judgment as to whether deviations from the gold standard are considered small or large. The figures provided in chapters three and four, however, allow to judge for oneself.

Dataset used for surveillance

Subsequently, EPICON was implemented in the Netherlands Information Network of General Practice (LINH), which enabled us to actually monitor health problems over time.[1;2] The dataset derived from this network has several strengths and limitations. First of all, this dataset contains the necessary information to study morbidity rates in general practice, namely:

a) Coded diagnoses of consultations, prescriptions, and referrals. Diagnoses of consultations, prescriptions and referrals are not only necessary to estimate the numerator of the morbidity rates, but they also provide important information on the demand for and supply of care.
b) Information whether or not a recorded diagnosis represents the start of a new episode, which is necessary to distinguish between incidence and prevalence rates.
c) Information about the size of the population at risk, which is needed to determine the denominator of the morbidity rates.
d) Basic patient characteristics, such as age and sex, which enable analysis at the patient level.
e) Basic practice characteristics, such as region of residence and degree of urbanization, which allow analyses at the practice level.

Second, the secondary use of data from EMRs in general practice is efficient and prevents or reduces certain forms of bias that may occur in primary data collection, such as bias as a result of non-response, differential recall or conducting the study (which may introduce a biased shift in attention). Furthermore, data from EMRs provide cases that are diagnosed by a physician, which provides more valid information than self-reporting of diseases. Third, major strengths of this dataset compared to other Dutch GP registries are the large size and the representativeness for the general population and for general practices in the Netherlands.

Two characteristics that ensure the representativeness of this dataset for all Dutch practices are worth mentioning. The first is that the data derive from general practice information systems of five different manufacturers. Although software designers have slightly adapted the systems of the LINH practices in order to achieve consistent recording,[1] the differences in design between these systems remain a largely unknown source of variability. An educated guess is that differences in design have a considerable impact on the accuracy and completeness of recording.

The second characteristic is that the dataset consists of a dynamic pool of practices, which differs from year to year as some practices leave the network and others join up. Although this limits the possibility to conduct longitudinal research, the turnover of the participating practices is small; in 2005, 89% of the practices had been participating for more than five years and 27% for more than ten years.[1] Hence, the dataset allows for longitudinal research based on a limited selection of practices that have participated over several years.

An important limitation inherent to the secondary use of data from EMRs is that research of diseases, exposures, and adjustment for confounders is limited to the available data. Furthermore, the possibility to control the quality of secondary data from EMRs is rather limited compared to quality control in primary data collection. Within LINH, much effort is put into ensuring the quality of the database. This requires a balance between interventions to control the quality of the data on the one hand and the extra workload for the participating GPs on the other. A detailed overview of all the quality measures that are taken is provided elsewhere.[1] Nevertheless, when we checked the data in order to monitor health problems over time, the quality of recording appeared to be insufficient for a number of practices. Therefore, we defined clear selection criteria for accuracy and completeness, resulting in a considerable loss of practices, especially in 2004 and 2005. Furthermore, the selected practices record not always 100% completely, which could mean that the calculated rates underestimate the true rate to a limited extent.

GENERALIZABILITY

Transportability of EPICON

The first step of the developed strategy for surveillance is the construction of episodes. To this end, we designed, built, and evaluated EPICON. The results from the external validation indicate that EPICON is transportable to other, similar datasets of diagnoses in general practice. In particular, EPICON can be applied to group diagnoses derived from contact-oriented EMRs that are coded by the International Classification of Primary Care (ICPC)[23] and that are characterized as either 'new' or 'ongoing'.

EPICON is also applicable to the new generation of episode-oriented EMRs.[3] These EMRs have been developed as a result of the 2005 guideline of the Dutch College of General Practitioners, which makes it compulsory that every newly introduced system provides the possibility to record episodes.[24] This guideline is now being implemented. At present, most Dutch practices use contact-oriented EMRs, while some of them use the new episode-oriented EMRs. A possible threat to the validity of morbidity rates derived from these episode-oriented EMRs is the unknown variation in the grouping rules used by different practices. The first experiences show variation between different general practice information systems in the user-friendliness of the possibility to record episodes as well as variation between GPs in the actual recording of episodes.[25] The results of this thesis indicate that rules for grouping diagnoses into episodes may have considerable impact on the morbidity rates. A possible application of EPICON in episode-oriented EMRs is to discover differences in (handling) the rules used by different GPs, i.e. EPICON could serve as a standard to compare with.

EPICON, which was originally developed using data from 2001, is applicable for the coming years. In the long run, however, changing medical insights that affect grouping rules may limit the usability of EPICON.

Generalizability of data derived from EMRs in relation to the health care system

Two characteristics of the Dutch health care system affect the generalizability of the data from the EMRs that were used in this study. These characteristics are the presence of a gatekeeping system and the availability of patient lists. In a gatekeeping system, morbidity rates in general practice provide a good indication of the health status of the general population.[13;26] In countries where patients have free access to specialist care, morbidity rates in general practice still provide an indication of the health status of the general population, but there is a higher chance of morbidity 'leaking' to other health care providers. Furthermore, we used patient lists to determine the size of the population at risk in the denominator. These lists are available in countries where every patient is registered with a GP,[27] which makes it easy to determine the denominator of the morbidity rates. Thus, the Netherlands has favorable conditions for surveillance of public health based on data from EMRs in general practice, because it utilizes both a gatekeeping and a list system.

Generalizability of the developed strategy to other GP registries

When there is a similar database in a country with a gatekeeping and a list system, the developed surveillance strategy is directly applicable to another GP registry. Either the strategy or the database could be adapted to accommodate departures from these conditions. In countries without a list system, for instance, other methods are available to determine the size of the population at risk.[28] In addition, when the characterization of a diagnosis as either 'new' or 'ongoing' is lacking, this field could be added to an existing EMR system.

Generalizability of the developed strategy to other disciplines

The developed strategy cannot be easily adopted by other disciplines. Compared to other health care providers, GPs (at least in the Netherlands) are advanced in the use of EMRs. For instance, the Dutch College of General Practitioners develops requirements for general practice information systems, produces guidelines for consistent use of these systems, and promotes the use of a coding system, i.e. the ICPC.[23-25] Such activities are not yet common among most other health care providers. Consequently, EMRs in general practice are at present and generally speaking more useful for extracting data for epidemiologic research than EMRs that are used by other disciplines.

Furthermore, the episode construction within the developed strategy, i.e. the relation between diagnoses, is specific to general practice. Compared to other medical doctors, GPs encounter many patients with a broad range of symptoms, signs and diseases, and will often have to make an educated guess about the relations between them. For instance, the probability that abdominal pain is caused by an appendicitis (i.e. both diagnoses should be grouped into one episode called appendicitis), might be lower in a general practice setting than in a hospital setting. The procedures followed to construct episodes, however, could be adopted by other disciplines, i.e. using manually grouped diagnoses to build another episode constructor or to develop episode-oriented EMRs. Once a database with episodes is available, the other steps of the developed strategy could be adapted to accommodate surveillance based on EMRs of other health care providers.

IMPLICATIONS FOR PUBLIC HEALTH, GENERAL PRACTICE, AND FUTURE RESEARCH

Public health

The developed strategy for continuous surveillance of health problems in general practice provides policymakers with relevant information that may serve a broad range of purposes, such as signaling striking and unusual disease patterns, targeting and evaluating interventions, and planning health care services. Overall, this thesis illustrates the potential use of data derived from EMRs for monitoring public health. Development of methods that disclose and use data from EMRs is an important topic for future surveillance of public health.

In this thesis, we applied the developed strategy to detect striking trends in the incidence of health problems in the Netherlands. Some of the results of this study merit attention of public health professionals. In particular, attention should be paid to the increasing concern about reactions to drugs among middle-aged and elderly patients in the Netherlands and the possible relation with the availability of over-the-counter drugs.[16;17] In addition, the increasing incidence of health problems in the urogenital area should be monitored closely as it could reflect increases in the incidence of sexually transmitted diseases and urinary tract infections.[20-22] Finally, the results of this thesis show a decrease in the incidence of adenotonsillar problems among children, which might be part of an overall decline in upper respiratory tract infections in the population.[29-32] Further insight into this issue is important for planning health care services, because upper respiratory tract infections form a substantial part of the workload in general practice, ear, nose, and throat surgery, and pediatrics.

General practice

An essential requirement for surveillance based on EMRs in general practice is adequate record keeping, especially accurate and complete recording of ICPC codes. An additional requirement for episode-oriented EMRs is the consistent use of uniform rules for grouping diagnoses into episodes.

Adequate medical record keeping is primarily important for the treatment and safety of patients. Problems in the transfer of information can jeopardize the quality of patient care. Consequences of faults in medical information transfer in curative care include wrong medication (44%), no treatment because of a lack of information (25%), and wrong operation or treatment (24%).[33] The increasing complexity of patient care increases the need of health care providers to get a quick overview of the diagnoses and treatments of a patient which, in turn, increases the necessity of adequate medical record keeping.

We expect that recent developments within Dutch general practice will further enhance the quality of medical record keeping. In order to improve the overview and exchangeability of information, the Dutch College of General Practitioners has produced a guideline for adequate record keeping within EMRs in general practice that advocates the routine use of ICPC codes.[34] In addition, the information systems of general practice out-of-hours cooperatives will be able to communicate with general practice information systems in the near future. The Dutch College of General Practitioners also produced a guideline for this exchange of information between the GP and the cooperative.[35] ICPC codes constitute an essential part of this information exchange, which will further enhance accurate and complete recording of these codes. Moreover, to improve adherence to these guidelines, a tool has been developed that enables GPs to acquire insight into their own recording habits. This tool, the general practice EMR scan, uses a set of indicators to assess the quality of recording within EMRs in general practice.[36] Probably, this tool will contribute to further enhancement of the quality of medical record keeping within general practice, which will serve both the primary and the secondary use of this data.

The results of this thesis, especially the findings in chapter four, could be used to discuss variation in the rules that are used to group diagnoses within episode-oriented EMRs. Agreement about those rules would further enhance the quality of medical record keeping. We recommend that the Dutch College of General Practitioners start this discussion in order to formulate general, uniform grouping rules, which could be included in the next update of the guideline for adequate record keeping.

Future research

EPICON

Future research should aim at a follow-up of the results of the evaluation of EPICON. The generalizability of EPICON in cumulative tests across diverse settings needs to be confirmed. Furthermore, the results of the evaluation indicate which aspects of EPICON could be improved, i.e. the grouping of unspecified and general diagnoses.

A first step in the further development of EPICON could be to refine this tool. To start with, the duration of a disease could be added to the method, which is an important criterion in deciding whether or not two diagnoses should be grouped together. The actual number of days between each pair of diagnoses in the same patient in the DNSGP-2 dataset can be used to make useful categories of disease duration. These categories could then be added to the case bases. Duration of disease could also be added to the decision table. Furthermore, the decision table could be revised by a panel of experts in order to establish broader consensus about the decisions in this table. Information about age and sex could also be included in the case bases. However, this addition will only be useful when age or sex is relevant for deciding whether or not diagnoses should be grouped together. For example, information about sex is not helpful when the signs, symptoms, and course of a disease are equal for both sexes.

A second step could be to add cases. New cases might be derived easily from data recorded in episode-oriented EMRs. A basic question that may help to decide whether or not new cases should be added is: Will these new cases help to improve the performance of EPICON? In particular, will the effect be large enough to have an effect on the generated morbidity rates? Which rules were used to group these new cases and are they similar to the grouping rules used in the DNSGP-2? The performance will not improve by adding cases that are based on different grouping rules. In general, adding cases could improve the performance where EPICONs' decisions are based on a small number of cases.

A third step could be to develop an extension that provides an indication as to whether or not a diagnosis represents the start of a new episode from the recorded disease history of a patient. This step would depart from the initial requirement for developing EPICON, i.e. that it should be in line with the basic methods used in the DNSGP-2. However, the suggested extension might produce more valid information about the start of a new episode than the field ‘new’ or ‘ongoing’. Moreover, it would enable the inclusion of diagnoses from prescriptions and referrals into the episode construction process. These advantages would probably enhance the validity of the morbidity rates derived from EMRs in general practice and we therefore advocate the development of such an extension.

EPICON uses only the first two processes of a general case-based reasoning cycle, i.e. the retrieve and the reuse process. EPICON could be extended to a learning system by building the other two processes, i.e. the revise and the retain process, into the system as well. For example, EPICON could suggest an episode for a diagnosis that needs to be grouped to a GP (retrieve and reuse) and incorporate the final decision of the GP (revise) into the case base for future use (retain). Hence, EPICON could be extended to a learning, and interactive decision support system within episode-oriented EMRs, that would assist the GP in the construction of episodes.

Epidemiologic surveillance

This thesis shows that data derived from EMRs in general practice are a valuable source for epidemiologic surveillance. However, this data is not readily available by pushing a single button. The use of this data for the purpose of surveillance requires a special strategy for processing, analyzing, and interpreting this information. Further research should focus on the development of methods that enable the use of EMRs in general practice for epidemiologic surveillance, in particular methods for ensuring the quality of this data.

In addition, more research is needed to examine differences in morbidity rates from various computerized GP registries. In the Netherlands, these registries differ in factors such as region of residence, patient population, classification system used to code diagnoses, measures taken to ensure the quality of the registration, practice software, and methods used to calculate morbidity rates.[13] All of these factors are potential sources of variation in morbidity rates derived from EMRs in general practice. The national public health agency in the Netherlands recently started a comprehensive study on the impact of these factors on morbidity rates from different GP registries. Gijsen and Poos provide various suggestions to improve the comparability of these rates, such as publishing the characteristics, rules, and quality procedures of the registries. One of their suggestions is to establish a countrywide database that includes data from all GP registries, or even from all GPs in the Netherlands.[13] This solution would enable control of relevant sources of variation and has the added advantage that consensus on definitions, rules, and procedures has to be reached, which will be a necessary requirement for such a project. We recommend to establish this countrywide GP database as it would enhance the validity of morbidity rates derived from EMRs in general practice.

Together with advances in informatics, the number of databases containing routinely collected (medical) data is growing rapidly. An important advantage of this development for epidemiologic surveillance is the possibility to link data from EMRs in general practice to other databases with relevant information, such as cause of death statistics. Box 7.1 presents a picture of a future that would be very advantageous for surveillance based on secondary data. An important disadvantage of this development is the threat to privacy and the opportunity for misuse. Researchers should be fully cognizant of this heightening risk and install advanced procedures to protect security and maintain the trust of the public.

Box 7.1 Where are we going?

"Imagine a country where all citizens are given a personal identification number at birth, which they keep for the rest of their lives, and where most written information generated by public authorities is stored in computers and is identifiable through this identification number. Imagine that this information includes an electronic medical file, all contacts to the health care system, all diagnoses made, all prescribed medicines, all social benefits, all birth defects, all immunizations, and more. Image that a similar registration system is used for income, work history, education, social grouping, and residence, and then envision a register system that can link family members together and link the members of society to huge biobanks that include everyone in the population. Image that all this information is stored and kept over time. In this vision, the entire country is a cohort." [37]

REFERENCES

1. Tacken MAJB. Quality of preventive performance in general practice: the use of routinely collected data. [dissertation]. Radboud University Nijmegen; 2005.
2. Verheij RA, Te Brake JHM, Abrahamse H, Van den Hoogen H, Braspenning J, Van Althuis T. Landelijk Informatie Netwerk Huisartsenzorg. Feiten en cijfers over huisartsenzorg in Nederland (Netherlands Information Network of General Practice. Facts and figures on Dutch GP care) [document on the internet]. Utrecht: Nivel; 2008 [cited 2008 Jun 22]. Available from: http://www.linh.nl.
3. Okkes IM, Groen A, Oskam SK, Lamberts H. Advantages of long observation in episode-oriented electronic patient records in family practice. Methods Inf Med. 2001;40:229-235.
4. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, Van der Zee J. Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health. 2005;15:59-65.
5. Westert GP, Jabaaij L, Schellevis FG, editors. Morbidity, performance and quality in primary care: Dutch general practice on stage. Oxon: Radcliffe Publishing; 2006.
6. Van der Linden MW, Westert GP, De Bakker DH, Schellevis FG. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Klachten en aandoeningen in de bevolking en in de huisartspraktijk (The second Dutch national survey of general practice. Health and problems in the Dutch population and in general practice). Utrecht, Bilthoven: NIVEL, RIVM; 2004.
7. Kolodner JL. Case-based reasoning. San Mateo: Morgan Kaufmann; 1993.
8. Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications. 1994;7:39-59.
9. Watson I, Marir F. Case-based reasoning: a review. The Knowledge Engineering Review. 1994;9:355-381.
10. Leake DB. Case-based reasoning. Experience, lessons & future directions. Menlo Park: American Association for Artificial Intelligence; 1996.
11. Van der Veen WJ, Meyboom-de Jong B. Age and Gender. In: Jones R, Britten N, Culpepper L, Gass DA, Grol R, Mant D, et al, editors. Oxford textbook of primary medical care. Oxford: Oxford University Press; 2004. 153-161.
12. Van der Werf GT, Smith RA, Stewart RE, Meyboom-de Jong B. Spiegel op de huisarts; over registratie van ziekte, medicatie en verwijzingen in de geautomatiseerde huisartspraktijk. Groningen: University of Groningen; 1998.

13. Gijsen R, Poos MJJC. Using registries in general practice to estimate countrywide morbidity in The Netherlands. Public Health. 2006;120:923-936.
14. Hedeker D, Gibbons RD. Longitudinal data analysis. New York: Wiley; 2006.
15. Twisk JWR. Applied multilevel analysis. Cambridge: Cambridge University Press; 2006.
16. College voor Zorgverzekeringen. Overzicht beleidsmaatregelen farmaceutische zorg 1995-2006 (Review policy measures pharmaceutical care 1995-2006) [document on the internet]. Diemen: CVZ; 2008 [cited 2008 Jan 14]. Available from: http://www.gipdatabank.nl.
17. Statistics Netherlands. Gezondheid en zorg in cijfers 2006 (Health and care in figures 2006). Voorburg/Heerlen: Statistics Netherlands; 2006.
18. Maag Lever Darm Stichting. Darmkanker.info (Coloncancer.info) [document on the internet]. Nieuwegein: MLDS; 2008 [cited 2008 Jan 14]. Available from: http://www.darmkanker.info.
19. Hill & Knowlton. Jaarverslag Maag Lever Darm Sichting 2004 (Annual report Stomach Liver Intestine Foundation 2004). Rotterdam: The happy horseman; 2004.
20. Huppert JS, Biro F, Lan D, Mortensen JE, Reed J, Slap GB. Urinary symptoms in adolescent females: STI or UTI? J Adolesc Health. 2007;40:418-424.
21. Van de Laar MJW. SOA nemen opnieuw toe (STDs increasing again). Infectieziekten bulletin. 2005;16:116-117.
22. Van de Laar MJW, Van Veen MG. Soa samengevat (STDs summarised) [document on the internet]. Bilthoven: RIVM; 2008 [updated 2008 Apr 17; cited 2008 Jan 14]. Available from: http://www.nationaalkompas.nl.
23. Lamberts H, Woods M. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.
24. Werkgroep revisie referentiemodel. HIS referentiemodel 2005 (Working group revision reference model. General practice information system reference model 2005). Utrecht: NHG; 2005.
25. Njoo K. ADEMD ondersteuning in HIS'en: een update (Supporting adequate record keeping using the electronic medical record in general practice information systems: an update). Utrecht: NHG; 2007.
26. Linden M, Gothe H, Ormel J. Pathways to care and psychological problems of general practice patients in a "gate keeper" and an "open access" health care system. A comparison of Germany and the Netherlands. Soc Psychiatry Psychiatr Epidemiol. 2003;38:690-700.
27. Boerma WG. Profiles of general practice in Europe. PhD [dissertation]. Utrecht: Nivel; 2003.
28. Bartholomeeusen S, Kim CY, Mertens R, Faes C, Buntinx F. The denominator in general practice, a new approach from the Intego database. Fam Pract. 2005;22:442-447.

29. Diepenhorst HIJ, Otters HBM, Van Suijlekom-Smit LWA, Schellevis FG, Van der Wouden JC. Tonsillitis acuta bij kinderen in de huisartsenpraktijk: verandering van incidentie en beleid? (Acute tonsillitis in children in the GP's practice: changes in incidence and policy?). Huisarts Wet. 2004;47:399-404.
30. Fleming DM, Ross AM, Cross KW, Kendall H. The reducing incidence of respiratory tract infection and its relation to antibiotic prescribing. Br J Gen Pract. 2003;53:778-783.
31. Gommers AM, Poos MJJC. Neemt het aantal mensen met infecties van de bovenste luchtwegen toe of af? (Is the number of people with upper respiratory tract infections increasing or decreasing? [document on the internet]. Bilthoven: RIVM; 2008 [updated 2008 Apr 17; cited 2008 Feb 28]. Available from: http://www.nationaalkompas.nl.
32. Kuyvenhoven M, Van Essen G, Schellevis F, Verheij T. Management of upper respiratory tract infections in Dutch general practice: antibiotic prescribing rates and incidences in 1987 and 2001. Fam Pract. 2006;23:175-179.
33. Foekema H, Hendrix C. Fouten worden duur betaald. Een onderzoek naar medische overdrachtsfouten (deel 2) (Faults are paid dearly. A study into faults in medical handovers (part 2)). Amsterdam: TNS NIPO; 2004.
34. Njoo KH. Richtlijn Adequate dossiervorming met het Elektronisch Medisch Dossier (Guideline adequate record keeping using the electronic medical record). Utrecht: NHG; 2004.
35. NHG/NICTIZ. Richtlijn gegevensuitwisseling huisarts en Centrale Huisartsen Post (CHP) (Guideline data exchange general practitioner and general practice out-of-hours cooperative). Utrecht, Leidschendam: NHG/NICTIZ; 2005.
36. Jabaaij L, Verheij R, Njoo K, Van den Hoogen H, Tiersma W, Levelink H. Het meten van de kwaliteit van de registratie in elektronische patiënten dossiers van huisartsen met de EPD-scan-h (Measuring the quality of recording within electronic medical records of general practitioners using the EMR-scan-GPs). Utrecht: Nivel; 2008.
37. Olsen J. Using secondary data. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3th ed. Philadelphia: Lippincott, Williams & Wilkins; 2008. 482.

SUMMARY

Epidemiologic surveillance is defined as *'the continuous and systematic process of collection, analysis, interpretation and dissemination of descriptive information for monitoring health problems'*. Surveillance is of major importance for various activities within public health, such as signaling unusual disease patterns, targeting and evaluating prevention and control measures, and planning health care facilities. In this thesis, we developed a strategy for continuous surveillance of health problems of the Dutch population based on data from Electronic Medical Records (EMRs) in general practice. This strategy entails four subsequent steps:

a. constructing episodes
b. analyzing the data
c. interpreting the results
d. follow-up of detected trends

The studies presented in this thesis encompass both the development and the application of this strategy. The thesis includes five studies, which are addressed in separate chapters.

Chapter 1 is an introduction to the topic. It starts with a description of the background and the history of surveillance. We explain two basic concepts from the field of epidemiology: the prevalence (a measure for existing cases of disease) and the incidence (a measure for new cases of disease). These measures are used to monitor the frequency and the distribution of health problems. Next, we describe the state-of-the-art of surveillance of health problems in the Netherlands based on data from EMRs in general practice. In order to calculate prevalence and incidence rates based on these records, the separate diagnoses of an individual patient need to be grouped into episodes. An episode comprises all diagnoses that refer to the same health problem. For instance, the general practitioner (GP) has assigned three separate, subsequent diagnoses to a particular patient: coughing, fever, and bronchitis, that together comprise one episode named bronchitis. Up until now these episodes were largely manually constructed.

Chapter 2 describes the development of EPICON, an application that automatically groups the diagnoses from EMRs in general practice into episodes. These episodes can be used to estimate morbidity rates (prevalence and incidence rates). This study builds upon an earlier large-scale investigation among Dutch GPs: the second Dutch National Survey of General Practice (DNSGP-2). In this previous study, a semi-computerized

method was developed to group diagnoses from EMRs in general practice into episodes. Using this latter method, episodes are generated automatically for easy-to-group diagnoses, whereas difficult-to-group diagnoses are grouped manually. EPICON was designed as an extension of this semi-computerized method. Within EPICON, the difficult-to-group diagnoses are automatically grouped using analogous, or 'case-based', reasoning, a method derived from the field of artificial intelligence. This way of reasoning assumes that a new problem can be solved by applying solutions from previous similar problems that have been stored in the memory (cases). EPICON uses the manually grouped diagnoses from the DNSGP-2 as 'cases' to automatically group diagnoses from a new dataset into episodes. A preliminary evaluation based on misclassifications indicated that EPICON performed sufficiently well to base morbidity rates upon.

In order to judge whether this application could actually be brought into use, EPICON had to be evaluated for the purpose for which it was to be used: estimating morbidity rates based on EMRs in general practice. Therefore, we performed two studies to evaluate EPICON at this point. In the first evaluation, described in *Chapter 3*, we studied the internal validity of EPICON, i.e., the performance of EPICON was tested using a sample from the dataset that was used to develop EPICON (the DNSGP-2 dataset). The results show that morbidity rates based on EPICON generally deviate only slightly from the gold standard (the DNSGP-2 rates). This allowed us to conclude that the internal validity of EPICON is adequate for estimating morbidity rates based on EMRs in general practice.

Chapter 4 describes the second evaluation: a study on the external validity (generalizability) of EPICON. In this investigation, the performance of EPICON was examined in an independent dataset whose data were recorded by GPs in episode-oriented EMRs. Within this new generation of EMRs, GPs manually group the diagnoses of a patient into episodes in the EMR of that patient. These episode-oriented EMRs are currently implemented in a number of Dutch general practices. In this study, the manually-grouped diagnoses were ungrouped and regrouped by EPICON. Next, we compared the morbidity rates based on EPICON with the rates based on the manually-grouped episodes, which we considered to be the gold standard in this study. EPICON proved to perform well for the great majority of the morbidity rates. However, a small part of the morbidity rates (5%) showed structural or large deviations. From this investigation we concluded that the external validity of EPICON is sufficient, but caution is required when EPICON is used to estimate morbidity rates of general or unspecified health problems.

Based on these results, EPICON was applied to construct episodes for a nationally representative database using data from EMRs in general practice (the LINH database, the Netherlands Information Network of General Practice). This implementation allows to routinely monitor health problems of the Dutch population on the basis of data from EMRs in general practice. To that end, a general surveillance strategy was needed that was useful for the whole dataset. *Chapter 5* describes this general strategy that includes the construction of episodes, the analyses of data, and the interpretation of the results.

The **first step** was the construction of episodes. The national database includes data from the usual (contact-oriented) EMRs as well as data from the new episode-oriented EMRs. For contact-oriented EMRs, episodes were constructed using EPICON; for episode-oriented EMRs, the episodes recorded by the GPs were used.

The **second step** was to analyze the data. An important question was how to deal with the large quantity of available data. Data were available for about 680 different health problems acquired over several years and it was impracticable to set up a separate analysis for each health problem. In addition, in the set-up of the analyses, we had to take into account the clustering of data within patients and within practices. Therefore, we developed a general, statistical model that takes into account clustering, age, and sex, and that can be applied to each health problem (a multilevel regression model).

The **third step** was to interpret the results emerging from these analyses. To that end, we had to consider various factors that could affect the registration of data within EMRs in general practice. Therefore, all possible events that could cause a change in morbidity rates were listed and grouped into four main categories: changes in recording, changes in the demand for care, changes in the supply of care, and 'real' changes in the number of new cases of disease.

We applied this general surveillance strategy to monitor health problems in the Netherlands over the period 2002-2005. In this study we detected several striking trends over time. It emerged that the concern of patients about reactions to drugs had increased, especially among patients aged 45 years and older. This might be explained by the increased availability and use of over-the-counter-drugs in this period of time. In addition, we found an increase in various urogenital problems, including prostate problems, fear of veneral diseases, and urination problems. Possible explanations for the increase in these types of problems are: the availability of, and campaigns for, new drugs for benign prostate hyperplasia, an increase in the number of sexually transmitted diseases, and an increase in the number of urinary tract infections. Another finding of this study was a decrease in the incidence of hypertrophy and recurrent infections of tonsils/adenoid.

Possible causes for this decrease in hypertrophy/recurrent infections of tonsils/adenoid among children were investigated in *Chapter 6*, which describes a follow-up strategy to find clues to possible causes for a detected trend. In this **fourth step**, we formulated specific hypotheses based on the list of possible causes for changes in morbidity rates. We examined these hypotheses based on the available information in the database, including data on consultations, referrals, and prescriptions of medication. Then, explanatory variables were constructed that were entered stepwise into the general statistical model. This fourth step allows to explore whether a detected trend is caused by 'artefacts' (like a change in recording) or whether there might be a 'real' change in the incidence of disease. This follow-up study was conducted among children aged 0-14 years, i.e., the age group in which hypertrophy and recurrent infections of tonsils/adenoid occur most frequently. The results show that part of the decline of this health problem could be explained by changes in the demand for care (fewer consultations) and in the supply of care (fewer referrals and prescriptions). However, no explanation was found for the residual decline. This decrease is possibly related to a 'real' decline in upper respiratory tract infections.

In *Chapter 7*, we first summarize the most important findings of this thesis before discussing various methodological issues. The surveillance strategy is developed from a broad approach that combines elements from different scientific fields. The drawback of this broad approach is that we could not elaborate on all elements in great detail. We also discuss the advantages and disadvantages of the particular database that was used. Important advantages are the representativeness for the Dutch population, the large amount of data, and the fact that the data has already been collected. A disadvantage is the limited possibility to control the quality of recording. Next, the applicability of the developed surveillance strategy is addressed. Under certain conditions, EPICON can be applied to other general practice registries. In addition, EPICON can serve as a standard against which to compare other episode constructions. When considering the surveillance strategy as a whole, it seems to be particularly suited for application in other general practice registries rather than for application in other disciplines. Finally, we discuss the implications of this thesis for public health, general practice, and future research. This thesis shows that data derived from EMRs in general practice are valuable for surveillance provided that methods are available for disclosing and using this data. Future surveillance should focus on the further development of methods for the secondary use of data derived from EMRs. For example, methods to enhance the quality of recording or the linking of data from EMRs to relevant information from other databases.

SAMENVATTING

Epidemiologische surveillance is gedefinieerd als *'het continue en systematische proces van verzameling, analyse, interpretatie en verspreiding van beschrijvende informatie voor het monitoren van gezondheidsproblemen'*. Surveillance is van groot belang voor allerlei activiteiten binnen de volksgezondheidszorg, zoals het signaleren van onverwachte ziektepatronen, het instellen en evalueren van maatregelen voor preventie of controle van ziekten en het plannen van gezondheidszorgvoorzieningen. Wij hebben in dit proefschrift een strategie ontwikkeld voor continue surveillance van gezondheidsproblemen van de Nederlandse bevolking op basis van gegevens uit Elektronische Medische Dossiers (EMDs) in de huisartspraktijk. Deze strategie bestaat uit vier opeenvolgende stappen:

a. constructie van episodes
b. analyse van gegevens
c. interpretatie van de resultaten
d. follow-up van ontdekte trends.

De studies in dit proefschrift omvatten zowel de ontwikkeling als de toepassing van deze strategie. Het proefschrift bestaat uit vijf studies die in aparte hoofdstukken worden behandeld.

Hoofdstuk 1 is een inleiding op het onderwerp. Het begint met een beschrijving van de achtergrond en de geschiedenis van surveillance. We lichten twee basisbegrippen uit de epidemiologie toe: de prevalentie (een maat voor bestaande ziektegevallen) en de incidentie (een maat voor nieuwe ziektegevallen). Deze maten worden gebruikt om de frequentie en de verspreiding van gezondheidsproblemen te monitoren. Daarna volgt de stand van zaken omtrent surveillance van gezondheidsproblemen in Nederland op basis van gegevens uit EMDs in de huisartspraktijk. Om prevalentie- en incidentiecijfers te kunnen berekenen op basis van deze dossiers, moeten de afzonderlijke diagnoses van een patiënt gegroepeerd worden in episodes. Een episode omvat alle diagnoses die betrekking hebben op hetzelfde gezondheidsprobleem. De huisarts heeft bijvoorbeeld drie afzonderlijke, opeenvolgende diagnoses bij een bepaalde patiënt gesteld: hoesten, koorts en bronchitis, die samen één episode met de naam bronchitis vormen. Tot nu toe werden deze episodes grotendeels handmatig geconstrueerd.

Hoofdstuk 2 beschrijft de ontwikkeling van EPICON, een applicatie om diagnoses uit EMDs in de huisartspraktijk automatisch in episodes te groeperen. Deze episodes kunnen gebruikt worden om morbiditeitscijfers (prevalentie- en incidentiecijfers) te berekenen. Deze studie bouwt voort op een eerder, grootschalig onderzoek onder Nederlandse huisartsen: de Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk (NS2). In dit eerdere onderzoek is een semi-automatische methode ontwikkeld om diagnoses uit EMDs in de huisartsenpraktijk in episodes te groeperen. Hierbij worden automatisch episodes gegenereerd voor makkelijk te groeperen diagnoses, terwijl moeilijk te groeperen diagnoses handmatig worden gegroepeerd. We hebben EPICON ontworpen als een uitbreiding van deze semi-automatische methode. De moeilijk te groeperen diagnoses worden binnen EPICON automatisch gegroepeerd met behulp van analoog ofwel 'case-based' redeneren, een methode die afkomstig is uit het vakgebied van de kunstmatige intelligentie. Deze vorm van redeneren veronderstelt dat een nieuw probleem opgelost kan worden door gebruik te maken van oplossingen voor eerdere, soortgelijke problemen, die in het geheugen zijn opgeslagen (cases). EPICON gebruikt de handmatig gegroepeerde diagnoses uit NS2 als 'cases' om diagnoses uit een nieuwe dataset automatisch in episodes te groeperen. Een voorlopige evaluatie op basis van misclassificaties suggereerde dat EPICON voldoende presteerde om morbiditeitscijfers op te baseren.

Om te beoordelen of deze applicatie daadwerkelijk in gebruik genomen kon worden, moest EPICON worden geëvalueerd ten aanzien van het doel waar het voor gebruikt zou gaan worden: het schatten van morbiditeitscijfers op basis van EMDs in de huisartspraktijk. We hebben twee studies uitgevoerd om EPICON op dit punt te evalueren. In de eerste evaluatie, beschreven in *hoofdstuk 3*, bestudeerden we de interne validiteit van EPICON, dat wil zeggen dat we de prestatie van EPICON onderzochten in een steekproef uit de dataset die gebruikt is om EPICON te ontwikkelen (de NS2-dataset). Uit de resultaten bleek dat de morbiditeitscijfers op basis van EPICON over het algemeen weinig afweken van de gehanteerde gouden standaard (de NS2-cijfers). Hieruit concludeerden we dat de interne validiteit van EPICON voldoende was voor het schatten van morbiditeitscijfers op basis van EMDs in de huisartspraktijk.

Hoofdstuk 4 beschrijft de tweede evaluatie: een studie naar de externe validiteit (generaliseerbaarheid) van EPICON. In dit onderzoek bestudeerden we de prestatie van EPICON in een onafhankelijke dataset, waarvan de gegevens door huisartsen waren vastgelegd in episode-georiënteerde EMDs. In deze nieuwe generatie van EMDs groeperen huisartsen de diagnoses van een patiënt handmatig in episodes in het EMD van die patiënt. Deze episode-georiënteerde EMDs wordt momenteel geïmplementeerd in een aantal Nederlandse huisartspraktijken. In dit onderzoek werden de handmatig gegroepeerde diagnoses ontkoppeld en opnieuw gegroepeerd door EPICON. Vervolgens vergeleken we de morbiditeitscijfers op basis van EPICON met de cijfers op basis van de handmatig gegroepeerde episodes, die we in deze studie als gouden standaard beschouwden. EPICON bleek goed te presteren voor het overgrote deel van de morbiditeitscijfers. Een klein deel van de morbiditeitscijfers (5%) liet echter structurele of grote afwijkingen zien. Uit dit onderzoek concludeerden we dat de externe validiteit van EPICON voldoende is, maar dat voorzichtigheid in acht genomen moet worden wanneer EPICON gebruikt zou worden voor morbiditeitscijfers van algemene of van weinig specifieke gezondheidsproblemen.

Op basis van deze resultaten werd EPICON in gebruik genomen om episodes te construeren voor een landelijk representatieve database met gegevens uit EMDs in de huisartspraktijk (de database van LINH, het Landelijk Informatie Netwerk Huisartsenzorg). Deze implementatie maakte het mogelijk om routinematig gezondheidsproblemen van de Nederlandse bevolking te monitoren op basis van gegevens uit EMDs in de huisartspraktijk. Hiervoor was een algemene surveillance strategie nodig, die bruikbaar was voor de gehele dataset. *Hoofdstuk 5* beschrijft deze algemene strategie, die de constructie van episodes, de analyse van gegevens en de interpretatie van de resultaten omvat.

De **eerste stap** was het construeren van episodes. De landelijke database bevat zowel gegevens uit de gangbare (contact-georiënteerde) EMDs als uit de nieuwe episode-georiënteerde EMDs. Voor contact-georienteerde EMDs werden episodes door EPICON gemaakt; voor episode-georienteerde EMDs werden de episodes gebruikt die de huisartsen geregistreerd hadden.

De **tweede stap** was het analyseren van de gegevens. Een belangrijke vraag was hoe de grote hoeveelheid beschikbare data het beste gehanteerd kon worden. Er waren gegevens beschikbaar over zo'n 680 verschillende gezondheidsproblemen over meerdere jaren en het was praktisch niet haalbaar om voor elk gezondheidsprobleem een aparte analyse op te zetten. Bovendien moesten we bij het opzetten van de analyses rekening houden met de clustering van gegevens binnen patiënten en binnen praktijken.

Daarom hebben we een algemeen, statistisch model ontworpen dat rekening houdt met clustering, leeftijd en geslacht en dat toegepast kan worden op elk gezondheidsprobleem (een multilevel regressie model).

De **derde stap** was het interpreteren van de resultaten uit deze analyses. Hierbij moesten we rekening houden met allerlei factoren die van invloed kunnen zijn op het vastleggen van gegevens in EMDs van huisartsen. Daarom maakten we een lijst van alle mogelijke oorzaken voor veranderingen in morbiditeitscijfers, die we onderverdeelden in vier categorieën: veranderingen in registratie, veranderingen in de vraag naar zorg, veranderingen in het aanbod van zorg en 'echte' veranderingen in het aantal nieuwe ziektegevallen.

We hebben deze algemene surveillance strategie toegepast om gezondheidsproblemen in Nederland te monitoren over de periode 2002-2005. In dit onderzoek ontdekten we verschillende opvallende trends in de tijd. Zo bleek dat de bezorgdheid van patiënten over (bij)werkingen van geneesmiddelen was toegenomen, met name onder patiënten van 45 jaar en ouder. Dit zou verklaard kunnen worden door de toegenomen beschikbaarheid en gebruik van zelfzorgmedicatie in deze periode. Ook vonden we een toename van verschillende urogenitale problemen, waaronder prostaatproblemen, angst voor geslachtsziektes en plasproblemen. Mogelijke verklaringen voor de gevonden toename van urogenitale problemen zijn: de beschikbaarheid van en reclame voor nieuwe geneesmiddelen voor goedaardige prostaatvergroting, een toename van het aantal geslachtsziektes en een toename van het aantal urineweginfecties. Een andere bevinding uit deze studie was dat de incidentie van vergrote en herhaaldelijk ontstoken amandelen was afgenomen.

De mogelijke oorzaken voor deze daling in vergrote/herhaaldelijk ontstoken amandelen bij kinderen is nader onderzocht in *hoofdstuk 6*, waarin we een follow-up strategie beschrijven om aanwijzingen te vinden voor mogelijke oorzaken van een ontdekte trend. In deze **vierde stap** formuleerden we specifieke hypotheses aan de hand van de lijst met mogelijke oorzaken voor veranderingen in morbiditeitscijfers. Deze hypotheses onderzochten we op basis van de beschikbare informatie in de database, waaronder gegevens over consulten, verwijzingen en voorschriften voor geneesmiddelen. We construeerden verklarende variabelen, die we stapsgewijs aan het algemene statistische model toevoegden. Met deze vierde stap kan onderzocht worden of een gevonden trend veroorzaakt wordt door 'artefacten' zoals een verandering in registratie of dat er mogelijk sprake is van een 'echte' verandering in de incidentie van een ziekte. Dit follow-up onderzoek vond plaats onder kinderen van 0 tot 14 jaar: de leeftijdsgroep waarin vergrote en herhaaldelijk ontstoken amandelen vaak voorkomt. Uit de resultaten bleek dat de daling van dit gezondheidsprobleem deels verklaard kon worden door veranderingen in de vraag naar zorg (minder consultaties) en in het aanbod van zorg (minder verwijzingen en voorschriften). Voor de resterende daling kon echter geen verklaring gevonden worden. Deze afname is mogelijk gerelateerd aan een 'echte' daling in bovenste luchtweginfecties.

In *hoofdstuk 7* geven we eerst een overzicht van de belangrijkste bevindingen van dit proefschrift. Vervolgens bespreken we een aantal methodologische kwesties. De surveillance strategie is ontwikkeld vanuit een brede aanpak, waarin elementen uit verschillende vakgebieden met elkaar gecombineerd worden. De keerzijde van deze brede aanpak was dat we niet uitvoerig op alle onderdelen in konden gaan. We bespreken ook de voor- en nadelen van de gebruikte database. Belangrijke voordelen zijn de representativiteit voor de Nederlandse bevolking, de grote hoeveelheid data en het feit dat de gegevens al zijn verzameld. Een nadeel is de beperkte mogelijkheid om controle uit te oefenen op de kwaliteit van de verslaglegging. Daarna komen de toepassingsmogelijkheden van de ontwikkelde surveillance strategie aan de orde. EPICON kan, onder bepaalde condities, toegepast worden in andere huisartsregistraties. Ook kan EPICON toegepast worden als een standaard waarmee andere episode-constructies kunnen worden vergeleken. Wanneer we de totale surveillance strategie beschouwen, dan leent deze zich met name voor toepassing in andere huisartsregistraties en niet zozeer voor toepassing in andere disciplines. Ten slotte bespreken we de implicaties van dit proefschrift voor de volksgezondheidszorg, de huisartspraktijk en toekomstig onderzoek. Dit proefschrift laat zien dat gegevens uit EMDs in de huisartspraktijk waardevol zijn voor surveillance op voorwaarde dat er methoden beschikbaar zijn om deze gegevens te ontsluiten en te gebruiken. Toekomstige surveillance zou zich moeten richten op de verdere ontwikkeling van methoden voor het secundair gebruik van gegevens uit EMDs. Hierbij kan bijvoorbeeld gedacht worden aan methoden om de kwaliteit van de verslaglegging te verbeteren of aan het koppelen van gegevens uit EMDs aan relevante informatie uit andere databases.

DANKWOORD

Met de voltooiing van dit proefschrift sluit ik een periode af waarin ik niet alleen veel heb geleerd, maar ook veel hulp en steun heb gekregen van een groot aantal mensen. Op deze plek wil ik iedereen bedanken die geholpen heeft bij de totstandkoming van dit proefschrift. Daarnaast wil ik graag een aantal mensen in het bijzonder noemen.

Om te beginnen wil ik alle huisartsen en praktijkmedewerkers bedanken die de gegevens zorgvuldig hebben vastgelegd in elektronische medische dossiers. Zonder al deze waardevolle gegevens had ik dit onderzoek immers niet kunnen uitvoeren.

Mijn promotoren, Pieter de Vries Robbé en Gerhard Zielhuis, wil ik hartelijk danken voor de deskundige begeleiding en de fijne samenwerking. Pieter, van jou heb ik veel geleerd over de informatiekundige kant van het onderzoek. Onze vele gesprekken hebben me geholpen om op dit terrein een visie te ontwikkelen. Daarnaast vond ik het bijzonder prettig dat ik altijd bij je terecht kon: je was steevast betrokken en bereid om naar me te luisteren en mee te denken. Gerhard, aan jou had ik een zeer deskundig begeleider voor de epidemiologische kant van het onderzoek. Ik vond het een voorrecht om met je te mogen samenwerken en ik heb erg veel van je geleerd. Je vermogen om tijdig de juiste beslissingen te nemen en helder prioriteiten te stellen heeft er zeker toe bijgedragen dat ik het proefschrift binnen de daarvoor gestelde tijd kon afronden. Ik verheug me op verdere samenwerking met jullie.

Dinny de Bakker, mijn co-promotor, wil ik allereerst bedanken voor zijn onvoorwaardelijke steun en vertrouwen in de goede afloop van dit project. Toen ik bij je kwam met het idee om de episodeconstructie te automatiseren (pas veel later, bij het schrijven van het eerste artikel, bedachten we de naam EPICON) was je direct bereid om dat te faciliteren. Bovendien heb je over de volle breedte (van case bases tot multilevel analyes) een inhoudelijke bijdrage geleverd aan dit proefschrift.

Tijdens dit onderzoek heb ik nauw samengewerkt met een aantal mensen van het NIVEL, die allen een belangrijke rol hebben gespeeld bij de totstandkoming van dit proefschrift: Robert Verheij (projectleider LINH), Jan Gravestein (programmeur EPICON) en Peter Spreeuwenberg (statisticus). Robert, veel dank voor de goede inbreng en de fijne samenwerking. En voor je gevoel humor; op de een of andere manier moesten we altijd wel een keer hard lachen tijdens een bijeenkomst, waardoor het dan even een klein feestje werd. Jan, wij hebben samen aan de wieg gestaan van EPICON. Ik wil je hartelijk danken voor het meedenken en voor je enthousiasme en inzet, wat ervoor zorgde dat het project niet alleen vaart, maar ook een 'flow' kreeg. Peter, veel dank voor je deskundige inbreng; ik heb veel van je geleerd over multilevel analyes en alles wat daarbij komt kijken. Ik verheug me op het samen verder uitbouwen en implementeren van de methoden die in dit proefschrift zijn ontwikkeld.

Geert Elbers en Ellen Theuns-Lamers, twee geneeskundestudenten die ik begeleidde bij hun afstudeerproject, wil ik hartelijk danken voor hun inzet, bijdrage en de leuke samenwerking. Bij een van deze onderzoeken hebben we dankbaar gebruik gemaakt van de gegevens van de RNG-praktijken. Willem Jan van der Veen, bedankt voor zowel het samenstellen van dit bestand als je inhoudelijke inbreng in dit onderzoek. Ook Alie Hiddema en de andere huisartsen van de RNG-praktijken: bedankt! In het andere onderzoek was Hans van der Wouden van de afdeling huisartsgeneeskunde in Rotterdam betrokken. Hans, jij ook bedankt!

Op deze plek wil ik ook een aantal andere mensen bedanken die bij NS2 of LINH betrokken waren of zijn. François Schellevis en Michiel van der Linden hebben samen de basis gelegd voor de NS2 episodeconstructie, waar een belangrijk deel van dit proefschrift op is gebaseerd. François, daarnaast wil ik je ook hartelijk danken voor je deskundige bijdrage aan de interpretatie van de incidentiecijfers. Ook wil ik Jan van der Steen, Han van Snellenberg, Tjeerd van Althuis en Waling Tiersma bedanken voor hun medewerking aan dit proefschrift.

Daarnaast wil ik een aantal deskundigen uit andere huisartsregistraties bedanken. Eloy van de Lisdonk en Toine Lagro-Janssen van de Continue Morbiditeitsregistratie Nijmegen: bedankt voor jullie inbreng bij de interpretatie van de incidentiecijfers. Furthermore, I want to thank Douglas Fleming, director of the Birmingham Research Unit of the Royal College of General Practitioners for his assistance with the comparison of LINH rates with the Weekly Return Services rates.

In dit project heb ik ook dankbaar gebruik gemaakt van de deskundige adviezen van Ameen Abu Hanna en Niels Piek omtrent het opzetten en evalueren van een case based systeem. Ameen en Niels, bedankt! Verder wil ik Steven Teerenstra bedanken voor zijn heldere uitleg en advies bij de multilevel analyses.

Mijn (ex)collega's op de afdeling medische informatiekunde wil ik bedanken voor de interessante discussies en de gezelligheid. Arjen, Wouter en Olga: met jullie heb ik het nodige promotie lief en leed gedeeld. Dank daarvoor! Tot deze laatste categorie behoren ook Marijn Brouwers en Marian Rombouts: bedankt voor alle steun en gezelligheid en vooral veel sterkte met de laatste loodjes!

Mijn lieve (schoon)familie en vrienden en vriendinnen wil ik bedanken voor hun belangstelling en alle vrolijkheid en ontspanning naast de promotie. Lieve Els en Virgie, wat fijn dat jullie op deze bijzondere dag mijn paranimfen willen zijn! Dit proefschrift was niet mogelijk geweest zonder alle mensen die het mij mogelijk maken om werk en gezin te combineren. Carla en Auk, hartelijk dank voor de jarenlange gezellige uitwisseling en zorg voor onze jongens. Lieve Gerard en Riet, wat is het heerlijk dat jullie altijd klaarstaan om op te passen en ondertussen ook nog zoveel klusjes voor ons doen, ik kan me geen betere schoonouders wensen. Mijn ouders, mijn zus en ook Hans en Marc wil ik hartelijk bedanken voor alle warme belangstelling en steun.

Lieve Pyke en Misha, bedankt voor alle keren dat jullie mij van mijn werk hebben gehouden! Jullie zorgen voor de nodige balans in mijn leven, wat is het fijn om bij jullie te zijn! En tenslotte, mijn lieve Walter, er zijn ontzettend veel kleine en grote dingen waar ik je dankbaar voor ben. Dit proefschrift heeft, vooral het afgelopen jaar, veel beslag gelegd op onze kostbare tijd samen. Bedankt voor al je liefde en steun (en voor alle goede, relativerende grappen over mijn proefschrift)! Ik draag dit boekje op aan jou.

ABOUT THE AUTHOR

Marion Biermans was born on March 16, 1967 in Eindhoven, the Netherlands. In 1986 she started Psychology studies at the Radboud University Nijmegen; during the last year of this study she worked as a research trainee at the March of Dimes Birth Defects Foundation in New York. She graduated in 1992 and then started work as research assistant at the department of Clinical Psychology and Personality at the Radboud University Nijmegen. She had several jobs between 1993 and 1997, including work in the field of computer education. In 1997 she traveled in Asia and decided to resume her scientific career after her return to the Netherlands. In 1998 she studied Epidemiology (postdoctoral education) at the Free University in Amsterdam and graduated that same year. Between 1999 and 2002 she was employed as a researcher at the Netherlands Institute for Health Services Research (NIVEL). In 2003 she joined the Department of Medical Informatics at the Radboud University Nijmegen Medical Centre, where she conducted the research described in this thesis. Since July 2008 she has been employed at that same institute in the department of Primary and Community Care, where she is involved in an inter-university project (the String-of-pearls initiative) aimed at the development of a joint IT infrastructure to collect and access patient data and biomaterials.

The gecko symbolizes a hidden pattern in the data; it refers to extracting useful information from a large dataset.