



Postprint Version	1.0
Journal website	http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T7S-4PP200C-1&_user=1421770&_coverDate=09%2F17%2F2007&_rdoc=1&_fmt=&_orig=search&_sort=d&_view=c&_acct=C000052679&_version=1&_urlVersion=0&_urlid=1421770&_md5=abf6d0edc4c4debece27e9562dcbfc27
Pubmed link	http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubmed&Cmd=ShowDetailView&TermToSearch=17870659&ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum
DOI	10.1016/j.ijmedinf.2007.08.002

This is a NIVEL certified Post Print, more info at <http://www.nivel.eu>

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

MARION C.J. BIERMANS^A, DINNY H. DE BAKKER^B, ROBERT A. VERHEIJ^B, JAN V. GRAVESTIJN^B, MICHIEL W. VAN DER LINDEN^B AND PIETER F. DE VRIES ROBBÉ^A

^aDepartment of Medical Informatics, Radboud University Nijmegen Medical Centre, 152 MI, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

^bNetherlands Institute for Health Services Research, Utrecht, The Netherlands

Received 30 May 2006; revised 3 August 2007; accepted 5 August 2007. Available online 17 September 2007.

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 EPICONs performance will probably be adequate for the purpose of estimating morbidity rates in general practice.

1. INTRODUCTION

1.1. 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.

1.2. Previous research

Our project builds upon the second Dutch National Survey of General Practice (DNSGP-2), which has been described elsewhere [6] and [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.

1.3. 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], [10], [11], [12], [13] and [14].

1.4. 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.

2. METHODS

2.1. 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] and [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 1 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.

2.2. 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] and [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.

2.3. 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 1 shows examples of these three types.

[TABLE 1].

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 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 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 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).

3. RESULTS

3.1. Development phase

3.1.1. 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. Fig. 1 shows this method, which consisted of five steps (shown in between parentheses).

[FIGURE 1]

- *Step 1.* The first consultation diagnosis of a patient in a 1-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 Section 2.1), 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). 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].

- *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.

3.1.2. 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.

3.1.3. Design

In line with the requirements, we designed the fully computerised procedure as an extension of the semi-computerised method. Fig. 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.

[FIGURE 2]

The black parts in Fig. 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 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 3.]

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 follows: 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.

3.1.3.1. 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 $\geq 70\%$ in case base 1 (see Table 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 $\geq 48\%$ for case base 1, $\geq 42\%$ for case base 2, and $\geq 71\%$ for case base 3.

3.2. Evaluation phase

Table 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 4].

Fig. 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.

[FIGURE 3]

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.

We also examined the number of misclassified diagnoses per ICPC-chapter. Fig. 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.

[FIGURE 4]

4. 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], [23] and [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 computerised 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.

Account of what this research has learned

What was known before the study:

- Data from electronic medical records (EMRs), recorded in routine care in general practice, are useful, but not directly suitable for epidemiologic research.
- Prevalence and incidence rates can be estimated by grouping diagnoses into episodes of care.
- Diagnoses can be grouped into episodes of care by using a semi-computerised method.
- Case-based reasoning is a promising method for classification in “weak theory” domains, such as medicine.

What the study has added to the body of knowledge:

- Development of a fully computerised method can disclose data from EMRs in general practice for estimating prevalence and incidence rates.
- Insight into the relation between diagnoses, episodes of care, and prevalence and incidence rates.
- A case-based method for grouping diagnoses from EMRs in general practice into episodes of care.
- A formative evaluation method for a case-based classification system.

TABLES AND FIGURES

Table 1 – Examples of misclassified diagnoses

Patient	Diagnosis ^a (ICPC)	Episode numbers			Misclassifications	
		Semi-computerised 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 combination

^a 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.



Fig. 1 – Semi-computerised grouping method.

Table 2 – Decision table for linking an ongoing diagnosis to a previous^a 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)

^a 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.

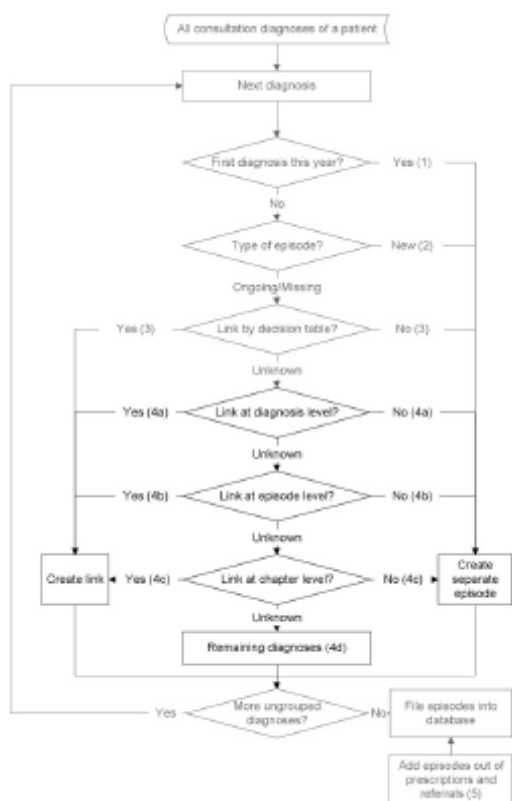


Fig. 2 – Fully computerised grouping method.

Table 3 – Case base 1: cases at diagnosis-level^a

Manually grouped diagnosis (ICPC-code)	Other diagnosis of the same patient (ICPC)	Total number 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)

^a 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).

Table 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 number of diagnoses	1,201,234	1,201,234

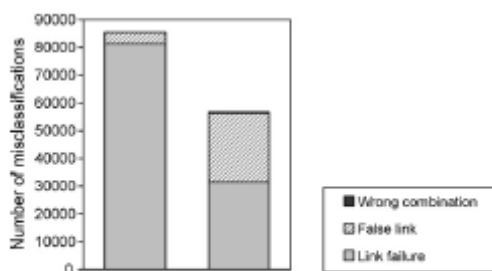


Fig. 3 – Type of misclassification in EPI-0 and EPICON.

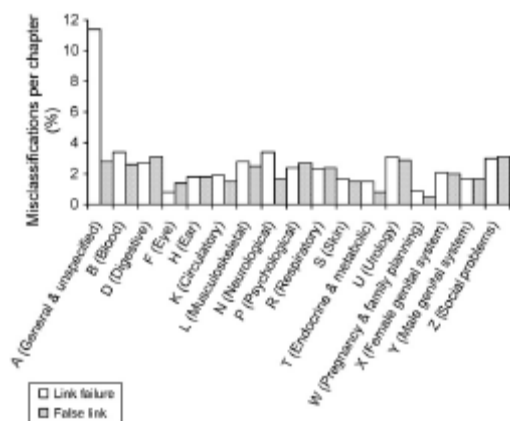


Fig. 4- Distribution of misclassifications across ICPC-chapters in EPICON.

REFERENCES

- [1] Classification Committee WONCA, An international glossary for general/family practice, *Fam. Pract.* **12** (1995), pp. 341–369.
- [2] L.L. Weed, Medical records that guide and teach, *N. Engl. J. Med.* **278** (1968), pp. 593–600.
- [3] I.M. Carey, D.G. Cook, S. De Wilde, S.A. Bremner, N. Richards, S. Caine, D.P. Strachan and S.R. Hilton, Developing a large electronic primary care database (doctors' independent network) for research, *Int. J. Med. Inform.* **73** (2004), pp. 443–453.
- [4] M. Lauteslager, H.J. Brouwer, J. Mohrs, P.J.E. Bindels and H.G.L.M. Grundmeijer, The patient as a source to improve the medical record, *Fam. Pract.* **19** (2002), pp. 167–171.
- [5] I.M. Okkes, A. Groen, S.K. Oskam and H. Lamberts, Advantages of long observation in episode-oriented electronic patient records in family practice, *Methods Inf. Med.* **40** (2001), pp. 229–235.
- [6] In: G.P. Westert, L. Jabaaij and F.G. Schellevis, Editors, *Morbidity, Performance and Quality in Primary Care: Dutch General Practice on Stage*, Radcliffe Publishing, Oxon (2006).
- [7] G.P. Westert, F.G. Schellevis, D.H. De Bakker, P.P. Groenewegen, J.M. Bensing and J. Van der Zee, Monitoring health inequalities through general practice: the second Dutch national survey of general practice, *Eur. J. Public Health* **15** (2005), pp. 59–65.
- [8] G.F. Luger and W.A. Stubblefield, Knowledge-intensive problem solving, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Addison Wesley Longman Inc., Reading (1998) pp. 207–246.
- [9] A. Aamodt and E. Plaza, Case-based reasoning: foundational issues, methodological variations, and system approaches, *AI Commun.* **7** (1994), pp. 39–59.
- [10] I. Bichindaritz, Case-based reasoning in the health sciences, *Artif. Intell. Med.* **36** (2005), pp. 121–125.
- [11] J.L. Kolodner, *Case-based Reasoning*, Morgan Kaufmann, San Mateo (1993).
- [12] R.T. Macura and K. Macura, Case-based reasoning: opportunities and applications in health care, *Artif. Intell. Med.* **9** (1997), pp. 1–4.
- [13] I. Watson and F. Marir, Case-based reasoning: a review, *Knowl. Eng. Rev.* **9** (1994), pp. 355–381.
- [14] R. Schmidt, S. Montani, R. Bellazzi, L. Portinale and L. Gierl, Case-based reasoning for medical knowledge-based systems, *Int. J. Med. Inform.* **64** (2001), pp. 355–367.
- [15] M.A.J.B. Tacken, Quality of preventive performance in general practice: the use of routinely collected data, Doctoral Dissertation, WOK, Radboud University Nijmegen, 2005.
- [16] R.A. Verheij, L. Jabaaij, H. Abrahamse, H. Van den Hoogen, J. Braspenning, T. Van Althuis, Netherlands information network of general practice: facts and figures on Dutch GP care, <http://www.linh.nl/>
- [17] M.W. Van der Linden, G.P. Westert, D.H. De Bakker, F.G. Schellevis, The Second Dutch National Survey of General Practice (NS2): Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Klachten en aandoeningen in de bevolking en in de huisartspraktijk (Health and Problems in the Dutch Population and in General Practice, in Dutch), NIVEL, RIVM, Bilthoven, Utrecht, 2004.
- [18] In: H. Lamberts and M. Woods, Editors, *International Classification of Primary Care (ICPC)*, Oxford University Press, Oxford (1987).



- [19] In: G.V. Nakamura, D.L. Medin and R. Taraban, Editors, *Categorization by Humans and Machines*, Academic Press, San Diego (1993).
- [20] R. Bareis, *Exemplar-based Knowledge Acquisition: A Unified Approach to Concept Representation, Classification and Learning*, Academic Press, Boston (1989).
- [21] Wet bescherming persoonsgegevens (Personal Data Protection Act, in Dutch), *Staatsblad*, 2000, p. 302.
- [22] A.K. Rosen and A. Mayer Oakes, Episodes of care: theoretical framework versus current operational realities, *Jt. Comm. J. Qual. Improv.* **25** (1999), pp. 111–128.
- [23] R.Y. Son, R.K. Taira, A.A.T. Bui, H. Kangaroo and A.F. Cárdenas, A context-sensitive methodology for automatic episode creation, *Proc. AMIA Symp.* (2002), pp. 707–711.
- [24] T.D. Wingert, J.E. Kralewski, T.J. Lindquist and D.J. Knutson, Constructing episodes of care from encounter and claims data: some methodological issues, *Inquiry* **32** (1995), pp. 430–443.
- [25] Anon., European physicians especially in Sweden, Netherlands and Denmark, lead U.S. in use of electronic medical records, *Health Care Res.* **2** (2002), pp. 1–3.