From the Department of Clinical Sciences, Family Medicine Stockholm, Karolinska Institutet, Stockholm and the Department of Biomedical Engineering, Division of Medical Informatics, Linköpings universitet, Linköping, Sweden

Classification and Reuse of Clinical Information in General Practice

Studies on diagnostic and pharmacological information in electronic patient record systems

Gunnar Nilsson

Stockholm 2002

All previously published papers were reproduced with permission from the publisher.

To Sophia, Patrik, Frida and Viktor

Abstract

The reuse of clinical information from the electronic patient record (EPR) for all health care areas – administrative, clinical, teaching and research – will constitute a challenge in coming decades. The classification and the coding of clinical information are essential for such data retrieval.

The prevalence, characteristics and management of patients with chronic heart failure were investigated in a four-year retrospective study including 46 general practitioners (GPs) (study I). Out of the registered population, 667 (0.7%) patients had chronic heart failure. About 99% had some chronic co-morbidity, and 27% were treated with angiotensin converting enzyme inhibitors. During the study period 66% had an echocardiograph. Gender differences were found concerning co-morbidity and management. It seemed that these patients could be more adequately managed, and the EPRs were feasible for research.

A working model for continuous medical education based on feedback on prescribing rates and educational outreach visits was evaluated in a randomised study, with three parallel intervention groups/fields, including 40 GPs (study II). In the hypertension field desired trends were recorded, with a significant effect on prescriptions for agents acting on the renin-angiotensin system. In the peptic ulcer/dyspepsia field desired trends were recorded, and in the depression group only minor changes. The model was promising, and the EPRs were feasible for retrieval of prescription data.

The Swedish primary health care version of International Statistical Classification of Diseases and Related Health Problems (ICD-10) was reclassified in a preceding study using a concept system approach, and implemented in a web-based browser. In a randomised experimental diagnostic coding trial, six GPs coded each of 152 medical problems in patient vignettes by means of three versions of ICD-10, one with a compositional structure (study III). At code level the reliability was poor and almost identical when comparing the three versions. At aggregated level the reliability was good, and it was somewhat better in the compositional structure.

Necessary conditions for the establishment of a database of diseases and health problems based on EPRs were investigated in a postal questionnaire study including 300 randomly selected GPs in Sweden (study IV). A total of 184 (61%) answered, and 92% used an EPR system. About 88% used an ICD based classification, and the classification in use was computerised for 74%. About 76% reported classifying at least one symptom or disease per encounter. Classification of diseases was considered important for follow-up by 83%. EPRs provide several fundamentals for a database of diagnostic data.

The textual content, health problems and diagnostic codes in EPRs were investigated in a retrospective database study including 20 randomly selected GPs and a total of 400 records (study V). About 15% of the records were problem-oriented. The mean number of words per record was 99.4. The mean number of problems managed per record was 1.3 and there were on average 1.1 diagnostic codes per record. The proportion of correct codes was estimated at 97%. The most frequent disease was essential hypertension (8%). It seemed that problem-oriented EPR systems enforced the coding activities.

In conclusion, data from EPRs in general practice was found feasible for research, education, and the establishment of a database of diagnostic data. Positive aspects of the reliability of diagnostic coding were found in new computerised versions of a traditional classification of diseases.

Key words: general practice, electronic patient record, congestive heart failure, data retrieval, prescribing, continuing medical education, classification, coding, reliability, ICD, medical modeling, postal questionnaire.

List of Publications

This thesis is based on the following papers, which will be referred to by their Roman numerals.

I.   Nilsson G, Strender L-E.
       Management of heart failure in primary health care. A retrospective study on electronic patient records in a registered population.
       Scandinavian Journal of Primary Health Care 2002;20(3):161-165.

II.   Nilsson G, Hjemdahl P, Hässler A, Vitols S, Wallén NH, Krakau I.
       Feedback on prescribing rate combined with problem-oriented pharmacotherapy education as a model to improve prescribing behaviour among general practitioners.
       European Journal of Clinical Pharmacology 2001;56(11):843-8.

III.   Nilsson G, Petersson H, Åhlfeldt H, Strender L-E.
       Evaluation of three Swedish ICD-10 primary care versions: reliability and ease of use in diagnostic coding.
       Methods of Information in Medicine 2000;39(4-5):325-31.

IV.   Nilsson G, Åhlfeldt H, Strender L-E.
       Computerisation, coding, data retrieval and related attitudes among Swedish general practitioners – a survey of necessary conditions for a database of diseases and health problems.
       International Journal of Medical Informatics 2002;65(2):135-43

V.   Nilsson G, Åhlfeldt H, Strender L-E.
       Textual content, health problems and diagnostic codes in electronic patient records in general practice.
       Submitted manuscript.

Contents

Errata in Published Papers

Paper II, page 846, in Fig. 1: *ACEIs* should be read *ACEIs/ARBs*.

Paper III, page 331: References 5 and 7 should change place.

Paper IV, in abstract, line 12, and on page 136, column 2, line 26: *972* should be read *973*.

List of Abbreviations

ACEI	Angiotensin converting enzyme inhibitor
ATC	Anatomic therapeutic chemical (classification system)
CHF	Chronic heart failure
CI	Confidence interval
DDD	Defined daily dose
EPR	Electronic patient record
GP	General practitioner
ICD	International (statistical) classification of diseases (and related health problems)
ICPC	International classification of primary care
K	Cohen's Kappa
K_G	Generalized Kappa
KSH97P	Klassifikation av sjukdomar och hälsoproblem 1997 primärvård (Classification of diseases and related health problems 1997 primary health care)
PHC	Primary health care
SNOMED	Systematized nomenclature of medicine
SOAP	Subjective, objective, assessment (analysis), plan (procedures)

Introduction

This thesis addresses the broad field of clinical information management in primary health care (PHC). The main scientific fields involved in the studies are PHC and medical informatics, including aspects of quality of health care and medical terminology. The principal methodologies involved in the studies are data retrieval, quality assurance, professional education, concept modeling and statistical analysis. The focus is on general practice and issues related to the electronic patient record (EPR).

Medical Informatics

The domain of medical informatics is very broad (van Bemmel 1997). It is defined as "the scientific field that deals with the storage, retrieval, and optimal use of biomedical information, data, knowledge for problem solving and decision making" (Shortliffle 1990). This field represents the link between medicine and information science, and its empirical objects are those of both medicine and information science (Seelos 1992). It is the rational study of the way we think about patients, and the way that treatments are defined, selected and evolved (Coiera 1997). The tools of informatics are clinical guidelines, formal medical languages, information systems and communication systems such as the Internet, rather than computers (Coiera 1997). The agenda for medical informatics should address improved quality of public health and possibly lower costs through implementation of evidence based health care.

Use and Reuse of Clinical Information

A major change in the next decade will centre on gathering, managing, and using clinical information (e.g. in the EPR systems) for all health care areas – administrative, clinical, teaching and research. The development of a new information management paradigm is a critical challenge facing the health care sector (Lorenzi 1995, Stead 1999). This includes development in areas such as structures for representing data and knowledge (so that complex relationships may be visualised), methods for acquisition and presentation of data (so that overload can be avoided), and management of changes related to people and information technology (so that the use of information is optimised). Three functions are of central importance regarding the use and reuse of clinical information: a personal clinical health record (one record for each patient), a health care data bank (e.g. a national database with interventions related to health problems and diagnoses), and

a health care knowledge base (including, for example, practice guidelines and decision support systems) (Stead 1999). Health care services and cost-effectiveness may be greatly affected as a result of attaining these goals. There is a need to evaluate the feasibility and possible effects of these proposed new functions.

Research on basic aspects of computers in PHC has been reported during the past decade. The attitudes of general practitioners (GP) towards the computerised prescribing of drugs have become more positive (Bolton 1999), and there is a high level of interest in learning various practice-related applications (Cook 1998).

In PHC it is possible for the EPR to support communication over organisational as well as professional borders, facilitating patient-centred care of patients. In such a vision, in a multidisciplinary perspective, a personal health record (i.e. only one record for each patient) is implicit. A generally accepted structure for patient records is the Subjective-Objective-Assessment-Plan (SOAP) structure. In addition, the problem-oriented medical record is considered to be a useful structure for records in PHC. The problem list in such a record and the terminology on which it is based are also of critical importance (Campbell 1998).

The reuse of data from EPRs is challenging, as much money is being spent on EPR systems that today are mainly used in the clinical care of patients and that offer little or nothing in terms of the future reuse of data. The EPR systems have the potential to significantly improve the way that clinical information is handled and clinical medicine is practised (Sujansky 1998). The advantages of computerised information systems in PHC with respect to obtaining epidemiological information and conducting audits have been demonstrated (Johnson 1991, Linnarsson 1993, Kinn 1995, Metsemakers 1996, Njálsson 1996). There are, however, several obstacles to the reuse of clinical data from EPRs (Crombie 1992, Hart 1999, Rector 2000, Grimsmo 2001). If this computer-stored clinical information is to be useful for all health care areas (clinical, research, teaching and administrative), the accuracy of data (correctness and completeness) in EPRs is of vital significance.

Terminology Issues in Health Care

Classification of Clinical Information

Identifying problems and diagnoses is vital in the medical problem-solving process, it constitutes a basis for decision-making, and it strongly influences the outcome of medical care. Proper classification of these clinical data is essential to the multipurpose reuse of information from EPRs (Rector 1991, Britt 1995), which

is mainly used to support quality assessment and research. Changes are currently taking place, with entry by coding staff now increasingly being done by healthcare professionals (Rector 2000). It seems to be of importance regarding the quality of the coding that doctors take an active part in the coding process (Earlam 1988). Due to the need for classification and coding, terminologies have become central elements in health information systems (Chute 1998). However, the major clinical classification in use today incompletely covers the clinical content of a patient record (Chute 1996). Today we also want coding and specialised terminologies to support information integration and messaging between health information systems.

Classification has two different meanings which must be differentiated (van Bemmel 1997). Firstly, there is the act or process of classifying, defined as "the coding of a description of an object by using codes or terms that are designators of the concept in a classification" (van Bemmel 1997). This is related to activities needed to assign an individual case to the right class and produce the right code in an efficient and reliable way. Problems with the process of classifying constitute a main concern of this thesis. Secondly, there is the process of designing a classification, which is only a partial concern of the thesis. Coding is defined as "the process of assigning an individual object or case to a class" (Marshall 1998). In this thesis, coding henceforth is used in the sense of 'classifying and coding'.

Coding is used mainly for two different purposes (Cimino 1996). Firstly, to abstract patient data, for example aggregated information from many patients into groups of diagnoses for the reporting of statistics. Secondly, to support coding of detailed clinical patient data in EPR systems in the care of an individual patient.

Information Accuracy

The accuracy of data in EPRs is critical to the optimal outcome of many health care activities. Ideally, the patient record consists of what clinicians have heard, seen, thought, and done (Rector 1991). Measuring and characterising the accuracy of data in EPRs and investigating ways to improve it has been declared imperative (Hogan 1997). Accuracy is calculated using two complementary measures: the correctness (the proportion of recorded observations that are correct compared to a gold standard – ideally the true state of the patient) and the completeness (the proportion of observations that are actually recorded). A few studies have recently been published on the accuracy of EPRs in general practice. The completeness and the correctness of EPRs were found to be at a high level for some chronic diseases (Nazareth 1993, Hassey 2001). The recording of data on acute illness and lifestyle

and socio-economic factors in general practice was found to be poor in one study in the UK (Ward 1994).

A correct diagnosis is important for patient care, and the validity (the extent to which an instrument measures what it is intended to measure, e.g. the extent to which a physician codes for the correct diagnosis of a patient) of physicians' diagnoses has been documented previously. In a recent study a completely correct diagnosis was found in only 43% of consultations in PHC (Ridderikhoff 1997). This is not surprising, as PHC deals with a blend of widely varying problems that are often in an early stage. Furthermore, the diagnosis is based on patient data such as symptoms and findings, which are also of limited validity (Gill 1973). The same range of diagnostic validity has been found in hospital settings (Zarling 1983, de Dombal 1991). The validity of codes concerns the process of making a correct diagnosis.

Problems with the process of coding concern the activities and support needed to assign an individual case to the right class and produce the right code in an efficient way with high reliability. If computer-stored information is to be useful for data retrieval, then the reliability (the extent to which the same measure will provide the same results under the same conditions, e.g. the extent to which two physicians code for the same diagnosis for a problem of a patient) of the patient data is of utmost importance (James 1991). This is due to the fact that information coded by different GPs is compared. Consequently, consistent use of the same code for the same clinical entity becomes important in order to facilitate data retrieval. Reliability is a useful measure when the true state of the patient is difficult to assess, or not well known. The reliability of coded diseases is, however, poor both in PHC and in hospital settings (Anderson 1980, Fortinsky 1981, Bridges-Webb 1986, Fisher 1992, Britt 1998, Dixon 1998).

Consequently, evaluation and improvement of the accuracy of coded clinical information have become important. The accuracy depends on a number of factors, such as coding schemes, coding tools, guidelines, routines, training, and attitudes towards information management as a whole. Evaluation and improvement of the accuracy of coded clinical information is therefore complex. Possible ways of improving the reliability of diagnostic coding are to improve coding routines, coding tools, and coding schemes. In addition, the physicians' motivation and the purpose of diagnostic coding are important factors to stress in order to improve the reliability. A model for evaluation of computerised coding systems has been proposed (Bolton 1997). This model includes evaluation of the structure of the coding system, the reliability of coding, the ease of use of the codes, the infrastructure supporting the coding system, and overall usefulness.

Terminologies and Coding Systems

The archetypal coding system for medical record abstraction is the International Classification of Diseases (ICD) (WHO 1993, Gersenovic 1995). Like most traditional classifications, the ICD is based on a strict hierarchy (one tree structure with chapters and subchapters), and problems recognised with such coding schemes are well described in the literature (Cimino 1996, Chute 1998, Rector 2000). Any one item can only be classified in one way, e.g. gout can be under musculoskeletal or metabolic disorders, but not under both. This is controlled by the priority rules in the ICD, which give infectious diseases (chapter I) and tumours (chapter II), for instance, higher priority than chapters related to organ systems (chapters III-XIV). Consequently, there have been compromises between chapters based on aetiology, anatomy and other grounds (Gersenovic 1995). The axes of classification are not consistent, e.g. at least thirteen different chapters must be combined to find all 'heart diseases' (Rector 2000). Adding finer grained details results in 'combinatorial explosion', e.g. adding a new qualifier to a set of codes doubles the number of codes. Further, much of the information in the ICD is contained only as language in the headings (e.g. chapter III, diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism).

The formal representation of diseases and health problems is in itself a complex chore (Cimino 1998, Flier 1998), and ambiguity is apparent when combining heterogeneous concepts such as location, symptoms, aetiology, syndrome, lesion, function, and process (Campbell 1979). Further, diseases and health problems are somewhat subjective descriptions of state of health, and may or may not include certain symptoms or signs.

Several common themes can be recognised in terminology development work (Cimino 1998), of which the following are partly addressed in this thesis:

- The content of the terminology – the content of the vocabulary and its coverage of particular domains.
- Concept orientation – when terms correspond to at least one and no more than one meaning.
- Multiple classification – allowing concepts to be classified according to more than one hierarchy.
- Multiple consistent views – multiple views of a terminology suitable for different purposes.
- Multiple granularity – when a terminology can be operated on multiple levels of details.

- Rejection of 'Not elsewhere classified' – terms that do not have a formal definition other than one of exclusion.

However, developing a truly concept-oriented terminology is not an easy task (Cimino 1998). Various approaches to solving these problems in medical (clinical) terminology are currently converging towards the use of a concept-based terminology with an underlying formal semantic terminological model (concept system) (Brown 2000). This approach is used to develop terminologies such as Systematized Nomenclature of Medicine (SNOMED) (Côté 1993), Clinical Terms Version 3 (formerly The Read Codes) (O'Neil 1995), the Logical Observation Identifier Names and Codes (LOINC) vocabulary (Huff 1998), the GALEN-project (Rector 1995), and SNOMED Clinical Terms (Price 2000). Terminologies based on a concept system can help us to more consistently represent clinical information. There is a need to investigate the effect of using a concept system approach to improve a traditional classification such as the ICD.

The development, use, and adaptation of concept systems have become central parts of all work in medical informatics (Musen 2001). Generally, models help us to better understand how our work and knowledge are organised, often exposing opportunities for simplification and identifying problems. In many respects concept systems therefore constitute a particularly interesting object of research. They could be regarded as systems consisting of several components (concepts) and their relations (concept relations). They are fundamental to information-based organisations and are especially important to specialists (Nuopponen 1994). Concept systems are of interest not only to terminology science but also, for example, to artificial intelligence, semantics, pedagogy, information science, classification theory, and database development.

Coding Tools

The accuracy of diagnostic coding also depends on the coding tools, and attention has therefore been directed toward the construction of computer-based tools for coding. For example, computerised coding tools have been found to save time and result in higher quality codes when medical concepts are extracted from narrative text from clinical cases (Hohnloser 1996). Coding, however, is time consuming, and much attention has also been directed toward the construction of computer-based tools for coding free-text reports (Spyns 1996). Natural language analysis methods have been found suitable for suggesting potentially correct codes from free text pathology reports (de Bruijn 1998). In addition, general-purpose natural language processing methods have been used to extract clinical data from narrative reports (Hripcsak 1995).

Application Areas

Primary Health Care

PHC is regarded as a fundamental constituent of the health care system in Sweden. General practice (family medicine) is the core medical speciality in PHC in Sweden, and the use of EPRs by GPs is almost universal (Spri 1995). PHC is provided by more than 4000 GPs in co-operation with district nurses and other health professionals at about 1400 health care centres. GPs carry out about 12.2 million consultations per year (Landstingsförbundet 1999). The health care system is undergoing rapid change, and we will continue to see changes involving organisation, new devices, new drugs and new techniques. The trend in health care is increasingly towards teamwork both within and between care units.

Terminologies in General Practice

Swedish general practice has a tradition of using a PHC version of the ICD as a controlled vocabulary in EPR systems in order to label diagnoses and health problems in encounters. This is in contrast to most other European countries, where the International Classification of Primary Care (ICPC) is used (Lamberts 1987). However, the extent to which the ICPC is used in everyday clinical practice has undergone little study. In some systems the ICD also supports a problem-oriented medical record (Weed 1968). In 1998, both a Swedish translation of the ICD-10 and a PHC version, Klassifikation av sjukdomar och hälsoproblem 1997 primärvård (KSH97P) (Socialstyrelsen 1997), were put into use. A PHC version of ICD-9 was used until the introduction of the KSH97P (Socialstyrelsen 1987).

The KSH97P was developed in 1997 by the Swedish Association of General Practice at the request of the National Board of Health and Social Welfare. The objective of the KSH97P is to reflect common diagnoses and symptoms within PHC. The main structure of the KSH97P is inherited from the ICD-10, i.e. the chapter structure reflecting the chosen principles of classification. The KSH97P uses 20 of the chapters of the ICD-10; chapter 20 is excluded due to the limited need for classification of injuries within the PHC sector. The total number of categories (disparate diagnoses) is 973. As the ICD has developed as a practical rather than a theoretical classification (for the single purpose of international statistics), it contains barriers to the re-use of data for the multiple purposes described above. The need has arisen for a more flexible classification system reflecting useful levels of diagnoses and health problems in PHC settings, and a more adequate representation system has been emphasised. Users may wish to be able to access diagnoses based on location or on aetiology in a poly-hierarchic system (Rector 1996).

Drugs are classified according to the Anatomical Therapeutic Chemical (ATC) classification system (WHO 1996). Drugs are automatically coded in all current EPR systems when using the integrated drug-prescribing module.

Today PHC needs a specialised terminology to support the following: clinical data capture (structured data entry), information integration and retrieval, messaging between software systems, as well as statistical reporting. A controlled vocabulary is needed in the EPR systems for diagnoses, health problems, findings, and interventions.

Management of Chronic Heart Failure

Chronic heart failure (CHF) is a common cause of death and morbidity with a reported prevalence in the population of between 0.4% and 5.2% (Parameshwar 1992, Schocken 1992, McDonagh 1997, Mosterd 1999, Davies 2001, Agarwal 2001). It is also responsible for a significant proportion of the national health care budget. This syndrome consists of symptoms and signs that may have potentially different causes and treatments. Echocardiography is therefore regarded as the key investigation in suspected CHF (Dargie 1994). It is well known that angiotensin converting enzyme inhibitors (ACEI) reduce morbidity and mortality (CONSENSUS trial study group 1987, SOLVID investigators 1991, Garg 1995, Cleland 1998), prevent the need for hospitalisation and save money in the treatment of patients with CHF (Kjekshus 1988, Hart 1993). Further, a nurse-directed, multidisciplinary intervention can improve quality of life and reduce hospital use and medical costs for elderly patients with CHF (Rich 1995).

Evidence suggests that GPs treat most of these patients (Wheeldon 1993). There are national and regional guidelines for GPs on the management of CHF (Läkemedelsverket 1992, European Society of Cardiology 1997, Remme 2002). The diagnosis and management of CHF are changing rapidly (Dargie 1994), and many practical problems occur in trying to implement and measure adherence to such recommendations in PHC (James 1997). A substantial proportion of CHF patients do not get the treatment they need (Clarke 1994, Hillis 1996, Morgan 1999). The barriers to effective management of CHF in general practice are complex (Hickling 2001). Knowledge concerning the management of CHF in Swedish PHC is limited. A recent study on prescribing indicated that the use of ACEIs in CHF is increasing very modestly (Burman 2001).

Changing Prescribing Behaviour

There is a steady increase in drug consumption in Western countries. New and more expensive drugs are continuously introduced, leading to increased drug costs,

even though the therapeutic gains with these newer drugs may be modest or even non-existent. Older drugs receive little promotion, but may deliver equivalent health outcomes at a much lower cost. Expenditures for pharmaceuticals consume an increasing proportion of the limited overall healthcare budget in Sweden, and the preference of GPs for new and more expensive drugs appears to be the main factor increasing costs (Hill 1997). In Sweden about 60% of all drugs are prescribed in PHC, and healthcare costs for pharmaceutical products have been increasing by approximately 13% annually (Burman 1999). On average, a GP prescribes drugs for almost 1.2 million SEK (Swedish Kronor) (approximate exchange rate: US dollar 1 = SEK 10) per year (Göransson 1997). Prescribing behaviour varies widely among doctors (Maronde 1971, McGavock 1988, Oxman 1995). The rates of uncashed prescriptions conceal important variations in both patient and doctor behaviour (Jones 1998). Many factors affect a doctor's prescribing behaviour, such as education, marketing, relationships to the pharmaceutical industry, colleagues, regulatory actions, demands from society and patients, and personal characteristics of the doctor (Hemminki 1975). Consequently, research into factors affecting prescribing behaviour has attracted increased interest in the medical profession and in society.

Changing prescribing behaviour is difficult and often requires complex interventions (Oxman 1995), and such efforts may have only modest effects on practice (Poses 1995, Freemantle 2000). Audit and feedback may improve clinical behaviour, but such attempts should not rely solely on this approach (Thomson O'Brien 2000a). For example, feedback on prescription rates and costs as a single intervention did not influence the prescribing behaviour of GPs (O'Connell 1999). It is not possible to recommend a specific complementary intervention to enhance the effectiveness of audit and feedback (Thomson O'Brien 2000b). Educational outreach visits may modify professional behaviour such as prescribing, but further research is needed to identify the key characteristics leading to success (Thomson O'Brien 2000c). For example, brief face-to-face educational visits by specially trained clinical pharmacists or physicians may reduce prescribing of inefficient or contraindicated drugs in the PHC setting (Soumerai 1990). Interactive education (Lundborg 1999), and "academic detailing" can also be effective educational methods for influencing prescribing in general practice (Diwan 1995, Tomson 1997). Local opinion leaders can sometimes change professional healthcare practice, but studies show mixed results (Thomson O'Brien 2000d). Audit and feedback, educational outreach visits, distribution of educational material, conferences, discussions with local opinion leaders, reminders, or local consensus

processes have modest or negligible effects when used alone. However, combinations of different strategies may be more effective.

Preceding Studies

Modeling and Reclassification of KSH97P

This preceding study addresses some of the problems described above that have been recognised with traditional coding schemes such as the ICD. The primary objectives were to create additional values for a traditional classification, and to test the feasibility of using a concept system (semantic terminological model) as an approach to improving a traditional classification (Petersson 1997). It was performed as a part of the Prim.dok project where a comprehensive concept model of multidisciplinary PHC was developed (Blomström 2001).

A terminology working-group with 11 GPs with extensive experience with EPR systems was formed within a network of eight health care centres in Stockholm County. Concept modeling and analysis of all concepts considered useful for follow-up of state of health in PHC were performed in a terminology working-group led by an experienced terminologist. Initially, a common goal for the modeling was defined, and limits were set for the modeling. International Organization for Standardization recommendations for work terminology were considered (International Organization for Standardization 1987). Twelve half-day seminars from March 1996 to May 1997 were held with members of the working-group. Concepts were described by means of terms, definitions, and their relations in a conceptual graph using Unified Modeling Language (Jacobson 1999). In the conceptual model all concepts were labelled, and they were defined by their relationships to other concepts. The Rational Rose® 98i programme was used as a tool for visual modeling. It facilitates the use of the Unified Modeling Language, and was used to create, analyse, modify and communicate concept models.

A basic concept system for PHC was initially developed and the basic concepts were found to be patient, care provider, procedure and state of health. Procedure and state of health form the base of the care process, represented in the concept system by the links between these concepts. Concepts related to state of health were identified, labelled and mapped (connected) according to their use as a new structure for the KSH97P. Highest priority was given to location, origin, and type, in favour of, for example, social context, severity and duration. In this sense, "origin" means "beginning of" as well as "aetiology". A concept system that was relevant for follow-up of state of health was developed with a total of 32 concepts (Fig. 1). The concept system contained two semantic relation types: generic

relations ('is-a' relations) and a number of non-generic relations (ontological relations). This concept model was used for the reclassification of the KSH97P.

A corresponding hierarchical three-dimensional concept representation system of the KSH97P was developed (Table 1). Since a three-axis representation corresponds to a cube, the model is called the Cube. The number of dimensions was limited to three. The axis of location describes functional systems of the human body. One subgroup of origin is mixed origin, which is used for codes with many types of origins (e.g. tumours). The axis of type corresponds to the generic part of state of health in the concept system. Here disease is defined as a concept with a prognosis (usually also a treatment and an aetiology). Symptom is defined as a complaint with a non-specific prognosis, treatment or aetiology. Risk is defined as a state of health without any symptom but with increased risk for disease (e.g. smoking).
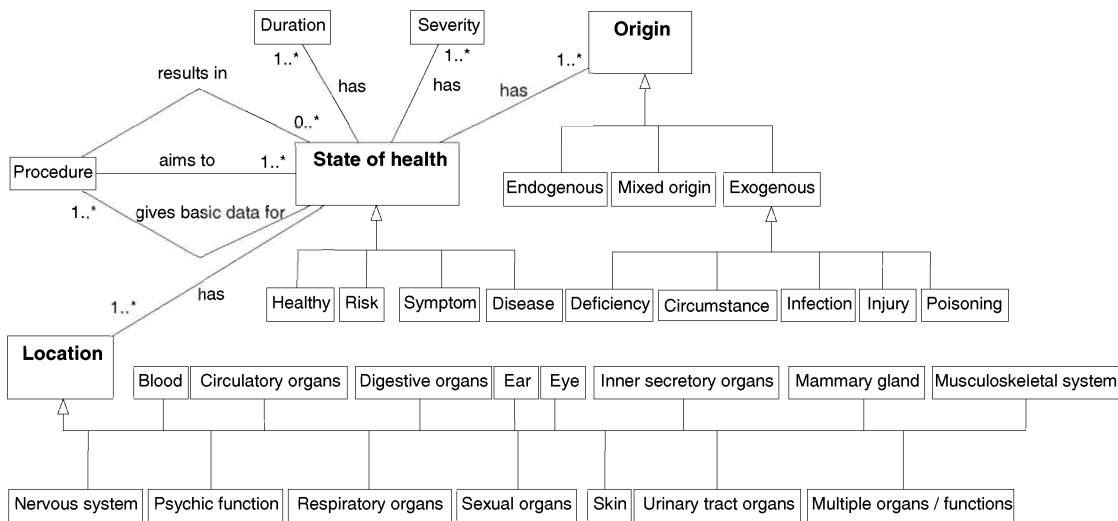

Figure 1. Concept system of state of health.

In order to improve the usefulness of the KSH97P, additional information about the codes was added. Each code within the KSH97P is connected to the structure for describing location in the Cube. Similarly, structures for describing origin and type were added. Additional definitions were given to all codes of the KSH97P as they were placed within the Cube. The statistical distribution of classifications

according to dimensions of the Cube is displayed in Table 1. It is possible to combine the elements from one, two, or three dimensions to retrieve or abstract a group of diagnoses from different ICD-chapters. Codes concerning procedures (e.g. vaccination) were handled separately, as they were not subordinated to the concept state of health (see Fig. 1). Diagnostic codes concerning procedures were found in 3.1% of the codes.

Location		**Origin**	
Multiple organs/functions	14.7	Endogenous	38.5
Musculoskeletal system	12.8	Mixed origin	30.8
Digestive organs	11.0	Infection	14.1
Skin	10.1	Injury	9.8
Sexual organs	8.2	Procedures	3.1
Circulatory organs	6.9	Circumstances	1.7
Respiratory organs	6.4	Poisoning	1.1
Nervous system	5.4	Deficiency	0.6
Psychic functions	5.2	Exogenous	0.2
Eye	3.7	*Total*	*100*
Urinary tract organs	3.5		
Procedures	3.1		
Blood	2.7	**Type**	
Inner secretory organs	2.7		
Ear	2.4	Disease	67.6
Mammary gland	1.2	Symptom	25.9
Total	*100*	Procedures	3.1
		Risk	2.9
		Healthy	0.5
		Total	*100*

Table 1. Three-dimensional concept representation system of the KSH97P, and the percentage distribution of codes (N=973).

This new (compositional) version of the KSH97P reflects common diagnoses and symptoms. It is based on sensible groupings that correspond to useful concepts

within PHC, and has implications for problems recognised with traditional classifications (Petersson 1997). The new version of the KSH97P can be defined as a second generation terminology system according to a proposed categorisation (Rossi Mori 1998), which can be briefly described as a compositional system with categorical structure and descriptors.

Implementation of a World Wide Web Accessible Database for KSH97P

This preceding study has been published previously (Petersson 1998). The objectives were to implement two computer-based versions of the KSH97P (one with the traditional structure, and one with the new compositional structure described above) by means of a relational database model accessible on the World Wide Web. A data model, developed by Spriterm (Junfors 1996), a Swedish national terminology project, was used for the technical implementation of the KSH97P. Implementation of the Spriterm model was performed by means of the Microsoft Access™ relational database management system (DBMS). A photograph of the interface of the World Wide Web application built upon the concept representation system is presented in Fig. 2.

The interface has four main frames:

1) The search frame, providing two search functions: free text search and browsing.
2) The presentation frame, used for presentation of search results and browsing.
3) The chapter frame, with functionality similar to the category frame described below.
4) The category frame, which is used for presentation and retrieval of information concerning diagnoses. The parent concepts are listed at the top and appear as hypertext links. To the right are links to the definition, mappings and semantic relations. The next row contains the concept's position within the Cube. Below this row is a list of original terms and alternative terms. There is also a sub frame used for presentation of additional information such as concepts connected via semantic relations.

The World Wide Web version of the KSH97P may promote the quality of classification in PHC, providing additional value with respect to clinical work, audit, research, teaching and administrative purposes.
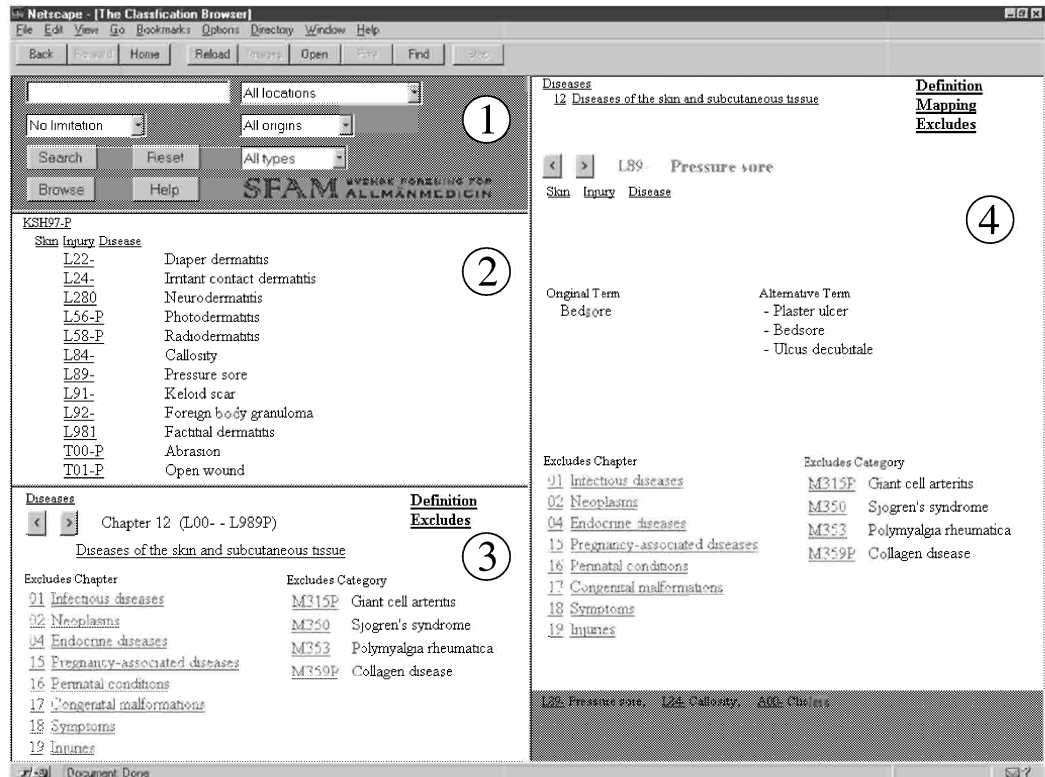
Figure 2. The Classification Browser interface (Petersson 1998).

Desirable features that are difficult to include in a paper-based terminology are multiple classification, multiple consistent views and multiple granularity. The Cube, a model for three-dimensional classification, has been added to the original hierarchical structure. Even though a generic structure is not implemented according to these characteristics, it is possible, through the semantic relations, to retrieve certain types of diagnoses such as those related to the skin. It is also possible to find diagnoses according to combined characteristics, e.g. infectious diseases of the skin, without having these combinations predefined in a complex and rigid chapter structure. The major difference between the information assigned according to the Cube and the original classification scheme is that the Cube classification is not affected by the priority rules inherited from the ICD-10. Despite the fact that the Cube classification may be more clinically relevant, it does not violate the priority rules of the original classification.

User acceptance of the browser, and the reliability of coding at different granularity levels of the original classification, as well as in the new compositional structure of the KSH97P, were evaluated in study III.

Aims

The aim of this thesis was to study aspects of the classification and reuse of diagnostic and pharmacotherapeutic information in general practice. The general goal of the studies was to improve the reuse of information from the EPR. The process of classifying was studied concerning the support needed to produce the right diagnostic codes in an efficient and reliable way.

The specific aims were:

- to study the prevalence, patient characteristics and management of patients with CHF (study I)

- to study the possibilities and limitations of large scale retrieval of data from EPR systems used in everyday clinical practice for purposes of research (study I)

- to develop and evaluate a working model for continuing medical education based on feedback of clinical information from EPR systems in combination with educational outreach visits (study II)

- to study the feasibility of using data from EPR systems for feedback on prescribing rates (study II)

- to study the influence of diagnostic coding tools and schemes on reliability and user acceptance (study III)

- to investigate necessary conditions for the establishment of a database of diseases and health problems for research and health care planning based on EPRs (study IV)

- to investigate textual content, health problems and diagnostic codes in everyday EPRs (study V)

Materials and Methods

The EPR Systems and Data Retrieval

In general, the EPR included the entire medical record and no paper records were used. The text-based version of Swedestar was used in studies I, II and III (Spri 1995). It is designed for PHC and comprehensively supports the problem-oriented medical record (Weed 1968). Further, it has a predefined dictionary used to index all narrative text (Linnarsson 1987). The Swedestar databases were accessible on line for searches and analyses through a statistical module of the EPR system based on Medical Query Language (Webster 1987). In study II, the BMS (Spri 1995), a graphic interface EPR system supporting the problem-oriented medical record, was also used. The BMS database was accessible for analyses through a separate programme (Query Manager™) based on Structured Query Language (SQL).

Coding

In studies I and II a PHC version of the ICD-9 was used for labelling health problems in both systems mentioned above (Socialstyrelsen 1987). Coding of health problems was done by the GPs. Drugs were classified according to the ATC classification system (WHO 1996). Drugs were automatically coded in the EPR system when using the integrated drug-prescribing module, which was normally in use, but were otherwise coded by the GP.

Statistical Methods

The data were analysed manually or using the SPSS® software program. A p-value < 0.05 was considered to be statistically significant. Ninety-five percent confidence intervals (CI) were used.

Independent two-sample t tests were used to compare interval scaled variables (studies I, IV and V). Chi-square and likelihood-ratio chi-square were used to test distributions of cross-classified nominal variables (studies I and IV). We used two-sample t tests for mean differences to compare changes in interval scaled variables (study II). The visual analogue scale data (study III), considered as an ordinal scale, and data that were not normally distributed were presented with medians and ranges and analysed with non-parametric statistical methods. Friedman's ANOVA was used for comparing related samples when there were more than two (study III). For comparisons between ordinal scaled variables for dependent samples, the

Wilcoxon signed-rank test was used (studies III and IV). One-Way ANOVA was used to compare interval scaled variables (studies IV and V). For comparisons between two independent groups and ordinal-scaled variables, the Wilcoxon rank-sum test was used (study IV). In order to take the large number of pairwise tests into consideration, a p-value < 0.01 was considered statistically significant (study IV). When there were more than two groups the Kruskal-Wallis test was applied, and if significant, the Dunn Procedure (p<0.05 was considered significant) was used for pairwise comparisons (study IV). Sidak's test was used for pairwise multiple comparisons (study V) (Kirk 2001). Pearson's correlation coefficient was used for measuring associations of interval scaled variables (study V).

The number of agreements, the percentage level of agreement, and Cohen's Kappa (K) were used to measure reliability (study III) (Altman 1991). The Kappa statistics can range from -1 to 1, but values for K will usually lie between 0 and 1, where 0 indicates only chance agreement and 1 indicates perfect chance corrected agreement. Suggested interpretations of K values are <0.20 poor, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 good, and 0.81-1.00 very good (Altman 1991). However, a sound interpretation of K has to take into account the prevalence of each category, the number of categories, as well as the number of agreements (Gjørup 1988, Brennan 1992). The confidence intervals of K were not calculated, as a much larger sample size than in this study is recommended (Fleiss 1978, Haas 1991). When measuring agreement among more than two examiners, as when comparing the reliability of the two aggregation schemes, Generalized Kappa (K_G) is used (Haas 1991).

Study I

Participants and Study Population

Of 102 permanently employed and active GPs in the Northwest Healthcare District of Stockholm, 46 GPs from 13 PHC centres (using Swedestar EPR system) were asked to participate, and all agreed to do so. The total registered population of the participating GPs comprised 100,222 individuals at the end of the study period, of which 5.6% were 75 years of age or older.

Study Design

A four-year, retrospective, observational study of EPRs was performed in parallel with a feasibility study to investigate aspects of patients with chronic heart failure. All EPRs of patients with a problem labelled heart failure (classified with code 428) who had been in contact with the GP during a one-year period (November 1994 to October 1995), and who were alive at the end of this period, were selected

in a primary data retrieval. Information about these patients concerning all current data on age, gender, date of latest contact, weight, blood pressure, serum creatinine, and date of death was extracted to a separate database. In this primary data retrieval, 968 patients had the diagnosis CHF. All these EPRs were reviewed manually. Of these, only patients whose CHF, according to the EPRs, was specifically managed by the GP were included. Furthermore, only patients with information in their records that fulfilled the criteria for CHF used in this study were included, a total of 667 patients.

Patients were judged to have CHF if they satisfied one or more of the following criteria used in a similar study (Clarke 1994): (1) exertional dyspnoea, orthopnoea, or paroxysmal nocturnal dyspnoea relieved by diuretic treatment, (2) dyspnoea with signs of elevated jugular venous pressure, or a third heart sound, or ankle oedema, or basal crepitations, (3) dyspnoea with an increased cardiothoracic ratio, upper lobe blood diversion, or frank pulmonary oedema on a chest radiogram, (4) dyspnoea with echocardiographic evidence of left ventricular dysfunction.

The EPRs of the 667 patients included in the study were reviewed retrospectively for additional information on (1) number of visits to GPs and district nurses (DN), (2) number of current chronic health problems, (3) occurrence of specific diseases related to CHF (e.g., coronary heart disease and hypertension), (4) specific investigations performed (e.g. chest radiogram and echocardiograph), (5) number of current pharmacological treatments, and (6) pharmacological treatment related to CHF (e.g. diuretics, cardiac glycosides and ACEIs).

In a random sample of 155 of the EPRs, additional information was also sought concerning (1) hospitalisation, (2) estimates of CHF function class, and (3) surgical treatment of heart disease.

Study II

Participants
Three established continuing medical education groups consisting of a total of 23 GPs, and six PHC centres with 27 GPs, were invited, and all 50 GPs agreed to participate. The PHC centres were chosen according to the number of GPs and the EPR system used.

Study Design
An experimental study with three parallel intervention groups was performed in parallel with a feasibility study to evaluate a working model for continuous medical education. The studied pharmacotherapy fields were hypertension, peptic

ulcer/dyspepsia and depression. One of the fields was randomised to each group, which also served as controls for the non-randomised fields.

The intervention consisted of feedback on individual prescribing rates and interactive problem-oriented educational outreach visits, and used educational material and local opinion leaders. A pharmacotherapy education group was formed consisting of physicians with teaching responsibilities, hospital specialists, and clinical pharmacologists. Educational material based on data from randomised clinical trials was compiled by the education group, and included recent national and local recommendations on treatment. A summary of the educational material was handed out. The GPs met three times with the education group, each session lasting 1-1.5 hours. The intervention took place between April 1997 and February 1998, and lasted between two and 11 months in the different groups.

The first visit concerned general information about the project and the pharmacotherapy field in question. The second visit focused on feedback on individual prescribing rates. The third focused on pharmacotherapy education and discussions in the specific pharmacotherapeutic field, and the goals for the specific field.

In the hypertension field we studied prescribing rates for the following main therapeutic groups of the ATC system: diuretics (except loop-diuretics); beta blocking agents; calcium channel blockers; agents acting on the renin-angiotensin system, i.e. ACEIs and angiotensin II receptor blockers; and other antihypertensive agents (ATC subgroup C02A, C02C and C02D). The goal was to favour the prescribing of beta blocking agents and diuretics rather than calcium channel blockers and agents acting on the renin-angiotensin system.

In the peptic ulcer/dyspepsia field we studied prescribing rates for the following therapeutic subgroups: proton pump inhibitors; H2-receptor antagonists; prostaglandins; and other drugs for treatment of peptic ulcer (i.e. ATC subgroup A02BX). The goals were to decrease prescribing in general and per prescription, and to increase the prescribing of H2-receptor antagonists at the expense of proton pump inhibitors.

In the depression field we studied the prescribing rates for tricyclic antidepressants (ATC subgroup N06AA), selective serotonin reuptake inhibitors, monoamine oxidase type A inhibitors, and other antidepressants (i.e. ATC subgroup N06AX). The goal was less specific than in the other groups, namely to focus greater attention on depression and thus increase prescribing, rather than to influence the choice of drugs. One might therefore consider this a non-specific control intervention arm of the study.

30

Source and Measures of Outcomes

The EPR was used as the source of information on drug prescription rates, as sales data from the pharmacy were considered less valid. Prescriptions and the total number of patients were thus collected from the EPR system for each GP. Prescribing rates and Defined Daily Doses (DDD) per prescription were analysed the year before and the year after the intervention. DDDs and DDD/1000 patients seen in one year were used as measuring units (WHO 1996). The units of analysis were the individual GPs.

Study III

Participants

Six GPs from different health care centres, three in each of two Swedish cities, were selected for the study. The subjects were not especially skilled Web surfers, which is the primary skill required for the application in this study. The subjects were considered representative of GPs in Sweden in terms of computer experience.

Study Design

A randomised and experimental multi-centre diagnostic coding trial was performed to evaluate the influence of diagnostic coding tools on reliability and user acceptance. Records from 89 encounters were used for the coding trial. They concerned 152 medical problems, and were randomly selected from a two-year period at one health care centre. The cases were considered fairly representative of Swedish general practice regarding distribution of age, sex and other population factors. The records were problem-oriented, i.e. all notes were recorded in the context of a specific problem (Weed 1968). The cases were presented on paper, and concerned one, two or three problems arranged according to the Subjective-Objective-Assessment-Plan structure. The records averaged 105 words per problem. The complete record of the encounter was presented, thus including prescriptions, laboratory tests, and referrals. Original diagnostic codes, assigned according to the Swedish PHC version of the ICD-9, were excluded.

Three versions of the KSH97P were used in the trial, (1) the Book - the official book version, (2) the Chapter Browser - a computerised version with the traditional ICD-structure, and (3) the Cube Browser - a computerised version with a new compositional structure. The computerised versions were modifications of the Web-based Classification Browser that has been described earlier (Petersson 1998). Two almost identical browsers, the Chapter Browser and the Cube Browser, were produced. The only differences between them were the browsing functions effected by the structure.

The encounters were divided into three sets of 50 or 51 problems. Each GP was asked to code one set of problems by means of each method (the Book, the Chapter Browser, and the Cube Browser) in an order that was determined randomly. They were instructed to select one code for each problem in the record. The GPs were randomly divided into three pairs in order to measure agreement, i.e. inter-rater reliability. One GP repeated the coding trial after four weeks to make it possible to measure intra-rater reliability. Sessions were held at two locations with the client and server computers connected over the Internet. Diagnostic codes for each case were recorded manually by the GPs and handed in for compilation at the end of each session.

The inter- and intra-rater reliability of diagnostic coding was measured on code level and on aggregated levels. The reliability of diagnostic code aggregation was measured on aggregated levels in the traditional ICD-chapter structure, and in the new compositional structure, thereby comparing the two methods with respect to the structuring of categorical data. The whole set of 912 codes was used to compare the reliability of the two alternative structures as aggregation schemes.

The ease of use was measured by the time used, the number of queries sent to different parts of the database, and subjective opinions. The time used was measured for each set of problems, and the number of queries was measured. Information about subjective opinions on the three versions, factors concerning the subjects' previous experiences and the experimental set-up were obtained through a survey and an individual interview. In the survey, a total of 44 questions were used, 18 of which were answered with 100-mm visual analogue scales (VAS) (Gaston-Johansson 1996). Multiple-choice items and comments comprised the remaining 26 questions.

Study IV

Participants

The study comprised 300 GPs randomly selected from 6072 registered specialists in general practice in Sweden who were under 65 years of age. Answers from GPs stating that they were active in general practice, defined as being active for more than six months in 1999, were used in the main data analysis.

Study Design

A descriptive study, using a postal questionnaire, was performed to investigate necessary conditions for the establishment of a database of diseases and health problems based on EPRs. The questionnaire comprised 15 multiple-choice

questions, each with one or several answers as appropriate to the specific question. One question concerned being active in general practice. Two dealt with the EPR system in use, and four concerned the classification systems in use. Three questions addressed the way the GP worked with classification and data retrieval of the codes. Two questions concerned attitudes toward the classification of diseases. Three questions dealt with opinions about the KSH97P.

Verbal rating scales were used as answer formats for nine of the questions. These ordinal scales consisted of one out of five alternative answers ranging from, for example, "Very seldom" to "Very often". The remaining six questions, regarding the classification system in use, data retrieval of codes, and the GP's practice, were nominal data scales with one or more out of four to eight alternatives with one open-ended alternative.

The study was conducted in February and March 2000. Written reminders were sent after two and four weeks when necessary. The first and second dispatches were sent to the GPs' home addresses, which were available from the National Board of Health and Welfare. The third dispatch was sent to the practice sites when the address was available from the national register of physicians (Läkarmatrikeln 2000), but otherwise to the GP's home.

To analyse whether non-responders differed from responders, data on the non-responders in the available registers were analysed according to age, sex, administrative province (county), and whether they were active in general practice according to the national register of physicians. Further, one fourth of all non-responders were randomly selected and their practice sites were contacted by telephone regarding whether or not they were active.

Study V

Participants
The GPs were selected from 6072 registered specialists in family medicine in Sweden who were under the age of 65 years. We randomly selected 26 who lived in Stockholm and who were professionally active in general practice in that city and they were asked to provide data for this study. Twenty GPs from 20 different PHC centres agreed to do so.

Study Design
A retrospective database study was performed to investigate the textual content, health problems and diagnostic codes in everyday EPRs. Basic data on recording routines were collected in a brief structured interview. Twenty face-to-face

encounters between November 1999 and October 2000 (one-year study period) were randomly selected and examined retrospectively for each GP, a total of 400 EPRs. All records were printed out on paper, anonymised, and thereafter scanned to text-files using the Optical Character Recognition technique.

The structure of each EPR was analysed according to use of the SOAP structure. Further, we registered the use of problem-oriented medical records, defined as records where the textual content was separated and oriented around each problem in a separate problem list. The number of words and number of clinical problems were analysed in the different parts of the SOAP structure of each record. Automatically produced texts such as prescriptions generated by prescribing modules in the EPR system were excluded. Further, the text for laboratory tests (produced after the encounter) and referrals (not considered a part of the encounter record) and similar information other than the core EPR were not analysed.

The diagnostic codes were registered, and the completeness and correctness of coding were evaluated by one of the authors. The completeness was estimated for each EPR as the number of clinical problems considered to be described by a diagnostic code. The correctness of coding was evaluated for each diagnostic code by judging whether or not it was a clinically relevant description of a problem described in the text of the record (yes or no). The diagnostic codes were analysed according to their frequencies and distribution in ICD chapters.

Results

Study I

Prevalence and Patient Characteristics

In the study population 968 (1.0%) patients had CHF, as seen in PHC and coded in the patients' records, and of these, 667 (0.7%) satisfied the study criteria for CHF. The prevalence of such patients among inhabitants 75 years or older was 8.9%. Of the CHF patients, 424 (59.6%) were female. The mean age was 79.3 years, 80.7 (CI 80.0; 81.5) for females and 77.0 (CI 76.0; 77.9) for males (p<0.0001). The median number of yearly health care contacts with the GP per patient for any health problem (the last year of the study period) was seven.

Out of a random sample of 155 of the 667 CHF patients, at least one period of hospitalisation for CHF was registered for 43.2%. The distribution of patients in CHF function classes, which was similar for males and females, was 11.6% in New York Heart Association function class I, 54.8% in class II, 32.3% in class III, and 1.3% in class IV (n=155).

Co-morbidity

The mean number of registered chronic diseases, CHF excluded, was 3.5 (CI 3.4; 3.6) per patient, and 658 patients (98.7%) had chronic co-morbidity of some type. The most frequent co-morbidities with clinical relevance to CHF were ischaemic heart disease (37.2%), hypertension (27.3%), chronic atrial fibrillation (23.7%), diabetes (22.3%) and valvular disease (5.1%). Four hundred and ninety-two patients (73.8%) had at least one of these, and 224 (33.6%) had two or more. Statistically significant gender differences were found concerning co-morbidity. Hypertension was more frequent in females (p=0.045), and diabetes (p=0.001) and valvular disease (p=0.048) in males.

Investigations, Laboratory Tests and Treatments

The most frequent major investigations were chest radiograms (66.3%), electrocardiograms (62.1%), echocardiographs (16.9%) and exercise tests (7.7%). The mean weight was 74.3 (kg) and the mean creatinine 101.4 (μmol/L). The mean blood pressure was 146.3/79.1 (mmHg), and the systolic pressure was significantly higher for males (p=0.009).

The mean number of pharmacological treatments ongoing at the end of the study (for any type of drug and indication) was 5.1 drugs. Frequently prescribed

pharmacological treatments indicated for cardiovascular diseases were diuretics (90.9%), cardiac glycosides (48.1%), platelet aggregation inhibitors (32.7%), vasodilators (31.6%), ACEIs (27.4%) (0.7% had tried ACEIs but were found to be intolerant), beta blocking agents (14.5%) and calcium channel blockers (12.8%). Statistically significant gender differences were found concerning management (Table 2).

Drug group	Females n=424	Males n=243
Diuretics (C03) [1]	93 (91; 96)	88 (84; 92)
Cardiac glycosides (C01A)	50 (45; 55)	45 (39; 51)
Platelet aggregation inhibitors (N02BA01) [2]	27 (22; 31)	42 (36; 48)
Vasodilators (C01D)	32 (27; 36)	32 (26; 37)
ACEIs (C09A) [3]	21 (17; 25)	37 (31; 42)
Beta blocking agents (C07) [5]	11 (8; 14)	20 (15; 25)
Calcium channel blockers (C08)	13 (10; 17)	12 (8; 16)
Vitamin K antagonists (B01AA)	7 (5; 10)	11 (7; 15)

Table 2. Pharmacological treatments with ATC classification codes used in patients with congestive heart failure, presented as percentage of patients in each drug group and 95% confidence intervals.

Statistically significant difference between genders (Chi-Square Test); [1] $p=0.016$, [2] $p<0.0001$, [3] $p<0.0001$, [5] $p=0.001$.

Two hundred seventy-six (41.4%) were in contact twice or more, for any reason, with a district nurse during the last year of the study period, and 58 (8.7%) had one contact.

Feasibility

EPRs from everyday clinical practice were feasible for research, although several areas need improvement.

Study II

Prescribing Rates

In the hypertension field prescriptions for diuretics were predominant before the intervention in both the intervention and control groups (43.2-48.7% of all antihypertensive drugs), followed by beta blocking agents (24.2-27.2%), calcium channel blockers (12.6-16.9%) and ACEIs/angiotensin II receptor blockers (14.1-15.0%). Prescribing behaviour for ACEIs/ angiotensin II receptor blockers was significantly influenced by intervention, as there was a slight decrease in the intervention arm (from 15.0 to 14.8%) whereas there was an increase (from 10.9 to 14.1%) in the control arm (p<0.05) (Fig. 3). Desired trends, although not significant, were seen in prescription rates of the other three main therapeutic groups: diuretics, beta blocking agents, and calcium channel blockers.
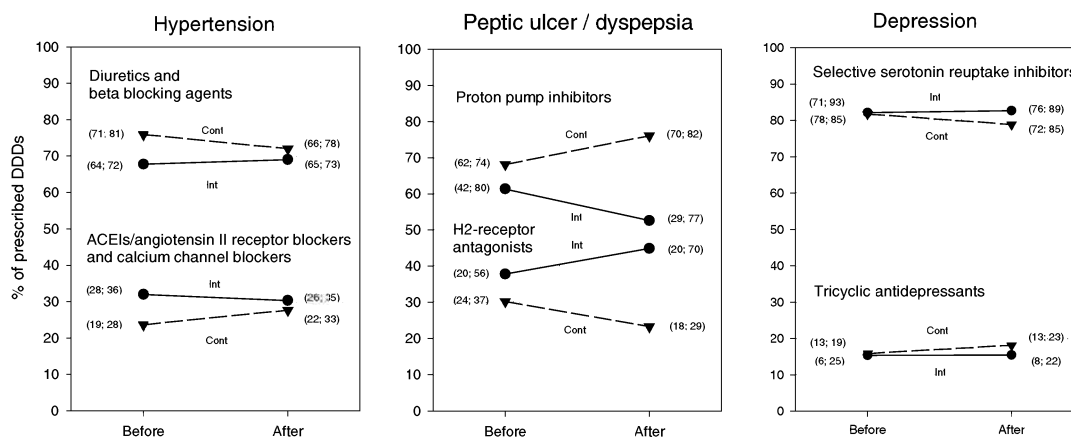

Figure 3. Fractional prescribing rates before and after intervention presented as mean percentage of prescribed Defined Daily Dosages (DDD) with 95% confidence intervals in the three intervention fields. *Int* Intervention group; *Cont* Control group

In the peptic ulcer/dyspepsia field the number of prescriptions (DDDs/1000 patients) increased in the intervention group, but remained fewer than in the control group after intervention. Before intervention, prescriptions for proton pump inhibitors were predominant (≈ 65%) in both groups of GPs. Desired trends were seen in the prescribing habits for both proton pump inhibitors and H2-antagonists, but the changes were not significant due to lack of power (Fig. 3). The individual

37

changes in prescribing habits varied in the intervention group: six out of eight GPs showed changes in line with the goals, and for three of them the changes were marked. Desired – but non-significant – changes were seen in mean DDDs per prescription, with a 3.3% decrease in the intervention group compared to a 0.2% increase in the control group of GPs.

In the depression field there was a 6.8% increase in mean prescribed DDDs/1000 patients in the intervention group compared to a 4.3% decrease in the control group. There was a marked predominance for selective serotonin reuptake inhibitors (81.7-82.1%) before the intervention. Only minor changes were seen in the fractional prescribing of tricyclic antidepressants and selective serotonin reuptake inhibitors in both the intervention group and the control group (Fig. 3).

Feasibility

EPRs from everyday clinical practice were feasible for obtaining feedback on prescribing rates.

Study III

Reliability

The number of agreements at code level showed only a small variation. The inter-coder reliability at code level was moderate (54-59%, K=0.53-0.58), and almost the same using the three versions. The inter-coder reliability at aggregated levels (e.g. an ICD chapter) was good or very good (80-89%, K=0.62-0.87), and there were no significant differences between the three versions.

The intra-coder reliability at code level was good, with a percentage level of agreement of 73-76% (N=51) and K=0.72-0.76. At aggregated levels the percentage level of agreement was 86-98% (N=51) and K=0.72-0.96.

The K_G scores for the whole set of 912 codes at aggregated levels were good or very good (0.66-0.86) (Table 3). The highest score was measured using the dimension of location in the compositional structure (K_G=0.86), followed by the ICD chapter structure (K_G=0.81). The number of classes ranged from 4-38.

Ease of Use

Differences were found in the required time for solving cases with the alternative versions, and the overall differences were significant (p<0.05). Most time was used with the Cube Browser, with a median time of 62 (range 40-104) minutes per set of 50-51 codes. The Chapter Browser required 58 (range 35-79) minutes, and the

Book 51 (range 35-70) minutes per set, respectively. Only the difference between the Book Browser and the Chapter Browser was statistically significant.

Categorisation scheme	Number of classes	K_G
The chapter structure	18	0.81
The compositional structure		
location	16	0.86
aetiology	8	0.74
type	4	0.66
location and aetiology	38	0.76
location and type	29	0.76
aetiology and type	15	0.71

Table 3. Comparison of aggregation based on the two structures. The number of classes and inter-coder reliability measured by the Generalized Kappa statistic (K_G).

The median number of queries to the database to code a specific problem was lower using the Chapter Browser, as compared to the Cube Browser, per set of codes, and this difference was statistically significant ($p < 0.05$).

In the survey, the scores for ease of use in practice did not vary much for the different methods. The scores for ease of getting an overview did vary, but the differences were not significant. The subjects also ranked the methods according to the overall criterion "best". The rank sums were higher for the two computerised versions compared to the Book, but this difference was not statistically significant. The Cube Browser was thought of as a feasible tool for finding uncertain diagnoses or when the GP does not really know what diagnosis he/she is looking for. The physicians found three dimensions in the compositional structure to be sufficient for retrieving diagnoses. Most GPs commented on the lack of a string matching function, with which their own EPR systems were provided. Some wanted to be able to have the diagnoses presented alphabetically instead of being ordered according to code. Coding tools in general practice require high

functionality in order to be accepted by the users, and the study explored the need for several different tools to retrieve diagnostic codes.

Study IV

Response Rates
Of the 300 randomly selected GPs, 50% answered without receiving a reminder, 8% after one reminder, and 18% after two reminders. A total of 228 GPs responded (76%), of which 184 (61% of the GPs) were active in general practice and thereby included in the study (participants). The internal dropout for different questions in the questionnaire did not exceed 3%. No significant differences between responders, non-responders and participants were found with respect to age, sex or being active in general practice.

Characterisation of Participants
The 184 participants represented all 21 counties in Sweden, and in the same proportion as the total number of registered GPs in each county. The mean age was 50.1 years, and 39.7% of the GPs were female. About 88% of these GPs had used EPRs for more than three years, while 8% did not use EPRs at all.

The most widespread EPR system, Profdoc, was used by 35% of the GPs, Medidoc by 16%, BMS by 15%, Biosis by 10%, and Swedestar by 7%. Each of the other 13 systems identified was used by fewer than 3%.

Classification Systems in Use
Some type of diagnostic classification was used by 93%, and some type of ICD based classification by 88%. About 68% of the GPs used only the KSH97P, 10% only the Swedish version of ICD-10, 5% only the Swedish PHC version of the ICD-9, and 3% only some other classification. GPs who used EPRs more frequently used some kind of diagnosis classification (p<0.001). None of the GPs used the ICPC.

Coding Tools
About 74% of the GPs used some form of computerised classification. About 57% used only a classification integrated in the EPR system. About 3% used only a computerised form in a separate programme, and 18% used only a paper-based classification. About 15% of the GPs used more than one form of computerised classification. A digest of the classification in abbreviated form was used "often" or "very often" by 39% of the GPs and "not at all" by 31%. GPs who used only a

paper version of the classification more frequently used a digest of the classification (p<0.001).

The 133 GPs with computerised classifications used different functions in order to find a specific code for a disease. An advanced string matching function was most frequently used, but a majority of GPs used only coding tools with fairly simple functions.

Classification Activities and Related Attitudes

About 76% of all GPs stated that they "almost always" classified some disease in relation to an encounter. Their mean age was higher than that of the remaining GPs, although the difference was not statistically significant. Twelve percent "mostly" classified a disease, 7% did so at an intermediate frequency, and 5% "never" classified a disease. GPs classified diseases at different frequencies based on the EPR systems they used. GPs who used BMS classified diseases more often (93% "almost always") than users of Profdoc (67% "almost always") (p=0.037).

Disease codes were infrequently used for data retrieval from the EPR systems. About 19% stated that they retrieved data containing disease codes "once a month" or more often. About 45% of the GPs retrieved data "once a year", 27% "not at all", 6% did not use EPRs, and 3% did not know.

Of the 137 GPs who retrieved data, 46% stated that the aim was multipurpose, 26% that it was for their "annual report", 20% for "quality assurance", 4% for "other purpose", 1% for "research", and 2% did not know.

Attitudes towards the importance of classification concerning annual reports and quality assurance were more positive than concerning the care of the patient (p<0.001).

Opinions regarding the KSH97P were significantly more positive concerning classifying a disease than a symptom (p<0.001). The size of the KSH97P was stated as being "much too long" by 7% (n=133), "too long" by 16%, "just long enough" by 46%, "too short" by 29%, and "much too short" by 2%.

Study V

Participants

Twelve of the 20 GPs were females, and the mean age was 51.3 years. Eighteen of the GPs usually had a secretary to transcribe dictated records, while two typed the records themselves. All GPs used a Swedish PHC version of the ICD-10 for labelling health problems (Socialstyrelsen 1997). Coding of health problems was

done by the GPs. All GPs used an EPR system with an integrated drug-prescribing module, where drugs were automatically coded according to the ATC classification system.

EPRs

Five (1.3%) of the 400 encounters were home encounters and the others were office encounters. One hundred and ninety (47.5%) of the encounters were planned, 168 (42.0%) were unplanned, and whether the others (10.5%) were planned or unplanned was indefinite. The mean age of the patients was 51.2 years and 60.0% were women.

Out of the 400 records, 391 (97.8%) were SOAP-structured. Fifty-eight (14.5%) records were problem-oriented medical records, and all of these were SOAP-structured.

The mean number of words in the complete EPRs (excluding prescribing, laboratory tests and referrals) was 99.4 (CI 92.0; 106.8), and the largest number of words was in the subjective part of the SOAP structure 48.4 (CI 44.3; 52.5). Assessment and plan were usually merged under a common heading (e.g. 'assessment and procedures'). When merged, this text was manually separated into its corresponding parts. The mean number of words in the complete EPR averaged from 52.3 to 175.9 per GP and from 61.6 to 139.6 in the different EPR systems (Fig. 4).

The total number of problems in the 400 EPRs was 494. The mean number of problems per EPR was 1.2 (CI 1.2; 1.3). It was not possible to divide the objective part into separate clinical problems, and the number of problems in this part was therefore not analysed. The mean number of problems in the different EPR systems averaged from 1.1 to 1.5 per EPR (Fig. 4).

The total number of diagnostic codes in the 400 records was 422. The mean number of codes per EPR was 1.1 (CI 1.0; 1.1), ranging from 0.1 to 1.8 per EPR for different GPs. The mean number of codes per EPR also differed for different EPR systems ($p < 0.001$), and was higher for Swedestar than for the other systems as a group ($p < 0.001$) (Fig. 4). In 75 (18.8%) of the 400 EPRs there was no coded diagnosis.

The mean completeness of diagnostic codes was 0.9 (CI 0.8; 0.9). The correctness of codes was 97.4%. There was a significant correlation between the number of words and problems (Pearson's correlation coefficient = 0.49, $p < 0.01$) and the number of problems and codes (Pearson's correlation coefficient = 0.65, $p < 0.01$).
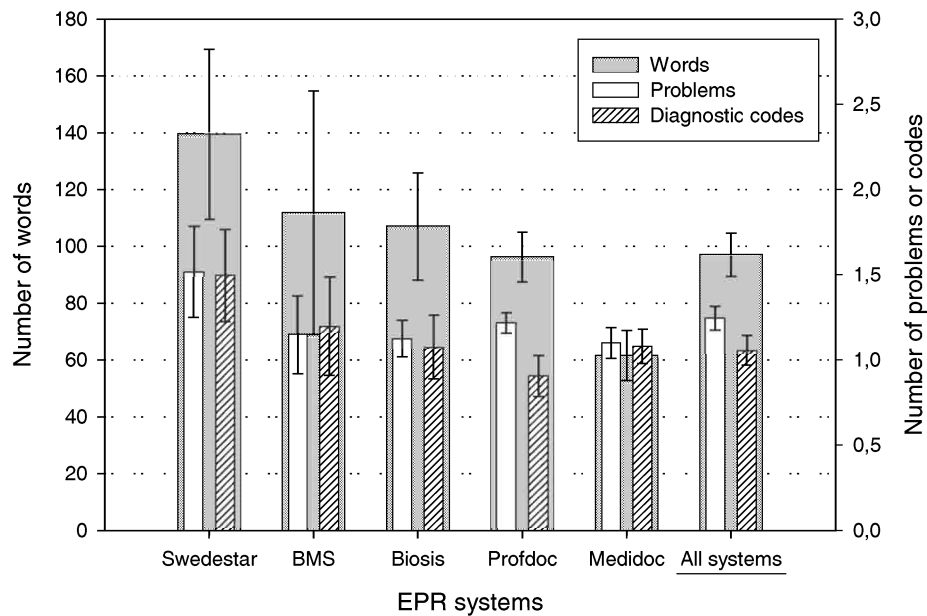
Figure 4. The mean number of words, problems and diagnostic codes in 400 EPRs, in different EPR systems and totals, with 95% confidence intervals. The number of EPRs from Swedestar = 60, BMS = 20, Biosis = 40, Medidoc = 60, and Profdoc = 220.

A total of 422 codes were collected. The most frequent diagnostic code was essential hypertension (7.8%), followed by acute upper respiratory infections of multiple and unspecified sites (5.9%), general medical examination (2.4%), heart failure (2.1%) and asthma (2.1%). The most predominant ICD-10 chapters were diseases of the musculoskeletal system (16.1%), diseases of the circulatory system (15.6%) and diseases of the respiratory system (14.2%).

Data concerning pharmacological prescriptions were found in 207 (51.8%) of the EPRs.

Discussion

The general goal in these studies was to improve the reuse of information from the EPR, as well as the process of coding. The general hypothesis investigated is that the collection and reuse of coded clinical information can be used for research and education in general practice.

Generalisation and Limitations

The setting is limited to general practice, but we believe that many results are applicable and useful to the health care sector as a whole. This thesis focuses on the reuse of data for research and education, but there are also implications for administrative and clinical areas.

In study I the GPs and the study population were not selected at random, but the number of GPs and their geographical distribution are such that we think that they are representative of Stockholm as a whole. In our view the population sample was such that it can be considered representative of the inhabitants of Stockholm. The observation time in the study was limited to four years. Both new and established patients with CHF were included, which is a limitation. This is a consequence of the limited quality of problem-oriented information in the EPR, making it difficult to identify patients undergoing primary investigations for CHF.

In studies I and II, our findings regarding the distribution of pharmacological treatment can be considered highly valid due to automatic coding of prescriptions in the present systems. The participants selected in study II seem fairly representative of GPs in Stockholm, as they represented existing practices and continuing medical education groups. One limitation in this study was the small number of participating GPs, especially in the peptic ulcer / dyspepsia group. The sample size was not based on power analysis, but was limited by workload implications. Furthermore, since diagnosis-related prescription data were not available, some prescriptions were undoubtedly issued based on other indications than those covered in the intervention. Lastly, the fact that the GPs functioned as participants in one part of the study and controls in another may have had non-specific effects that were not evaluated. These limitations might have reduced our possibilities of detecting differences in prescribing behaviour in the study.

In study III the numbers of practitioners and encounters were considered adequate for the objectives of this study. However, it is difficult to draw conclusions reflecting PHC in Sweden in general, even if the subjects are

considered representative. As noted by some of the subjects, the coding results might have been different under normal circumstances, when actually seeing the patient. Some physicians thought they needed more time to really be able to compare the versions. However, the effects of any improved learning on the results of the experiment were minimised by the random order of the methods used.

In study IV there seems to be no reason to believe that the selected GPs were not representative for Sweden, or that there were systematic differences between responders and non-responders. In study V we investigated EPRs of GPs in Stockholm, and the results are considered fairly representative for the area, and partially representative for GPs in Sweden.

Coding of Clinical Information

Terminologies and Coding Systems

As shown in study IV, the KSH97P is presently available in most EPR systems in PHC in Sweden. The frequent use of ICD-related classifications in PHC in Sweden instead of the ICPC, for example, is the result of recommendations from the Swedish National Board of Health and Welfare. It is considered advantageous to have the same classification of diseases for both PHC and hospital care.

How extensive a PHC adaptation of the ICD-10 actually needs to be is, however, a matter of debate. It seems that GPs have very disparate personal opinions in this regard. In the future, it is possible that improved coding tools could overcome the need for the ICD-10 in a shortened version.

The KSH97P was not as well accepted for classification of symptoms as for classification of diseases. This could be due to the fact that the chapter structure in the ICD mainly supports categorisation of diseases. It is possible that the forthcoming Swedish version of the International Classification of Functioning, Disability and Health could fulfil the needs GPs have for a workable classification of symptoms and related concepts (WHO 2001). However, the concept of disease is logically complex (Campbell 1979), and classifying diseases, syndromes, complaints or health problems with one classification is a challenge. Most widely used terminologies are based on a hierarchical structure, designed for a single purpose, and an item can only be classified in one way. Consequently, there are barriers to the re-use of codes for multiple purposes (Rector 2000). There is a need for richer clinical coding systems, and coding tools in general practice, and these require high functionality to be accepted by the users. Future coding needs to include structured data for all sorts of information, not only diagnoses and health problems, including problem-oriented codes for interventions. It is possible that

forthcoming coding systems such as SNOMED Clinical Terms can fulfil these needs (College of American Pathologists 2000, Price 2000).

Coding Tools

As the time for coding activities in everyday clinical practice is limited, user-acceptance of coding tools is dependent on time-effectiveness. Diagnostic coding without computer-based coding tools, e.g. automatic coding of free text, is time consuming. A shorter time for coding was expected for the two browsers compared to the Book (study III). The time did correlate to the version, but the Book was faster. Coding performance would probably improve after longer implementation. The time-savings reported in one study (Hohnloser 1996), due to the linking of the users' vocabulary to the target term, could probably have been demonstrated by using the string matching function and the alternative terms in the KSH97P, but was not done in the present study.

User-acceptance should have high priority in the development of EPR facilities (van Ginneken 1995). Study III supports the idea that coding tools in general practice require high functionality to be accepted by the users. The subjects expressed the need for several different tools to retrieve diagnosis codes, and they thought it was important to study these. Depending on personal preferences and the type of encounter, different methods are preferred. These include an alphabetical list, a personalised or local short list, the string matching function, the three versions in this study, parts of the Cube Browser in paper form, as well as others. Personal preferences often favour alphabetical or short lists, and this needs to be evaluated, as such lists probably affect reliability negatively.

As shown in study IV, computerised coding tools are widely used and are mainly limited to tools with fairly simple functions. They can be improved by means, for example, of a more advanced string matching function, and additional views based on a compositional concept system. Further, different kinds of diagnostic classification support, such as explanations of a textbook type and rule-based representations of diagnostic criteria, could be advantageous.

The computerisation of classification seems favourable with respect to user acceptance and the reliability of codes, as it was associated with less use of digested or abbreviated forms. Like the introduction of EPRs, the introduction of computerised coding systems requires a new form of discipline when entering clinical information.

Information Accuracy

There is a need for high quality in data retrieval from the EPR, and ideally, all clinically relevant information should be of high accuracy. The fact that information about many day-to-day activities in general practice is of low accuracy is due in part to characteristics of the information itself. One example of clinically relevant data with low reliability is found in the distinction between the concepts of symptom and disease. Some aspects of improved correctness and completeness in coding were highlighted in study II. This was further investigated in study III and will undergo additional study in forthcoming research.

Trying to improve the reliability of diagnostic data is complex. The results in study III point out the difference between the coding system as a coding tool for code generation and as a categorisation scheme for code aggregation. Many factors, such as the purpose of recording the data, the recorders themselves, the number of diagnoses, diagnostic criteria, and the coding methods, all have the potential to influence the reliability (Anderson 1980). For example, many problems in a problem list could be coded both as a disease and as a symptom, as they often appear simultaneously.

In study III, the moderate level of inter-rater reliability at code level indicates the limitations with respect to diagnosis-related data retrieval in Swedish PHC settings. The reliability of diagnostic coding was not improved by the new compositional structure of the classification. However, used as an aggregation scheme it seemed to have advantages over the traditional ICD-chapter structure, and was consequently more suitable for diagnosis-related data retrieval. This supports evidence indicating the advantage of compositional terminologies in terms of improved quality of data retrieval (Brown 2000). Further development of the compositional structure should focus on improvement of the reliability of diagnostic data, but the extent to which this is possible is uncertain. The results indicate that the dimensions of both location and origin may be more reliable. The reliability that can be achieved at code level is a matter for discussion, however. The study explored the need for research on the reliability at code level for different types of diagnostic codes, e.g. codes for well defined diagnoses such as diabetes that may have higher reliability than diagnoses that include vague descriptions such as dyspepsia. The basic problem regarding high reliability in diagnostic coding involves clarifying which set of states of health can be denoted by each code, and making that information explicit and easily accessible. By extending the compositional structure with multiple classification, for example, each state of health can be associated with several locations in the compositional structure, and this could perhaps improve the reliability.

The compositional structure used in study III was created to cover all clinical concepts of state of health that are relevant to data retrieval for purposes of follow-up. The chosen dimensions in the Cube cover the most essential aspects of the parts of an arbitrary process, i.e. where, how, and what (Wartofsky 1968). Extending the description of diagnoses, therefore, requires a thorough approach in order to have high reliability. For example, for a state of health with two locations it is necessary to specify the quantity as well as the quality of how the locations are affected, i.e. to state whether both locations are affected at the same time, or only one of them, and in what way each location is influenced. The two structures tested in our study, the traditional and the compositional, are to some extent a first generation structure (an enumerated term list) and a second generation structure (a pre-defined structure of concepts), respectively (Rossi Mori 1998). The presented concept system of state of health was used to deal with some of the problems that are common in terminology work as described in the introduction (e.g. concept orientation and multiple classification) (Cimino 1998), and which are also of importance regarding the reliability of coding. Concept orientation has been achieved in the Cube by giving each concept in the model a single coherent meaning, e.g. codes related to procedures were kept separate. The Cube also allows multiple hierarchies, which is probably useful in general practice, but its inclusion in a paper-based terminology is problematic.

It seems that the use of EPRs, especially certain specific EPR systems, enforces both recording and coding activities in general, as shown in study IV. This can be attributed to characteristics of the systems, such as support of the problem-oriented medical record, which is a characteristic of the two systems in this study that were associated with frequent classification of diseases (BMS and Swedestar) (Spri 1995).

In study V we investigated EPRs of GPs in Stockholm. We found the EPRs to be extensive, and a majority of the problems managed were coded with a high level of completeness and correctness. As expected, however, the subjective, objective and planning parts of the records were not coded, except for prescriptions, which is a limitation with respect to data retrieval. EPRs in general practice and their content have received little attention, and it is therefore difficult to make comparisons with other health care areas (e.g. hospital care) and countries. In one study of the accuracy of EPRs, high recording practices showed no important lapses in the recording of diagnoses, prescriptions or referrals when compared with the actual process of the consultation (Pringle 1995). One could assume that the text in the EPRs in study V was also fairly accurate in terms of the actual care process. It seems that most EPRs in general practice are quite extensive in terms of the textual

content, and frequently consist of what clinicians have heard, seen, thought, and done. Almost all the EPRs in our study were dictated by the GP and typed by a secretary, and this most likely contributed to the large number of words. The mean number of diagnostic codes per problem (0.9) and the mean number of codes per encounter (1.1) can be considered fairly high, as there are no financial incentives for coding diagnoses for GPs in Sweden. These findings are somewhat higher than in one similar study where the completeness of diagnoses was estimated to be 0.7 (Whitelaw 1996). The importance of the GP and the EPR systems in attaining accurate diagnostic coding is emphasised in study V, where there were great variations among GPs as well among the EPR systems.

Reuse of Clinical Information

Monitoring Clinical Practice

Study I indicates that identification of CHF patients based on the PHC version of the ICD-9 could be advantageous, but this is still problematic. The problems of some CHF patients are probably labelled in the EPR with a symptom (e.g. dyspnoea) or a diagnosis that is only related to CHF (e.g. ischaemic heart disease). Further research is needed on the reliability and validity of these data. Registration of diagnosis-related data should, if possible, be used in future evaluations of prescribing. Collection of diagnosis-related data, however, requires improved routines and tools for diagnostic coding.

The ATC classification system was used for prescriptions in studies I and II, but as the recommended standard for prescribing rates, the DDD, was not registered in the EPR systems, it is therefore difficult to calculate. This would be of advantage, as the number of DDDs per prescription varies considerably between drug groups and GPs, and we suggest that evaluations of prescription behaviour should focus on measurements of DDDs (e.g. DDDs/1000 patients/physician/year). The proper source of data for studies of prescribing behaviour can be discussed; we propose that prescribing rates collected from the EPR are the most valid ones in terms of behaviour of the physician. Diagnosis-oriented prescribing data were not available in the EPR systems in study II. In our study this data would have revealed additional information about rational prescribing (i.e. diagnosis-related prescribing rates) in relation to medical factors, which would probably be useful in the intervention.

Clinical information from EPR systems in PHC, with a registered population, is likely to provide data of fairly high accuracy regarding both the prevalence and management of most common disorders. The prevalence figures in study I (1.0 and

0.7%) were somewhat higher than in some other studies (Parameshwar 1992, Wheeldon 1993), but lower than in one study that estimated the prevalence at 2.0% (Schocken 1992). However, in our study there are probably few patients with CHF who have never been in contact with the GP, but the prevalence may have been underestimated. Prevalence figures are often not mentioned in similar studies, as a registered population is frequently lacking (Krogh-Jensen 1983), and these figures are an advantage of our study.

Age, gender, co-morbidity and blood pressure in study I differed markedly compared to the largest studies evaluating treatment with ACEIs in CHF (CONSENSUS trial study group 1987, SOLVID investigators 1991). These differences indicate the importance of evaluating treatments in their main setting (i.e. PHC), and the potential of reusing EPRs for this purpose. Chest radiograms and electrocardiograms were performed fairly frequently, in line with findings in one similar study (Clarke 1994), although the frequency of echocardiography was lower than the 31% reported in that study. In a recent study on the use of echocardiography in PHC in Australia, the frequency was 64% (Krum 2001), and in hospital settings in Sweden it is performed in 59% of CHF patients (Mejhert 1999). Due to improved access, its use has most likely increased in PHC since completion of study I (Murphy 1996), although probably not dramatically.

In study I there were gender differences, with more pronounced suboptimal pharmacological treatment in women; these findings are in line with findings in one other study (Mejhert 1999, Johansson 2002). The use of ACEIs was greater than in one similar study (Clarke 1994), but much lower than in two recent studies (Muntwyler 2000, Krum 2001). The limited use of ACEIs in general and the predominance of males could be explained partly by age factors and co-morbidity. The importance of ACEI treatment for patients with extensive co-morbidity can be discussed, and needs further evaluation.

In study IV, the infrequent retrieval of codes is somewhat disappointing, even though it is known that few doctors make use of data collection in their practice (Webb 1991). This may possibly be due to the lack of time for assessment activities, limitations in the functions of the EPR systems with respect to data retrieval, and limited knowledge concerning data retrieval. These are barriers to be overcome in more comprehensive efforts concerning data retrieval.

The results concerning distribution of codes in ICD chapters in study V are similar to those obtained in one study (Njálsson 1996), although the use of a supplementary classification (chapter XXI) is less common in our study. The most common single diagnostic code, essential hypertension, is more predominant in

our study, and the frequencies of certain other diseases such as heart failure and polymyalgia rheumatica are higher.

Changing Clinical Behaviour

Studies I and II deal in part with the difficulties involved in implementing practice guidelines in everyday clinical practice. This may be as complex as the acquisition of the new knowledge itself, and can be supported by research based on data from the EPR systems. In study II, reuse of data from EPR systems was found feasible, and seems to be useful for behavioural change when combined with educational outreach visits. It was possible to influence prescribing behaviour regarding antihypertensive and peptic ulcer/dyspepsia drugs. To our knowledge there are no studies in which educational outreach visits are combined with feedback, educational material, and participation by local opinion leaders. Few studies have investigated the effect of educational outreach visits among GPs in a randomised, controlled manner. In the Cochrane review on educational outreach visits (Thomson O'Brien 2000c), only 18 studies fulfilled the inclusion criteria, and 13 of them used some complementary intervention. Three studies used educational outreach visits in combination with feedback (McConnell 1982, Stergachis 1987, Feder 1995), and two of them aimed at improving prescribing behaviour, both with observed positive effects (McConnell 1982, Stergachis 1987). However, in these studies outcomes were based on dispensed prescriptions rather than on prescriptions from the physician and DDDs for the drugs. In the Cochrane review on feedback, which included 37 studies (Thomson O'Brien 2000a), seven aimed at improving prescribing behaviour and showed modest effects in terms of number or cost of prescriptions, and were based on dispensed drugs or claims data.

Factors such as the characteristics of the information provided (e.g. whether it is of little or great interest to GPs), the motivation of the recipient, and the clinical setting also need to be considered when a change in behaviour is desired (Kanouse 1988). The educational model presented in study II was well accepted by the GPs and seems to be efficient, but if supported by complementary intervention it might result in more noticeable effects on prescribing behaviour. The intervention may also have had other beneficial effects not specifically evaluated in this study. Strategies such as computerised reminders, social marketing strategies, regulatory actions, economic incentives and non-commercial marketing directed towards patients may be considered. Our results in study II indicate that different strategies are required in different fields of pharmacotherapy.

Health Care Database

The potential for data retrieval in general in audit and feedback activities was demonstrated in studies I and II. Several areas were found to need improvement, some of which were addressed in study III. Studies I and II indicate that systematic searches of EPRs are feasible for large-scale data retrieval from EPRs in order to identify subgroups of patients in PHC who will benefit from research and continuing audit, especially regarding the highly reliable data on pharmacological treatments. Implementation of clinical guidelines can be supported in this way, and health care services and cost-effectiveness may be greatly affected as a result. However, although most doctors collect practice activity data, only a few make use of it (Webb 1991). This can be altered by efforts to improve the accuracy of medical records, and an increased focus on the reuse of data from the EPRs, which are areas for future research.

In study IV the EPRs in everyday clinical use in Swedish general practice were found to provide several fundamentals for a database of diagnostic data for research and health care planning. The use of EPRs by GPs in Sweden was found to be extensive, and more frequent than reported in recent studies (Bolton 1999, Lenhart 2000, Waring 2000). In this regard, it would be possible to realise a patient-centred longitudinal database on diagnoses and health problems (Stead 1999). Due to the high frequency found for classification of diseases, and its stated importance both for the care of the patient and for follow-up, diagnostic data are probably fairly complete and correct.

In study V, almost all EPRs (97.8%) were structured in SOAP. However, the problem-oriented medical record has not been widely accepted as was recently pointed out (Takeda 1999), and in our study it was adhered to by only 14.5% of the EPRs. This is somewhat disappointing, as this structure is considered to be the most ideal way of describing the care process (Takeda 1999), and it also enforces diagnostic coding. Information systems based on problem-oriented medical records have not been widely accepted in Sweden, possibly due to the increased workload involved in recording.

There are several barriers to the establishment of a valid and reliable database of diseases and health problems, such as the limitations of the classifications, coding tools and routines for data retrieval. Classification and coding in the EPRs are essential, and in the future they should include structured data elements for all sorts of information, including details of family history, risk factors, prescribed drugs, therapeutic procedures, and results. Possibly an extended ICPC, the International Classification of Functioning, Disability and Health, or the forthcoming more comprehensive SNOMED Clinical Terms could improve the quality and quantity

of clinical coding and provide us with high quality coded information concerning a variety of clinical concepts (College of American Pathologists 2000). SNOMED Clinical Terms is a comprehensive coding system designed for EPRs in general practice, and could be useful as such a terminology as it is aimed at both PHC and EPR systems. Classification and coding of all parts of the problem-oriented medical record would, however, result in a significant workload for GPs.

The following efforts directed towards the establishment of a valid, reliable and comprehensive health care database are proposed:

–   Implementation of EPR systems that enforce problem-oriented medical records with classification and coding through the use of structured encounter forms with a problem list as a prominent component.
–   Implementation of richer clinical coding systems, possibly based on a concept model incorporating ICD-10, International Classification of Functioning, Disability and Health, ICPC and SNOMED Clinical Terms.
–   Development of more advanced computerised coding tools for just-in-time coding.
–   Establishment of a multidisciplinary perspective with 'one record-one patient' as the working model.
–   Education and regular training in classification and coding.

Many challenging possibilities are already provided by everyday EPRs in general practice, and a new information management paradigm is underway in PHC. There is an urgent need for co-operative efforts towards a valid and reliable comprehensive health care database to support systematic reuse of clinical information for research, clinical audit, teaching and health care planning, and further research, governmental actions and financial issues are also of vital concern and need to be included.

Conclusions

EPRs in Swedish general practice provide several fundamentals for a valid and reliable database of diagnostic data and prescribing for use in research and health care planning. Information management among GPs in this area has reached a high degree of computerisation. Further, there is conformity in the use of classifications, frequent coding activities are carried out, there are positive attitudes to classification, some promising systems are in use, and there is a workable PHC version of the ICD-10.

The EPRs in general practice in Stockholm are extensive in terms of the textual content. A vast majority of the problems in EPRs are coded with ICD-10 codes, and the levels of completeness and correctness of diagnostic codes in EPRs are high. It seems that problem-oriented EPR systems enforce recording and coding activities.

The reliability of diagnostic coding is not improved by the new compositional structure of the KSH97P as compared to the traditional ICD-structure. At aggregated levels the reliability is good, and it is somewhat better in the compositional structure, which makes it suitable for diagnosis-related data retrieval. Some positive effects of the computerised versions of the classification on user acceptance are found. Coding tools in general practice require high functionality to be accepted by the users.

CHF is a common problem among the elderly in PHC in Stockholm, and the majority of these patients present a complex clinical picture with several chronic diseases and treatments. It seems that CHF patients would be more adequately investigated with echocardiography and would also receive more optimal treatment with ACEIs. There are gender differences among CHF patients regarding age, co-morbidity, investigation and treatment. Everyday EPRs provide a feasible source of clinical information for research, although several areas need improvement.

Feedback on individual prescribing rates, combined with problem-oriented educational outreach visits, can be used to improve prescribing behaviour so that there is a more rational use of drugs. EPRs are feasible for feedback on and follow-up of individual prescribing rates.

There are already many challenging possibilities provided by everyday EPRs in general practice for supporting systematic reuse of clinical information for research, clinical audit, teaching and health care planning.

Forthcoming Research

Further research is needed in the field of classification and coding in general practice, especially concerning aspects that can improve the accuracy of clinical information. Many factors related to the multipurpose reuse of clinical information also need evaluation and improvement.

Studies are underway on the reuse of data, or data mining, in EPR systems in PHC. The aim of one of the studies is to investigate detailed aspects of warfarin treatment. The primary objectives are to study the prevalence and incidence of different treatment diagnoses, the management of warfarin treatment in chronic atrial fibrillation, and methods for data quality checks. A retrospective health economic appraisal study based on EPRs, diagnosis-related production statistics and mortality figures has also been initiated. The aim of this other study is to assess the cost of skin diseases caused by ultraviolet radiation in inhabitants in Stockholm.

Development and research regarding SNOMED Clinical Terms will be carried out at the Centre for Medical Terminology at Karolinska Institutet. The main clinical content of SNOMED Clinical Terms will be translated and evaluated in a recently initiated project. Hopefully, implementation in EPR systems in PHC will follow, making use of experiences from the UK (Iggulden 2001). It is possible that collection of clinical data based on this and other terminologies can be supported by structured data entry. Initiatives to develop and implement such computer aids are underway in Stockholm County. The implementation of such efforts and their implications constitute challenging areas for future research.

Initiatives have also been taken toward the establishment of regional health care databases based on clinical information retrieved from EPRs in general practice. The implications of such comprehensive health care databases involve a number of opportunities and challenges. Some of these comprise the expected increase in associations between medical informatics and bioinformatics (Kulikowski 2002), and opportunities to undertake comprehensive, national audits and epidemiological research.

Acknowledgements

I wish to express my sincere gratitude to everyone who has helped me during my work on this thesis. In particular, I want to thank:

References

Agarwal AK, Venugopalan P, de Bono D. Prevalence and aetiology of heart failure in an Arab population. Eur J Heart Fail 2001;3(3):301-5.

Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991. p. 403-9.

Anderson JE. Reliability of morbidity data in family practice. J Fam Pract 1980;10(4):677-83.

van Bemmel JH. Medical informatics, art or science? Methods Inf Med 1996;35(3):157-72.

van Bemmel JH, Musen MA. Handbook of medical informatics. Heidelberg: Springer-Verlag; 1997. p. 81-98.

Blomström E, Malmberg B-G, Morgell R, Månsson I, Nilsson G, Rising I, et al. Prim.dok projektrapport [Prim.dok project report, in Swedish]. Stockholm: Stockholm County Council; 2001.

Bolton P, Mira M, Usher H, Prior G. A model for the evaluation of computerised codes. The Gabrieli Medical Nomenclature as an example. Aust Fam Physician 1997;26 Suppl 2:76-8.

Bolton PG, Usher HE, Mira M, Harding LM, Prior G. Information technology and general practice. A survey of general practitioner attitudes towards computerisation. Aust Fam Physician 1999;28 Suppl 1:19-21.

Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ 1992;304(6840):1491-4.

Bridges-Webb C. Classifying and coding morbidity in general practice: validity and reliability in an international trial. J Fam Pract 1986;23(2):147-50.

Britt H, Beaton N, Miller G. General practice medical records: why code? why classify? Aust Fam Physician 1995;24(4):612,614-5.

Britt H, Angelis M, Harris E. The reliability and validity of doctor-recorded morbidity data in active data collection systems. Scand J Prim Health Care 1998;16(1):50-5.

Brown PJ, Sönksen P. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. J Am Med Inform Assoc 2000;7(4):392-403.

de Bruijn LM, Hasman A, Arends JW. Automatic coding of diagnostic reports. Methods Inf Med 1998;37(3):260-5.

Burman K, editor. Svensk läkemedelsstatistik 1998 [Swedish pharmaceutical drug statistics 1998, in Swedish]. Stockholm: Apoteket AB; 1999. p. 32.

Burman K, editor. Svensk läkemedelsstatistik 2000 [Swedish pharmaceutical drug statistics 2000, in Swedish]. Stockholm: Apoteket AB; 2001. p. 250.

Campbell EJ, Scadding JG, Roberts RS. The concept of disease. Br Med J 1979;2(6193):757-62.

Campbell JR. Strategies for problem list implementation in a complex clinical enterprise. In: Chute CG, editor. Proceedings of the 1998 AMIA Annual Symposium; 1998 Nov 7-11; Orlando (FL), US. Philadelphia: Hanley & Belfus; 1998. p. 285-9.

Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. J Am Med Inform Assoc 1996;3(3):224-33.

Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. J Am Med Inform Assoc 1998;5(6):503-10.

Cimino JJ. Review paper: coding systems in health care. Methods Inf Med 1996;35(4-5):273-84.

Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37(4-5):394-403.

Clarke KW, Gray D, Hampton JR. Evidence of inadequate investigation and treatment of patients with heart failure. Br Heart J 1994;71(6):584-7.

Cleland JG, Swedberg K, Poole-Wilson PA. Successes and failures of current treatment of heart failure. Lancet 1998;352 Suppl 1:19-28.

Coiera E. Guide to medical informatics, the internet and telemedicine. London: Chapman & Hall; 1997. p. xxi.

College of American Pathologists. SNOMED CT. The College; 2000. Available from: URL: http://www.snomed.org/snomedct.html

[The] CONSENSUS trial study group. Effects of enalapril on mortality in severe congestive heart failure: results of Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). N Engl J Med 1987;316(23):1429-35.

Cook MC, Hartman JA, Russell LR. Integrating personal computers into family practice: a comparison of practising physicians and residents. Bull Med Libr Assoc 1998;86(3):316-25.

Côté RA, Rothwell DJ, Palotay JL, Brochu L. The Systematized Nomenclature of Human and Veterinary Medicine: SNOMED International. Northfield: College of American Pathologists; 1993.

Crombie DL, Cross KW, Fleming DM. The problem of diagnostic variability in general practice. J Epidemiol Community Health 1992;46(4):447-54.

Dahlin B, Hässler A. Problemorienterad journal [Problem oriented record, in Swedish]. Stockholm: Spri; 1986.

Dargie HJ, McMurray JJ. Diagnosis and management of heart failure. BMJ 1994;308(6924):321-8.

Davies MK, Hobbs FD, Davis RC, Kenkre JE, Roalfe AK, Hare R, et al. Prevalence of left-ventricular systolic dysfunction and heart failure in the Echocardiographic Heart of England Screening study: a population based study. Lancet 2001;358(9280):439-44.

Diwan VK, Wahlström R, Tomson G, Beermann B, Sterky G, Eriksson B. Effects of "group detailing" on the prescribing of lipid-lowering drugs: a randomized controlled trial in Swedish primary care. J Clin Epidemiol 1995;48(5):705-11.

Dixon J, Sanderson C, Elliot P, Walls P, Jones J, Petticrew M. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. J Public Health Med 1998;20(1):63-9.

de Dombal FT, Dallos V, McAdam WA. Can computer aided teaching packages improve clinical care in patients with acute abdominal pain? BMJ 1991;302(6791):1495-7.

Earlam R. Korner, nomenclature, and SNOMED. Br Med J (Clin Res Ed) 1988;296(6626):903-5.

European Committe for Standardization [Technical Committee 251]. Medical informatics - Categorical structures of systems of concepts - Model for representation of semantics [PrENV 12264]. Brussels: European Committee for Standardization; 1995.

European Society of Cardiology. The treatment of heart failure. Task force of the Working Group on Heart Failure of the European Society of Cardiology. Eur Heart J 1997;18(5):736-53.

Feder G, Griffiths C, Highton C, Eldridge S, Spence M, Southgate L. Do clinical guidelines introduced with practice based education improve care of asthmatic and diabetic patients? A randomised controlled trial in general practices in east London. BMJ 1995;311(7018):1473-8.

Fleiss JL, Cicchetti DV. Inference about weighted kappa in the non-null case. Appl Psychol Meas 1978;2(1):113-7.

Freemantle N. Implementation strategies. Fam Pract 2000;17 Suppl 1:7-10.

Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health 1992;82(2):243-8.

Flier FJ, de Vries Robbé PF, Zanstra PE. Three types of IS-A statement in diagnostic classifications: three types of knowledge needed for development and maintenance. Methods Inf Med 1998;37(4-5):453-9.

Fortinsky RH, Gutman JD. A two-phase study of the reliability of computerized morbidity data. J Fam Pract 1981;13(2):229-35.

Garg R, Yusuf S. Overview of randomized trials of angiotensin-converting enzyme inhibitors on mortality and morbidity in patients with heart failure. Collaborative Group of ACE Inhibitor Trials [published erratum appears in JAMA 1995;274(6):62]. JAMA 1995;273(18):1450-6.

Gaston-Johansson F. Measurement of pain: the psychometric properties of the Pain-O-Meter, a simple, inexpensive pain assessment tool that could change health care practices. J Pain Symptom Manage 1996;12(3):172-81.

Gersenovic M. The ICD family of classifications. Methods Inf Med 1995;34(1-2):172-5.

Gill PW, Leaper DJ, Guillou PJ, Staniland JR, Horrocks JC, de Dombal FT. Observer variation in clinical diagnosis - a computer-aided assessment of its magnitude and importance in 552 patients with abdominal pain. Methods Inf Med 1973;12(2):108-13.

van Ginneken AM. The structure of data in medical records. In: van Bemmel JH, McCray AT, editors. Yearbook of Medical Informatics 1995. Stuttgart: Schattauer; 1995. p. 61-70.

Gjørup T. The Kappa coefficient and the prevalence of a diagnosis. Methods Inf Med 1988;27(4):184-6.

Grimsmo A, Hagman E, Falko E, Matthiessen L, Njálsson T. Patients, diagnoses and processes in general practice in the Nordic countries. An attempt to make data from computerised medical records available for comparable statistics. Scand J Prim Health Care 2001;19(2):76-82.

Göransson P. Ronjastudien 1996 [The Ronja Study 1996, in Swedish]. Stockholm: Apoteksbolaget AB; 1997.

Haas M. Statistical methodology for reliability studies. J Manipulative Physiol Ther 1991;14(12):119-32.

Hart HE, van der Woulden JC, Höppener P, van Schendel GJ, Knottnerus JA. General practice registration network in the Netherlands: a brief report. J Am Med Inform Assoc 1999;6(2):173-5.

Hart W, Rhodes G, McMurray JJV. The cost effectiveness of enalapril in the treatment of chronic heart failure. Br J Med Econ 1993;6:91-8.

Hassey A, Gerrett D, Wilson A. A survey of validity and utility of electronic patient records in a general practice. BMJ 2001;322(7299):1401-5.

Hemminki E. Review of literature on the factors affecting drug prescribing. Soc Sci Med 1975;9(2):111-5.

Hickling JA, Nazareth I, Rogers S. The barriers to effective management of heart failure in general practice. Br J Gen Pract 2001;51(469):615-8.

Hill SR, Henry DD, Smith AJ. Rising prescription drug costs: whose responsibility? Med J Aust 1997;167(1):6-7.

Hillis GS, Trent RJ, Winton P, MacLeod AM, Jennings KP. Angiotensin-converting-enzyme inhibitors in the management of cardiac failure: are we ignoring the evidence? QJM 1996;89(2):145-50.

Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med Inform Assoc 1997;4(5):342-55.

Hohnloser JH, Kadle P, Peurner F. Coding clinical information: analysis of clinicians using computerized coding. Methods Inf Med 1996;35(2):104-7.

Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med 1995;122(9):681-8.

Huff SM, Rocha RA, McDonald CJ, De Moor GJ, Fiers T, Bidgood WD, et al. Development of the Logical Observation Identifier Names and Codes (LOINC) vocabulary. J Am Med Inform Assoc 1998;5(3):276-92.

Iggulden P, Price C. SNOMED Clinical Terms - looking forward to the next phase. Br J Healthcare Comput Info Manage 2001;18(10):20-2.

Illingworth V, editor. A dictionary of computing. Oxford: Oxford University Press; 1996. Available from: URL: http://www.oxfordreference.com

International Organization for Standardization. Terminology work - Principles and methods [ISO/CD 704:1987]. Wien: The Organisation; 1987.

Jacobson I, Booch G, Rumbaugh J. The Unified Modeling Language reference manual. Massachusetts: Addison-Wesley; 1999.

James NK, Reid CD. Plastic surgery audit codes: are the results reproducible? Br J Plast Surg 1991;44:62-4.

James PA, Cowan TM, Graham RP, Majeroni BA, Fox CH, Jaen CR. Using a clinical practice guideline to measure physician practice: translating a guideline for the management of heart failure. J Am Board Fam Pract 1997;10(3):206-12.

Johansson S, Wallander MA, Ruigómez A, García Rodríguez LA. Treatment patterns among newly diagnosed heart failure patients in general practice. Eur J Clin Pharmacol 2002;57(11):813-7.

Johnson N, Mant D, Jones L, Randall T. Use of computerised general practice data for publication surveillance: comparative study of influenza data. BMJ 1991;302(6779):763-5.

Jones I, Britten N. Why do some patients not cash their prescriptions? Br J Gen Pract 1998;48(426):903-5.

Junfors A, Wallin S-B, Thurin A, Carlsson M. A conceptual model for a terminological system for healthcare in Sweden. In: Brender J, Christiansen JP, Scherrer J-R, McNair P, editors. Medical Informatics Europe '96. Amsterdam: IOS Press; 1996. p. 203-7.

Kanouse DE, Jacoby I. When does information change practitioners' behavior? Int J Technol Assess Health Care 1988;4(1):27-33.

Kinn S, Lee N, Millman A. ABC of medical computing. Using computers in clinical audit. BMJ 1995;311(7007):739-24.

Kirk RE, Natanegara F. Three programs for computing Dunn-Sidak critical values. Psychol Rep 2001;88(3 Pt 2):1067-70.

Kjekshus J, Swedberg K. Tolerability of enalapril in congestive heart failure. Am J Cardiol 1988;62(2):67A-72A.

Krogh-Jensen P. The denominator problem. Scand J Prim Health Care 1983;1(2):53.

Krum H, Tonkin AM, Currie R, Djundjek R, Johnston CI. Chronic heart failure in Australian general practice. The Cardiac Awareness Survey and Evaluation (CASE) study. Med J Aust 2001;174(9):439-44.

Kulikowski CA. The micro-macro spectrum of medical informatics challenges: from molecular medicine to transforming health care in a globalizing society. Methods Inf Med 2002;41(1):20-4.

Lamberts H, Wood W. ICPC International Classification of Primary Care. New York: Oxford University Press; 1987.

Landstingsförbundet. Basårsstatistik 1998 [Base year statistics 1998, in Swedish]. Stockholm: Landstingsförbundet; 1999.

Lenhart JG, Honess K, Covington D, Johnson KE. An analysis of trends, perceptions, and use patterns of electronic medical records among US family practice residency programs. Fam Med 2000;32(2):109-14.

Linnarsson R. Development and evaluation of a complete computerbased problem oriented medical record system for primary care. In: Serio A, O'Moore R, Tardini A, Rogers FH, editors. Proceedings of Medical Informatics Europe '87. Rome: European Federation for Medical Informatics; 1987. p. 209-14.

Linnarsson R. Methods, design and components for a computer-based patient record to promote quality care in general practice [dissertation]. Linköping: Linköping University; 1993.

Lorenzi NM, Gardner RM, Pryor TA, Stead WW. Medical informatics: the key to an organization's place in the new health care environment. J Am Med Inform Assoc 1995;2(6):391-2.

Lundborg CS, Wahlström R, Oke T, Tomson G, Diwan VK. Influencing prescribing for urinary tract infection and asthma in primary care in Sweden: a randomized controlled trial of an interactive educational intervention. J Clin Epidemiol 1999;52(8):801-12.

Läkarmatrikeln [The Swedish register of physicians]. Göteborg: Mediahuset i Göteborg AB; 2000.

Läkemedelsverket. Behandling av akut och kronisk hjärtsvikt [Treatement of acute and chronic heart failure, in Swedish]. Information från Läkemedelsverket 1992;4:221-39.

Maronde RF, Lee PV, McCarron MM, Seibert S. A study of prescribing patterns. Med Care 1971;9(5):383-95.

Marshall G, editor. A dictionary of sociology. Oxford: Oxford University Press; 1998. Available from: URL: http://www.oxfordreference.com

McConnell TS, Cushing AH, Bankhurst AD, Healy JL, McIlvenna PA, Skipper BJ. Physician behaviour modification using claims data: tetracycline for upper respiratory infection. West J Med 1982;137(5):448-50.

McDonagh TA, Morrison CE, Lawrence A, Ford I, Tunstall-Pedoe H, McMurray JJ, et al. Symptomatic and asymptomatic left-ventricular systolic dysfunction in an urban population. Lancet 1997;350(9081):829-33.

McGavock H. Some patterns of prescribing by urban general practitioners. Br Med J 1988;296(6626):900-2.

Mejhert M, Holmgren J, Wändell P, Persson H, Edner M. Diagnostic tests, treatment and follow-up in heart failure patients - is there a gender bias in the coherence to guidelines? Eur J Heart Fail 1999;1(4):407-10.

Merriam-Webster's collegiate dictionary. Springfield (MA): Merriam-Webster; 2001. Available from: URL: http://www.m-w.com/dictionary.htm

Metsemakers JF, Knottnerus JA, van Schendel GJ, Kocken RJ, Limonard CB. Unlocking patients' records in general practice for research, medical education and quality assurance: the Registration Network Family Practices. Int J Biomed Comput 1996;42(1-2):43-50.

Morgan S, Smith H, Simpson I, Liddiard GS, Raphael H, Pickering RM, et al. Prevalence and clinical characteristics of left ventricular dysfunction among elderly patients in general practice setting: cross sectional survey. BMJ 1999;318(7180):368-72.

Mosterd A, Hoes AW, de Bruyne MC, Deckers JW, Linker DT, Hofman A, et al. Prevalence of heart failure and left ventricular dysfunction in the general population; The Rotterdam Study. Eur Heart J 1999;20(6):447-55.

Muntwyler J, Follath F. Management of heart failure in Switzerland. Eur J Heart Fail 2000;2(1):113-5.

Murphy JJ, Frain JP, Ramesh P, Siddiquin RN, Bossingham CM. Open-access echocardiography to general practitioners for suspected heart failure. Br J Gen Pract 1996;46(409):475-6.

Musen MA. Creating and using ontologies: what informatics is all about. In: Patel V, Rogers R, Haux R, editors. Medinfo 2001. Proceedings of the 10th world congress on medical informatics; 2001 Sep 2-5, London, UK. Philadelphia: Hanley & Belfus; 2001. p. 1514.

Nazareth I, King M, Haines A, Rangel L, Myers S. Accuracy of diagnosis of psychosis on general practice computer system. BMJ 1993;307(6895):32-4.

Njálsson T, Sigurdsson JA, McAuley RG. Health problems in family practice. An Icelandic multicentre study. Scand J Prim Health Care 1996;14(1):4-12.

Nuopponen A. Begreppsystem för terminologisk analys [Concept systems for terminological analysis, in Swedish with summary in English, dissertation]. Vaasa: University of Vaasa; 1994.

O'Connell DL, Henry D, Tomlins R. Randomised controlled trial of effect of feedback on general practitioners' prescribing in Australia. BMJ 1999;318(7182):507-11.

O'Neil M, Payne C, Read J. Read Codes Version 3: a user led terminology. Methods Inf Med 1995;34(1-2):187-92.

Oxman AD, Thomson MA, Davis DA, Haynes RB. No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. CMAJ 1995;153(10):1423-31.

Parameshwar J, Shackell MM, Richardson A, Poole-Wilson PA, Sutton GC. Prevalence of heart failure in three general practices in north west London. Br J Gen Pract 1992;42(360):287-9.

Petersson H, Nilsson G, Åhlfeldt H, Malmberg B-G, Wigertz O. Design and implementation of a World Wide Web accessible database for the Swedish ICD-10 primary care version using a concept system approach. In: Masys DR, editor. Proceedings of the 1997 AMIA Annual Symposium; 1997 Oct 25-29; Nashville (TN), US. Philadelphia: Hanley & Belfus; 1997. p. 885.

Petersson H, Nilsson G, Åhlfeldt H, Malmberg B-G, Wigertz O. Semantic modeling of a traditional classification: results and implications. In: Cesink B, McCray AT, Scherrer J-R, editors. MEDINFO '98. Proceedings of the 9th World Congress on Medical Informatics; 1998 Aug 18-22; Seoul, Korea. Amsterdam: ISO-Press; 1998. p. 613-7.

Poses RM, Cebul RD, Wigton RS. You can lead a horse to water - improving physicians' knowledge of probabilities may not affect their decisions. Med Decis Making 1995;15(1):65-75.

Pringle M, Ward P, Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. Br J Gen Pract 1995;45(399):537-41.

Price C, Spackman K. SNOMED Clinical Terms. Br J Healthcare Comput Info Manage 2000;17(2):27-31.

Rector AL, Nowlan WA, Kay S. Foundations for an electronic medical record. Methods Inf Med 1991;30(3):179-86.

Rector AL, Solomon WD, Nowlan WA, Rush TW, Zanstra PE, Claassen WM. A Terminology Server for medical language and medical information systems. Methods Inf Med 1995;34(1-2):147-57.

Rector AL. Faithfulness or comparability [editorial]. Methods Inf Med 1996;35(3):218-9.

Rector AL. Terminology, codes, and classifications in perspective: the challenge of re-use. Br J Healthcare Comput Info Manage 2000;17(2):20-3.

Remme WJ, Swedberg K. Comprehensive guidelines for the diagnosis and treatment of chronic heart failure. Task force for the diagnosis and treatment of chronic heart failure of the European Society of Cardiology. Eur J Heart Fail 2002;4(1):11-22.

Rich MW, Beckham V, Wittenberg C, Leven CL, Freedland KE, Carney RM. A multidisciplinary intervention to prevent the readmission of elderly patients with congestive heart failure. N Engl J Med 1995;333(18):1190-5.

Ridderikhoff J, van Herk E. A diagnostic support system in general practice: is it feasible? Int J Med Inf 1997;45(3):133-43.

Rossi Mori A, Consorti F, Galeazzi E. Standards to support development of terminological systems for healthcare telematics. Methods Inf Med 1998;37(4-5):551-63.

Schocken DD, Arrieta MI, Leaverton PE, Ross EA. Prevalence and mortality rate of congestive heart failure in the United States. J Am Coll Cardiol 1992;20(2):301-6.

Seelos HL. A new paradigm of medical informatics. Methods Inf Med 1992;31(2):79-81.

Shortliffle EH, Perreault LE, Wiederholt G. Medical informatics: Computer applications in health care. Reading (MA): Addison-Wesley Publishing Company; 1990.

66

Socialstyrelsen. Klassifikation av sjukdomar 1987 primärvård [Classification of diseases 1987 primary health care, in Swedish]. Stockholm: Socialstyrelsen; 1987.

Socialstyrelsen. Klassifikation av sjukdomar och hälsoproblem 1997 primärvård [Classification of diseases and related health problems 1997 primary health care, in Swedish]. Stockholm: Socialstyrelsen; 1997.

[The] SOLVID investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 1991;325(5):293-302.

Soumerai SB, McLaughlin TJ, Avorn J. Quality assurance for drug prescribing. Qual Assur Health Care 1990;2(1):37-58.

Spri [Hälso- och sjukvårdens utvecklingsinstitut]. Patientjournaler med datorstöd [Computerised patient record systems, in Swedish]. Stockholm: Spri; 1995.

Spyns P. Natural language processing in medicine: an overview. Methods Inf Med 1996;35(4-5):285-301.

Stead WW, Lorenzi NM. Health informatics: linking investment to value. J Am Med Inform Assoc 1999;6(5):341-8.

Stergachis A, Fors M, Wagner EH, Sims DD, Penna P. Effect of clinical pharmacists on drug prescribing in a primary-care clinic. Am J Hosp Pharm 1987;44(3):525-9.

Sujansky WV. The benefits and challenges of an electronic medical record: much more than a "word-processed" patient chart. West J Med 1998;169(3):176-83.

Takeda H. What has changed after Dr Lawrence Weed's paper in 1968? In: van Bemmel JH, McCray A, editors. Yearbook of Medical Informatics. Stuttgart: Shattauer; 1999. p. 218-20.

Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Audit and feedback: effects on professional practice and health care outcomes (Cochrane Review). The Cochrane Library, Issue 1. Oxford: Update Software; 2000a.

Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Audit and feedback versus alternative strategies: effects on professional practice and health care outcomes (Cochrane Review). The Cochrane Library, Issue 1. Oxford: Update Software; 2000b.

Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Educational outreach visits: effects on professional practice and health care outcomes (Cochrane Review). The Cochrane Library, Issue 1. Oxford: Update Software; 2000c.

Thomson O'Brien MA, Oxman AD, Davis DA, Haynes RB, Freemantle N, Harvey EL. Local opinion leaders: effects on professional practice and health care outcomes (Cochrane reveiw). The Cocrane Library, Issue 1. Oxford: Update Software; 2000d.

Tomson Y, Hasselström J, Tomson G, Åberg H. Asthma education for Swedish primary care physicians - a study on the effects of "academic detailing" on practice and patient knowledge. Eur J Clin Pharmacol 1997;53(3-4):191-6.

Ward P, Morton-Jones A, Pringle MA, Chilvers CE. Generating social class data in primary care. Public Health 1994;108(4):279-87.

Waring N. To what extent are practices 'paperless' and what are the constraints to them becoming more so? Br J Gen Pract 2000;50(450):46-7.

Wartofsky MW. Conceptual foundation of scientific thought. Toronto: Collier-Macmillan; 1968. p. 8.

Webb SJ, Dowell AC, Heywood P. Survey of general practice audit in Leeds. BMJ 1991;302(6773):390-2.

Webster S, Morgan M, Barnett GO. Medical Query Language: Improved access to MUMPS databases. In: Stead WW, editor. Proceedings of the 11th Annual Symposium on Computer Applications in Medical Care 1987. Washington (DC): IEEE Computer Society Press; 1987. p. 306-9.

Weed LL. Medical records that guide and teach. N Engl J Med 1968;278(11):593-600,652-7.

Wheeldon NM, MacDonald TM, Flucker CJ, McKendrick AD, McDevitt DG, Struthers AD. Echocardiography in chronic heart failure in the community. Q J Med 1993;86(1):17-23.

Whitelaw FG, Nevin SL, Milne RM, Taylor RJ, Taylor MW, Watt AH. Completeness and accuracy of morbidity and repeat prescribing records held on general practice computers in Scotland. Br J Gen Pract 1996;46(404):181-6.

WHO. International statistical classification of diseases and related health problems 10th revision. Geneva: WHO; 1993.

WHO. Guidelines for ATC classification and DDD assignment. Oslo: WHO Collaborating Centre for Drug Statistics Methodology; 1996.

WHO. International Classification of Functioning, Disability and Health. Geneva: WHO; 2001.

Zarling EJ, Sexton H, Milnor P. Failure to diagnose acute myocardial infarction. The clinicopathologic experience at a large community hospital. JAMA 1983;250(9):1177-81.

Glossary and Definitions

Classification	1. The act or process of classifying. 2. A systematic arrangement in groups or categories according to established criteria. (Merriam-Webster's collegiate dictionary 2001)
Coding	The transformation of observations into categories and classifications, assigning a number or symbol to each item of information or section of a statement, to enable quantitative analysis to be carried out subsequently. (Marshall 1998)
Compositional structure	A terminological system with categorical structure and descriptors. (European Committe for Standardization 1995)
Concept	1. Something conceived in the mind. 2. An abstract or generic idea generalised from particular instances. (Merriam-Webster's collegiate dictionary 2001)
Concept system	A system where terms (symbols) are separate from concepts (what is being symbolised). There are relations between concepts and between concepts and terms. (European Committe for Standardization 1995)
Generic relation	A relation between two concepts that share an identical set of characteristics but one of them, the subordinate concept, has at least one additional delimiting characteristic. (European Committe for Standardization 1995)
Multiple classification	Allowing concepts to be classified according to more than one hierarchy. (Cimino 1998)
Modeling	The act of creating a model of something for a particular purpose, such as to describe it, understand it, or derive some properties. (Illingworth 1996)
Reliability	The extent to which an experiment, test, or measuring procedure yields the same results on repeated trials.

(Merriam-Webster's collegiate dictionary 2001)

Term

A word or expression that has a precise meaning in some uses or is peculiar to a science, art, profession, or subject. (Merriam-Webster's collegiate dictionary 2001)

Terminology

1. The technical or special terms used in a business, art, science, or special subject.
2. Nomenclature as a field of study.
(Merriam-Webster's collegiate dictionary 2001)

Validity

The property of being genuine, a true reflection of attitudes, behaviour, or characteristics. A measure (such as a question, series of questions, or test) is considered valid if it is thought to measure the concept or property which it claims to measure. (Marshall 1998)

Vocabulary

1. A list or collection of words or of words and phrases usually alphabetically arranged and explained or defined.
2. A sum or stock of words employed by a language, group, individual, or work or in a field of knowledge.
(Merriam-Webster's collegiate dictionary 2001)

Summary in Swedish – Sammanfattning på svenska

Under de kommande decennierna blir återanvändning av klinisk information som finns i elektroniska patientjournaler en intressant utmaning. Det gäller återanvändning för olika ändamål – administrativa, kliniska, för utbildning och forskning. För sådan återanvändning av data krävs att den kliniska informationen klassificerats och kodats. Föreliggande avhandling gäller området medicinsk informatik – insamling, lagring, återhämtning och optimal användning av information – inom primärvård.

I studie I undersöktes prevalensen av, kännetecken för och vården av patienter med kronisk hjärtsvikt, samt möjligheten att i forskningen utnyttja elektroniska patientjournaler som upprättats i den normala mottagningsverksamheten. Fyrtiosex allmänläkare ingick i en fyraårig retrospektiv studie av elektroniska patient-journaler. Av den registrerade befolkningen hade 667 personer (0,7%) diagnosen hjärtsvikt och uppfyllde de diagnostiska kriterier som ställts upp för vår undersökning. Av dessa patienter hade 658 stycken (99%) minst en annan kronisk samtidig sjukdom. Under undersökningstiden genomgick 66% av patienterna med hjärtsvikt en hjärt-lungröntgen, och 17% undersöktes med ekokardiografi. Vanligt förekommande läkemedel, med indikation för kardiovaskulära sjukdomar, var diuretika (91%), hjärtglykosider (48%), kärlvidgande medel (32%) och ACE-hämmare (27%). Skillnader noterades mellan könen i fråga om såväl förekomst av andra kroniska sjukdomar som läkemedelsbehandling. Kronisk hjärtsvikt var vanlig bland de äldre patienterna inom primärvården, och utgör ett komplext kliniskt problem. Det förefaller som om behandlingen av dessa patienter kunde förbättras. Elektroniska patientjournaler var användbara för forskning, även om flera områden kan förbättras.

En modell för medicinsk vidareutbildning baserad på återföring av läkemedelsförskrivningsstatistik och problemorienterad undervisning på mottagningen utvecklades och analyserades i studie II. Fyrtio allmänläkare deltog i en randomiserad experimentell studie, omfattande tre parallella interventions-grupper, som också fungerade som kontrollgrupper för varandra. Tre terapi-områden för behandling med läkemedel valdes ut – hypertoni, magsår/dyspepsi och depression. Förskrivningsdata för ett år hämtades från de elektroniska patientjournalerna före och efter interventionen. För terapiområdet hypertoni registrerades önskvärda tendenser i förskrivningen, med ökat inslag av diuretika

och betablockerare, och med en signifikant minskad relativ förskrivning av medel som påverkar renin-angiotensinsystemet (p<0,05). För området magsår/dyspepsi registrerades önskvärda trender för båda de typer av läkemedel som interventionen riktade sig mot. För terapiområdet depression som inriktade sig på ökad uppmärksamheten på denna sjukdom, registrerades endast små förändringar. Återföring av individuell förskrivningsdata, i kombination med problemorienterad undervisning, utgjorde en lovande modell för förbättring av förskrivningsvanor. Data från de elektroniska patientjournalerna kunde användas för återföring av förskrivningsstatistik, även om diagnosrelaterad statistik inte fanns att tillgå.

Den svenska primärvårdsversionen av ICD-10 datoriserades, omklassificerades baserat på ett begreppssystem, och implementerades i ett webb-baserat sökprogram i tidigare studier. I studie III analyserades de nya versionerna av klassifikationen – med avseende på kodverktygens betydelse för reliabilitet och användaracceptans – i en randomiserad, experimentell diagnoskodningsstudie. Sex allmänläkare kodade 152 medicinska olika tillstånd som presenterades i kliniska fallbeskrivningar, med hjälp av tre versioner av klassifikationen, varav en hade sammansatt struktur. På kodnivå var reliabiliteten låg och den var nästan identisk för de tre versionerna. På aggregerad nivå var reliabiliteten god, och den var något bättre i den sammansatta strukturen. Vissa positiva effekter på användaracceptans noterades för de nya versionerna av klassifikationen. Idéer för förbättringar av användaracceptansen väcktes, och studien belyste behovet av flera olika verktyg för att hämta fram diagnoskoder.

I studie IV undersöktes nödvändiga förutsättningar för upprättandet av en databas över sjukdomar och hälsoproblem för forskning och sjukvårdsplanering, utgående från elektroniska patientjournaler hos allmänläkare. Tre hundra slumpvis utvalda allmänläkare i Sverige deltog i en enkätstudie med frågor om datorisering av patientjournaler, klassificering av diagnoser, kodningsverktyg, kodnings-frekvens, återhämtning av koder samt angränsande attityder. Sammanlagt kom 184 allmänläkare (61% av de 300 som valts ut) att ingå i studien. Omkring 92% använde ett elektroniskt journalsystem, och 88% använde ICD-baserade klassifika-tioner. Den använda klassifikationen var datoriserad hos 74%, och i första hand användes enkla kodningsverktyg. Omkring 76% rapporterade att de klassificerade minst ett symtom eller en sjukdom per patientbesök. Koderna återhämtades "en gång i månaden" eller oftare av 19%. Klassifikation av sjukdomar bedömdes av 83% som viktigt för uppföljning, och av 75% som viktigt för behandlingen av patienten. Omkring 31% av läkarna ansåg att den svenska primärvårdsversionen av ICD-10 är alltför begränsad i storlek. Elektroniska patientjournaler från

73

primärvården i Sverige utgör i flera avseenden en bra grund för en databas över läkemedelsförskrivning, sjukdomar och hälsoproblem.

I studie V undersöktes det textuella innehållet, hälsoproblem och diagnoskoder i elektroniska patientjournaler. Tjugo slumpvis utvalda allmänläkare i Stockholm deltog i en retrospektiv undersökning. Sammanlagt undersöktes 400 journaler, med avseende på struktur, innehåll samt korrekthet i diagnoskodning. Omkring 98% av de journaler som ingick tillämpade strukturen subjektivt-objektivt-bedömning-plan, och 15% var problemorienterade. I medeltal, var antalet ord per journal 99,4 och antalet medicinska problem per journal 1.3. Det fanns i medeltal 1,1 diagnos-koder per journal, och det varierade både mellan allmänläkare och mellan de olika journalsystemen. Antalet koder per behandlat problem var i medeltal 0,9 och andelen korrekta koder bedömdes vara 97%. Den vanligaste medicinska diagnoskoden var hypertoni (8%) och akut övre luftvägsinfektion (6%). De dominerade kapitlen i ICD-10 var sjukdomar i muskuloskeletala systemet och bindväven (16%), cirkulationsorganens sjukdomar (16%) och andningsorganens sjukdomar (14%). De elektroniska patientjournalerna i primärvården var omfattande och diagnoskoderna var i stor utsträckning kompletta och korrekta. Det förefaller som om de problemorienterade journalsystemen uppmuntrar till kodning. Det bedömdes vara möjligt att skapa en databas med diagnosdata utgående från elektroniska patientjournaler i primärvården.

Sammanfattningsvis gjordes bedömningen att det är möjligt att hämta data från elektroniska patientjournaler för forskning och medicinsk vidareutbildning. Datoriserade versioner av en traditionell klassifikation utvärderades, och positiva aspekter noterades avseende användaracceptans och reliabilitet vid diagnoskodning. Elektroniska patientjournaler inom den svenska primärvården är omfattande, har korrekta diagnoskoder och tillhandahåller flera av grundvalarna för att upprätta en databas med klinisk information – ett viktigt område som kräver samverkan och fortsatt forskning för att utveckla dagens primärvård.

Publications