



SemanticMining
NoE 507505
**Semantic Interoperability and Data
Mining in Biomedicine**

Deliverable D23.1

Health Statistics

Delivery date: month 11

Report Version: 1

Report Preparation Date: 2005.01.28

Dissemination level: RE

Associated work package: WP10, WP21

Lead contractor: LiU-IMT, Linköpings universitet

**Project funded by the European Community
under the FP6 Programme “Integrating and
Strengthening the European Research Area”
(2002-2006)**



Table of Contents

0	ADMINISTRATIVE INFORMATION	1
0.1	SUMMARY.....	1
0.2	OVERVIEW.....	2
1	INTRODUCTION	3
2	HEALTH STATISTICS AND HEALTH CARE STATISTICS	4
2.1	THE RANGE OF HEALTH STATISTICS.....	4
2.2	LEVELS OF FOLLOW-UP.....	5
2.3	LEVELS OF AGGREGATION.....	6
2.4	WHAT ARE THE PROBLEMS.....	6
3	HEALTH STATISTICS TODAY	7
3.1	MORTALITY DATA.....	8
3.2	MORBIDITY DATA.....	9
3.3	HEALTH INTERVENTIONS.....	10
3.4	HEALTH OUTCOMES.....	11
3.5	NOMESCO STATISTICS.....	11
3.6	QUALITY INDICATORS AND QUALITY REGISTERS.....	12
4	THE HOSPITAL DATA PROJECT	15
4.1	MAIN OBJECTIVES.....	15
4.2	METHODOLOGY.....	15
4.3	DATA ITEMS.....	15
4.4	ORGANISATION.....	15
4.5	SYSTEMATIC APPROACH.....	15
4.6	RESULTS.....	16
4.7	CONCLUSIONS.....	16
5	EUROPEAN DEVELOPMENT IN HOSPITAL STATISTICS	18
5.1	HOSPITAL DATA PROJECT, PHASE 2.....	18
5.2	EUROSTAT TECHNICAL GROUP.....	18
5.3	EUROSTAT TASK FORCE ON HOSPITAL STATISTICS.....	18
5.4	EUROSTAT WORKING GROUP ON PUBLIC HEALTH STATISTICS.....	18
5.5	OECD.....	19
5.6	WHO REGIONAL OFFICE FOR EUROPE.....	19
5.7	WHO-FIC HOSPITAL DATA WORKING GROUP.....	19
6	VALIDITY AND RELIABILITY	21
6.1	ERROR TYPES IN ICD CODING.....	21
6.2	POTENTIAL SOURCES OF ERRORS.....	23
6.3	VALIDITY IN MORTALITY CODING.....	25
6.4	RELIABILITY IN MORBIDITY DIAGNOSIS CODING.....	25
6.5	IMPACT ON CODING: 'WHO IS CODING'.....	27
7	ENHANCING CLASSIFICATION THROUGH TERMINOLOGY	29
7.1	A FINNISH SERVICE CLASSIFICATION.....	29
7.2	TERMINOLOGY SYSTEMS FOR DATA AGGREGATION.....	31
	REFERENCES	34



0 Administrative information

Lead contractor/partner for WP/Deliverable: LiU (IMT)

Assisting partners for WP/Deliverable: –

Author list:

Lars Berg	National Board of Health and Welfare, Sweden
Matti Ojala	National Research and Development Centre for Welfare and Health, Finland
Håkan Petersson	Linköping University, Sweden
György Surján	National Institute and Library for Health Information, Budapest, Hungary
Ole Terkelsen	National Board of Health, Denmark
Martti Virtanen	WHO Collaborating Centre for Classification of Diseases in the Nordic countries, Uppsala University, Sweden

0.1 Summary

The first phase of this WP has been devoted to compilation of background and baseline material in the field of health statistics, with a natural focus on the situation in Europe. Issues in focus are the scope of health and health care statistics, the tools used for coding and classification, as well as problems of quality and comparability of data. A basic research question is how the move from traditional classifications to reference terminologies may improve the quality of health statistics. A specific aspect of this is the use of SNOMED CT as an aggregation tool in the production of reliable health statistics. An extensive report has been written by the WP.

0.2 Overview

0.2.1 Objectives

<i>Objectives</i>	<i>Progress towards achieving objectives</i>
The main objective of this research activity is to share experience, understanding and development of statistical methods for measuring information quality, ontologies for health indicators, and methods for quantification of semantic distance. Moreover, the objective is to encourage sharing of data material (e.g. quality registries and coded patient data) applicable for development and evaluation.	Three tasks are associated with this WP: 1) Documentation of problems in European health statistics 2) Documentation of ontologies for health indicators 3) Proposal for methods for measuring reliability and semantic distance This far we have completed the first task, and this constitutes the major part of Deliverable D23.1. By giving examples of the connection between classification, terminology and ontology, we have also begun to look into the second task. Task three remains to its full extent.

0.2.2 Project meetings

<i>Meeting</i>	<i>Date and place</i>	<i>Comments</i>
First working session	10 July, 2004 Balatonfüred, Hungary	8 participants in conjunction with NoE summer school
Second working session	15 October, 2004 Stockholm, Sweden	9 participants
Third working session	24-30 October, 2004 Reykjavik, Iceland	5 participants in conjunction with the WHO-FIC network meeting

0.2.3 Deviations from Plan

<i>Causes and Description</i>	<i>Corrective actions</i>
No major deviations from plan.	



1 Introduction

Data compilation has a long history in the field of medicine, especially in the domain of mortality statistics. As early as in the 16th century, mortality statistics were compiled in London during the plague, and in 1662 the pioneering statistical work of John Graunt, which was based on the London Bills of Mortality, was published [Nyström 1986]. While these statistics were not based on systematic classifications, such systems were developed later, and in 1893 the International List of Causes of Death was adopted by the International Statistical Institute. This was the first edition of the International Classification of Diseases (ICD) [Gersenovíc 1995], and after the sixth revision in 1948 the ICD was expanded so that it would also be suitable for morbidity statistics.

In addition to classification, i.e. grouping phenomena according to resemblance, medical data abstraction relies on coding, which means that concepts are represented with a combination of numbers and characters. Both principles have the aim of simplifying patient information for storing and manipulation. Domains for which data are pooled and coded include, among others, epidemiological studies, medical insurance registers, reimbursement, discharge summaries, and quality assessment [Gersenovíc 1995].

Based on their intended use, coding systems have been divided into those that abstract clinical data and those that preserve clinically useful detail in a standardized way [Cimino 1996]. The ICD coding systems and other ‘traditional’ classifications belong to the former category, while systems in the latter category for example constitute an important component in the interconnection of medical records and knowledge-based decision support systems [Cimino 1998]. SNOMED CT is one such example, and section 7.2 describes a pilot study where ICD-coded data is aggregated with the help of SNOMED CT.

2 Health Statistics and Health Care Statistics

Health and health care are not only important for each individual but also important indicators of the state of a society. Therefore statistics about health are an important part of the information system of any developed country.

2.1 The Range of Health Statistics

The basis of all health statistics is in vital statistics—in the simple figures about number of births and deaths, the mean age at death and in the age and gender distribution of the population. These ‘simple’ statistics are however not simple to produce, and reliable data about vital statistics are still rare in the world. In addition to vital statistics, cause of death is the oldest regularly reported health indicator. The International Classification of Diseases, which is currently in its 10th revision (ICD-10), has its roots in mortality coding that started more than 150 years ago. The distribution of ‘underlying cause of death’ is still the most significant single variable in statistical analysis of health.

However, after the Second World War the reporting has increasingly concentrated in hospital inpatient statistics. Most developed countries and among them almost all European countries follow the utilization of hospital care as a health indicator. The statistics are usually concentrated to the principal diagnosis of the hospital stay at issue, although systems usually also collect a number of secondary diagnoses. Surgical procedures referring to extended operation room procedures are usually also registered, but there is less agreement regarding the coding systems used in their classification. The current situation is further outlined in chapter 3.

The development of health care has resulted in shortening of the inpatient treatment periods and increasingly to outpatient surgery where the patient is admitted to the hospital and returns home on the day of operation. Obviously, a similar development is happening in the ‘non-surgical’ activities of hospitals, which means that the inpatient statistics do not any more reflect the health of the population in the same way they used to do. On the local level, outpatient contacts have been registered in most systems for many years. Because of the nature of the activity, the procedures are relatively more important than the diagnoses in the outpatient setting. Interest in national level registration is increasing, but there are many practical problems. International comparisons are currently difficult because of the huge differences in the registration systems.

The case of primary care is different from the secondary level hospital care. For example, the number of contacts is very large and therefore person-based registers exist only on local level. Furthermore, tools for data collection and aggregation (ICD-10 and different procedure classifications) are mainly developed for hospital use and are regarded as difficult to use in primary care settings. One reason is that the granularity of the classifications in a primary care setting should be lower and the reporting system



less complex. Nevertheless, the health problem in the form of diagnosis and procedures is important and has been collected in different local systems. However, international comparisons are—due to variation in registration practice—difficult in primary care as well [Grimsmo 2001].

Statistics on retirement for sickness and long-term sick-leave is usually available through the insurance companies. In addition to demographic information, usually only diagnoses are available; functioning and disability information are usually recorded as free text. Information about short term sick-leave is usually not available outside the local setting.

The statistics are usually also reported as the activity of the different involved health care units. Such reports are usually limited to procedures rather than patient problems. The number and type of staff are also central items.

2.2 Levels of Follow-up

The World Health Organization (WHO) collects and publishes worldwide statistics on causes of death. This work has a long tradition, it is extensive and the comparability is fairly good. However, a very large proportion of the WHO member countries do not report causes of death information to the WHO. The fact that these countries are those who are the most exposed with regard to health problems is called an ‘information paradox’ by WHO. Another paradox is that the World Health Organization collects much data on death but very little on health and wellbeing. The World Health Report, last published in year 2000, has been a trial towards real ‘health reporting’. The idea of DALY’s (Disability-Adjusted Life Years) is important in this respect but practical solutions are not clear yet.

Although it is an organisation for economical cooperation, the Organisation for Economic Co-operation and Development (OECD) collects important information about health. Among other things it seems to be the only organisation reporting case mix (Diagnosis Related Groups, DRG) based statistics. Although this is an interesting approach, the quality of the data seems too heterogeneous for reasonable analysis.

The European Union (EU) reports health statistics through the Statistical Office of the European Communities (Eurostat). The problem is poor comparability and as an example, the EU has initiated a special Hospital Data Project to standardize the data collection. The statistics are based on diagnosis classification on ICD-level and on different national procedure classifications. The Hospital Data Project has recommended short lists for both diagnoses and procedures, which would standardize the different lists existing in this area. The Hospital Data Project is further described in chapter 4.

National statistics are collected in all 15 old EU countries and probably also in the new member countries. The principles and methods of collecting data vary, and the coverage of the data is not always complete. The general comparability of the statistics is rather poor as reflected in the Eurostat and OECD reports.

2.3 Levels of Aggregation

Because of the nature of statistical analysis the groups used in the process need to be of acceptable size. One main reason for this is confidentiality—an individual may never be identified in the tables; another main reason is statistical—any statistical analysis demands a sufficient number of observations in the analyzed groups to have reasonable statistical power. Usually this means aggregation to a level where the number of groups is much less than in for example ICD-10. For example, the recommend diagnosis shortlist of the Hospital Data Project contains 130 entries [HDP 2003].

2.4 What Are the Problems

The main and by far most important problem is data quality, which has to do both with coverage and actual quality of data. Generally one can say that data is recorded in health care only because it is demanded, and in most cases this means that data is entered to avoid financial losses to the institute at issue. Thus economic considerations may also distort the resulting statistics. Quality in terms of validity and reliability is discussed in section 6.1 and 6.4 respectively.

The second important problem in the development of effective and flexible health statistics systems is confidentiality—any misuse of personal data must be absolutely impossible, which complicates the development of the systems.

As a matter of fact, most of the available statistics are health care statistics about the use of health care facilities. According to a widely accepted WHO definition of health, health is not just the absence of disease, which means that the health of a population should also be described in terms of functioning and ability to work and enjoy life. The International Classification of Functioning, Disability and Health (ICF) published by the WHO in 2001, is the first attempt to create a tool for systematic collection of this type of information.

3 Health Statistics Today

This part of the report will give a short overview of the use of health statistics, the problems and possibilities with statistics and the classifications used. The focus will be on health care statistics, not on death statistics.

Health statistics can be divided in

- Mortality statistics (death statistics)
- Morbidity statistics (health care statistics)
 - Inpatient care
 - Day care
 - Outpatient care

When considering outpatient care it can consist of statistics from the visits to

- Hospital doctors/specialists
- Family Physicians/General Practitioners
- Nurses
- Other health care personnel (occupational therapists, physiotherapists etc.)

The health care statistics can be used to get information about different aspects of the health care, for a variety of purposes and for international or national comparisons. Within countries, health care statistics can for example be used to analyse regional performance and to identify areas that may require action. It can be used for information about infrastructure, activities, personnel and costs, and for purposes such as supporting activity monitoring, performance measurement, case mix-based funding, service planning and epidemiological analysis.

However, in a global perspective, we have an *'information paradox'*; countries with the greatest health burdens and needs have the biggest information gaps. They not only have the least information but also limited capacity (skills, systems) to generate, analyze, present and disseminate information.

Health information is the glue that holds a health system together. In most countries stronger, more integrated information systems are required. One example is vital registration systems—the ability to count births and deaths. These systems are still missing for most of the countries' population, especially in countries with high disease burdens. To make people count, we first need to be able to count people.

J. W. Lee - WHO Director General [WHO 2004, page 10]

Walker and McKenzie, report results from a current survey study on coding practices [Walker 2004]. The survey was distributed by the WHO-FIC Education Committee in early 2004, and was sent to the WHO member states to request information about the international coder workforce. Data about morbidity coders and mortality coders were collected separately, and the survey included a broad range of questions covering topics such as implementation and use of coding systems, characteristics and responsibilities



of the coders, educational background and training, development of an international coder credentialing process and support to further promote coding in each country.

Approximately 100 survey responses each were returned for morbidity coders and mortality coders. This represents approximately a 50 % overall response rate from the 192 WHO member states.

3.1 Mortality Data

Of the 192 WHO Member States, 83 are not in a position to provide recent mortality data. Of these 83 countries, 70 are located in three regions: 40 in the African Region, 10 in the Eastern Mediterranean Region and 20 in the Western Pacific Region. The 83 countries which cannot report basic mortality figures, in fact, have an estimated burden of mortality that is 4.17 times that of those who report [WHO 2004].

In addition, even in those countries where ICD has been implemented, there is a wide divergence of approach and consistency in its use for something as fundamental as the recording of causes of death. Many countries continue to use ICD-9, and some use short lists rather than the full classification. While some countries have moved largely to automated coding, others rely wholly or in a large part on manual coding. Where manual coding is used, in some countries it is performed by trained clerical or higher level coders and in others by certifying physicians.

Of 99 responding countries in the Walker and McKenzie study, 93 reported that they code mortality data to some extent. Of those countries, 77 (83 %) code and report the underlying cause of death for all occurrences, while 16 (17 %) countries code only a portion of deaths, e.g. those that occur in hospitals. [Walker 2004]

ICD-10 is the most used classification, and it is used in 80 (86 %) of the coding countries. Eight countries (9 %) use ICD-9. The remaining countries use older ICD versions, ICD modifications or shortlists.

Job titles of people who complete the coding of mortality data varied. Responses were as follows, noting that multiple responses were permitted:

- 19.4 % reported clinical coders
- 33.3 % reported administration officers
- 33.3 % reported health information or medical record managers
- 40.9 % reported statistical assistants
- 24.7 % reported medical officers
- 18.3 % reported other professional groups (including nurses, epidemiologists and nosologists)

Automated coding is not widely implemented. Among the countries that code mortality, only 7 (8 %) use an automated system exclusively. Another 20 (22 %) use a combination of automated and manual coding methods, while 70 (75 %) countries use coding books to code manually.

3.2 Morbidity Data

Beyond mortality, ICD can be used to code diseases in health care settings. To serve this need several countries have found it necessary to develop clinical modifications beyond the international core. Examples include the United States (ICD-9-CM and ICD-10-CM); Australia (ICD-10-AM), Canada (ICD-10-CA), and Germany (ICD-10-GM). In addition, many other countries try to adopt one or more of these systems to have an operational system for reimbursement purposes and better planning of purchasing of care.

The modifications are typically the basis for case mix groupings (DRGs), which are used for reimbursement and related purposes. Their use for this purpose has justified the investment necessary to build the clinical modification, but the modification is then used for routine reporting and so has far wider application than case mix groupings.

Clinical modifications are extensions of the reference ICD classification. These allow new content to be added to the derived classification which has not yet been included in the reference classification. As such, these modifications provide ready material for the updating or revision of the reference classification.

Each country must invest heavily to produce and maintain its clinical modification, and yet there is much common content between them. The different ICD-10 clinical modifications are, however, neither compatible with each other nor fully compatible with ICD-10. The coding standards for at least some of them differ from the international rules e.g. regarding the definition of the principal diagnosis for morbidity coding. This situation is a threat to the comparability of morbidity data at the international level.

Walker and McKenzie found that out of 96 respondents, 90 reported that morbidity data is coded in their country. In 59 (66 %) of those countries all hospital discharges from all hospitals are coded, and in 19 (22 %) countries, all discharges from some hospitals are coded. [Walker 2004]

Considering diagnosis coding, 74 (82 %) countries use ICD-10 while 8 (9 %) countries use ICD-9. Other ICD versions are used in 14 (16 %) countries. As for mortality coding, the job titles of people who complete the coding of morbidity data varied. Responses were as follows, noting that multiple responses were permitted:

- 26 (29 %) reported clinical coders
- 31 (34 %) reported administration officers
- 48 (53 %) reported health information or medical record managers
- 33 (37 %) reported statistical assistants
- 36 (40 %) reported medical officers
- 22 (24 %) reported nurses
- 10 (11 %) reported other professional groups (including nosologists, statisticians and technical personnel)

In 54 (60 %) countries, coding is manual through the use of coding books. 34 (37.8 %) countries use a combination of manual coding and coding software, and 5 (6 %) countries use coding software exclusively.

3.3 Health Interventions

Taking stock of health interventions is important because they reflect what the health systems actually deliver. To achieve this goal one needs a relevant framework. There is, however, no international classification of health interventions. WHO's International Classification of Procedures in Medicine (ICPM) which was developed in 1979 is now out of date.

The need, however, still exists as evidenced by the fact that individual countries have produced their own classifications of interventions. These include, for example, the Canadian Classification of Interventions (CCI), variations of NOMESCO's Classification of Surgical Procedures (NCSP), the French Classification Commune des Actes Médicaux (CCAM), the UK classification of interventions (OPCS-4) and the Australian Classification of Health Interventions (ACHI). In the US, the ICD-10 Procedure Classification System (ICD-10-PCS) has been prepared as a replacement for ICD-9-CM, Volume 3 and for use in conjunction with ICD-10-CM.

Walker and McKenzie also report a greater diversity in the use of classifications in procedure coding than in diagnosis coding [Walker 2004]. ICD-9-CM is used in 20 (22 %) countries, and ICPM is used in 19 (21 %) countries. 26 (29 %) countries use other classifications including combinations of classifications at different sites, and 20 (22 %) countries either do not code procedures or did not answer that question.

Most national classifications serve reimbursement schemes and/or case mix (diagnostic related) groups which are basically driven by the need to explain health care expenditures. These classifications are particularly focused on medical or surgical interventions and mostly on inpatient care. Given the local developmental base and that their content and style are different, international comparisons are not possible.

There is a gap within the WHO suite of classifications to address health interventions. Such a classification is needed to compare international data and to respond to the needs in countries without an intervention classification of their own for use in conjunction with other members of the WHO Family of Classifications (WHO-FIC).

Given the recognition of this need, the WHO-FIC network agreed to produce and maintain an internationalized version of the Australian Classification of Health Interventions as a starting point towards the International Classification of Health Interventions (ICHI). ICHI is currently in its beta version for field tests that will elicit the feasibility and utility with respect to the needs of the users in different parts of the world.

Currently it is a summary list (e.g. like the ICD Special tabulation lists for mortality) and there may be a need for an expanded and detailed classification of health interventions. A flexible data base may allow zooming in and out between the large groupings for international comparisons and a detailed finer granularity for national modifications. The international comparisons need not be as detailed as the national versions. For example, at the international level there may be a desire to know how many caesarean sections or hysterectomies are done, but less of a desire to know the very detailed procedures.

3.4 Health Outcomes

As a successor to the International Classification of Impairments, Disabilities and Handicaps (ICIDH), and after ten years of successive field tests, the WHO published ICF at the end of 2001. It was developed in multiple languages and published simultaneously in the six WHO official languages: Arabic, Chinese, English, French, Russian, and Spanish. Another 28 language versions have been prepared since 2001.

ICF has an appeal for a much needed common framework in this field. However, implementation needs more than just a conceptual tool. The spectrum of ICF use is very large and includes surveys, censuses and statistics, health care, social security, education, labour and many others. Work is required to ensure information sharing, development of educational materials, calibration of assessment tools, evaluation of use in the field, and overall, the development of a dynamic implementation strategy.

3.5 NOMESCO Statistics

The Nordic Medico-Statistical Committee (NOMESCO) was set up in 1966, following a recommendation by the Nordic Council. When the Nordic Council of Ministers was established in 1978, the Committee was made a permanent statistical committee with separate funding and was placed under the auspices of the Nordic Committee on Social Policy.

Since 1979, the Committee has had a permanent secretariat in Copenhagen, since January 2002 administratively connected to the Danish National Board of Health. The main aim of the activities of the committee is to ensure comparability among the Nordic countries in statistics on health, diseases, mortality and expenditure in the health sector. This is done partly by working on the establishment of a uniform basis for definitions, registration and delimitation of individual areas, so that the results can be reasonably compared between countries.

The results are published by NOMESCO through its main yearly publication *Health Statistics in the Nordic Countries* as well as in special publications such as *Nordic/Baltic Health Statistics* (with the third publication in 2004) and with supplementary information on the Committee's homepage (<http://www.nom-nos.dk/>). The tables and figures in the statistics are only roughly commented, which means that no real analysis is done. Each year a theme section on actual topics is selected and special areas of health statistics are studied more thoroughly.

Theme sections from 1996 to 2003 have been:

- Cardiovascular diseases (1996)
- Occurrence and survival of low birth weight births in the Nordic countries 1980-1994 (1997)
- Patients, diagnoses and activities in general practice in the Nordic countries (1998)
- Cause-of-death statistics (1999)
- Validity and comparability of Nordic hospital discharge Statistics (2000)
- Medicine consumption in the Nordic countries (2001)
- Validity and comparability of Nordic day surgery statistics (2002)
- Equal access to care (2003)

3.5.1 Conclusions of the 2000 NOMESCO Theme Report

The aim of the NOMESCO work on Nordic health statistics is to enable comparisons of morbidity and utilisation of care between the Nordic countries. There is a demand for comparative health statistics and the Nordic health statistics publications are important in this context.

The examination of a few selected diagnoses indicated that there are several reasons for variation. It is often necessary to use supplementary data in order to clarify whether the differences observed in published data are due to misprints, reporting or registration errors, organisational differences, or variation in medical practice or in the occurrence of disease.

3.5.2 The Need for Validity Studies

This overview has indicated the general uncertainty surrounding the application of the classifications of diseases and procedures within the Nordic countries. Random sampling of discharge data and medical records for quality control is not being done on a regular basis in the Nordic countries. Quality control of the assignment of diagnoses and coding requires access to complete medical record data for individual patients and discharges as well as professional medical reviewers, and is therefore very demanding. A complementary approach is prospective studies using standardized case descriptions to be diagnosed separately in the different countries.'

3.6 Quality Indicators and Quality Registers

Good care, which means care of high quality and on equal terms for the whole population, is the ultimate quality goal for all health care and medical services. There is a need for systems for planning, implementing, following up and continuously developing quality in those activities.

Quality indicators form an important foundation in the continuous process of quality work in health care and medical services. The concept of quality can be looked at from three closely related perspectives: quality in terms of patient satisfaction, quality as experienced by health-care professionals and quality as experienced by the organisation.

3.6.1 Quality Indicators

There are a large number of different definitions of the term '*quality indicator*'. These definitions are often linked to the purpose and areas of the use of the indicators. As defined by the Swedish National Encyclopaedia, a quality indicator is a measurable property or circumstance that articulates the level of quality in an activity. In health care examples include the risk of bedsores and the risk of injuries due to falls. With the aid of quality indicators, areas can be identified which should be studied more closely with regard to causal connections and opportunities for improvement.

3.6.1.1 Indicators of Quality of Care

Health care and medical services must be operated so that they meet the requirements for good care. This means that they have to be of good quality and satisfy the patient's

need for security in care and treatment. These factors are measured in terms of structure, process and results, e.g. from quality registers or the national health data registers. Clinical indicators must be pure measures of outcome or reflect structural and process dimensions in health care and medical services which have a proven correlation with medical outcomes.

3.6.1.2 Overall Health-Related Indicators

Overall health-related indicators are also important for an overview assessment of changes in the health situation, development of disease etc., even if the correlation of such indicators with quality of health care and medical services is not self-evident.

3.6.1.3 Choice of Quality Indicators

The following factors can guide the choice of quality indicators:

1. *Relevance.* The quality indicator should focus on health care with sufficiently large volume, i.e. large patient groups, resource-demanding areas and/or observed differences in practice and quality.
2. *Validity.* Account must be taken, in processing and interpreting, of the extent to which the indicator actually measures the quality of the care and reflects differences in patient composition and organisation. Process and structural measures should be evidence-based.
3. *Measurability.* The indicator must be formulated so that it can measure in practice and be clearly and unambiguously defined.
4. *Possibility of influencing.* A quality indicator should measure something which can be influenced e.g. by efforts in medical services. To take an example, bedsores in principle can be eliminated with the right care. On the other hand, average life expectancy can only be explained to a small extent by the efforts of medical services.
5. *Interpretability/unambiguousness.* The measurements of outcome must be possible to interpret in an unambiguous manner: What is good and what is poor? It is important to indicate for individual quality indicators how high the bar should be set. The baseline should be evidence-based and rely on current knowledge and good medical practice.

3.6.2 Quality Registers

In Sweden there are today over forty quality registries, which either have achieved or are in the process of achieving nationwide coverage. They were started up by representatives of the medical profession and established to support efforts to improve the quality of clinical work. Their purpose is to support learning and development and they are not intended for supervisory or similar purposes. The registry managers are distributed among a variety of hospital departments administered by many different health authorities. In most cases, the development from a local to a national registry has taken place gradually.

The key to all quality management is to follow up your own results. When a national quality registry is established, this is the result of a consensus in the medical speciality



concerned on important concepts and quality indicators and a conviction that the registry provides a quality measurement tool based on these indicators. This tool may then be developed and refined from year to year. Individual departments that participate in one or more registries can thus follow up their own results and see how patients benefit from the care and treatment they receive.

3.6.3 Performance Indicators

There is ongoing work in for example England, Canada, Australia, USA and Norway with 'performance indicators' and 'performance rating'. The health agencies have actively been working with indicators about access to health care, effective delivery of appropriate care, efficiency of care and health outcome.

More information on this topic is available on the Web

- England: www.dh.gov.uk
- Canada: www.cihi.ca
- Australia: www.health.gov.au

4 The Hospital Data Project

The Hospital Data Project (HDP) is a project of the European Union Health Monitoring Programme (HMP). Hospital activity data was one of the first data sets loaded onto the Commission's pilot system for the telematic exchange of health information (HIEMS), and HDP was designated to demonstrate the feasibility of hospital activity data dissemination at the level of raw aggregated data sets and to highlight the very low level of comparability between the national data sets. The project report [HDP 2003] was delivered in June 2003, and this section is an abstract of that report.

4.1 *Main Objectives*

The main objectives were to find a methodology for the collection of comparable hospital activity data across Europe and to produce a pilot data set with focus on inpatients and day cases. The areas of infrastructure (e.g. beds), personnel and outpatients were excluded since achieving comparability on these would require separate projects and alternative approaches.

4.2 *Methodology*

The methodology included inventory of national hospital activity data sets; prior initiatives and current projects; defining coverage (inpatients and day cases with outpatient care, palliative care, healthy babies excluded, and all hospitals with psychiatric, maternity and geriatric patients included); and collection of metadata (i.e. information about the data).

4.3 *Data Items*

The selected data items for inclusion in the common data set (CDS) were: country, year, type of admission, age, gender, diagnosis/external cause/procedure category, numbers of inpatient discharges, numbers of bed days, mean length of stay, median length of stay, and numbers of day case discharges.

4.4 *Organisation*

All member states, Iceland and the WHO participated in the project, and the work of coordination, research, and data collection/validation has been shared between Ireland and the United Kingdom.

4.5 *Systematic Approach*

The difficulties in achieving data comparability in an area recognised to be diverse and complex were well known. Previous efforts to a great extent attempted to harmonise high level indicator definitions irrespective of differences in health systems, coverage and coding between countries. Inevitably, the levels of comparability achieved were very low.

What the project required was a methodology which systematically took into account all the real and potential causes of non-comparability. Essentially, this meant basing the methodology on a detailed inventory of patient level data in each participating country.

In addition, full account was taken of past and current work in the area of hospital data and of new work on the functional specification of health systems developed by the OECD in the context of health accounting. Account was also taken of current work in the area of health indicators such as other HMP projects and, in particular, the work of the European Community Health Indicators (ECHI) project.

A special expert group was commissioned to develop a new diagnosis shortlist appropriate for hospital activity and based on ICD-10. Transformations from ICD-9 into the new shortlist were also developed. A selection of hospital procedures was also agreed in order to collect test data in this area, although it was recognised that given the diversity of procedure coding systems in use throughout Europe an additional project would be required to attempt to achieve comparability of procedure data.

4.6 Results

The methodology has been developed and implemented. The CDS has been specified and data have been collected in the required format from 16 countries. This includes both numeric activity data covering inpatients and day cases as well as detailed metadata describing national data sets and potential causes of variation from common definitions. The constituent countries of the United Kingdom are treated separately in the CDS. While all countries but Spain submitted data, there were difficulties in the specification and formatting of data from Greece and Denmark which were not possible to resolve within the time constraints of the project.

The expert group set up to devise an ICD-10 based shortlist for hospital activity also achieved its aims. The shortlist was implemented in all countries. Following further testing with data from other countries, the list may be recommended for international use at the autumn 2003 meeting of the Heads of WHO Collaborating Centres for the Classification of Diseases.

4.7 Conclusions

The project should be seen as the essential first step in a continuing process of making available and improving comparable hospital activity data throughout Europe. Many challenges remain. These include addressing the numerous areas of health system variation and differences in coding and definition which continue to mask true differences in hospital activity. They also include investigating the possibility of extending the approach and the Common Data Set to time series data, to candidate countries, to other categories of patients (eg. outpatients), to measurement of patient mobility, to look at sub-national data and to facilitate the development of output and performance indicators.

WHO is of the view that further development of this work may facilitate the adoption of a standard hospital discharge data reporting format for use by various international



organisations. This would help to harmonise hospital discharge data currently being collected by Eurostat, WHO and OECD and would remove the unnecessary burden caused by countries having to report the same data in different formats to different international agencies.

5 European Development in Hospital Statistics

There is presently an active development in several organisations in Europe in the area of comparative hospital statistics aiming at a harmonisation of routine data collection and analysis.

5.1 Hospital Data Project, Phase 2

The EU Hospital Data Project has got approval for a continuation in a second phase (HDP2) that will start in 2005. Among the questions that the extended project will be working on is the list of ‘sentinel procedures’ that needs further consideration.

5.2 Eurostat Technical Group

Eurostat—the Statistical Office of the European Communities—has a technical group called Care, which has proposed a Minimum Data Set on hospital statistics.

5.3 Eurostat Task Force on Hospital Statistics

An ad hoc Task Force on Hospital Statistics has been established for this purpose and met in April and September 2004 and is planning for another meeting in October 2004. The recommendations of the HDP on which data should be collected have been tentatively accepted by the Task Force.

An interesting development in the discussions of the Task Force has to do with the data collection method. There is a proposal asking member states to hold data in a common format from which a number of tables can be requested by Eurostat and other international organisations such as OECD and WHO. This would ease the burden on national staff in providing data to different organisations and at the same time give desired flexibility to the data collection.

The Task Force is also discussing the shortlists for diagnoses and procedures suggested by HDP. Some people in the Task Force have felt that the HDP shortlist for diagnoses is too long.

5.4 Eurostat Working Group on Public Health Statistics

Eurostat also has a Working Group on Public Health Statistics, which has set up a project aiming at a morbidity list for Europe. The focus of this project is on a list of groups of diseases which can eventually form the basis for regular reporting across the EU. The aim is to use the best source or combination of sources to provide estimates of incidence and/or prevalence (depending on the specific disease and availability of data). In many instances, hospital data will be an important source, but for others primary

care data or multi-source registry information will be the key to obtaining the best available estimates.

5.5 OECD

OECD has written comments to a range of the proposals for the data collection being discussed by the Task Force with an aim to try to harmonise as much as possible the Eurostat and OECD routine data collection in these areas. These comments include a detailed comparison of the diagnostic shortlist proposed by the HDP and the Task Force and the one presently being used by OECD. There are also comments on the HDP1 shortlist of 'sentinel procedures' in relation to OECD's and NOMESCO's current data collection.

5.6 WHO Regional Office for Europe

The WHO Regional Office for Europe has for several years been considering to start regular annual collection of detailed hospital discharge data from member states in its geographical area, which is wider than the EU countries. WHO/EURO is planning to test such data collection, even with more detailed diagnostic categories than the HDP shortlist. With its broad responsibility as an international technical agency, WHO has of course an interest in standardized international data collection and is therefore also taking part in the discussions in the Eurostat Task Force.

5.7 WHO-FIC Hospital Data Working Group

The Hospital Data Working Group, a subgroup of the Family Development Committee, was set up at the WHO Family of International Classifications (WHO-FIC) Network Meeting WHO-FIC Brisbane Meeting in 2002. This group met at the 2003 Meeting in Cologne, when it was decided to collect data from some non-European countries to be added to data already available for fifteen European countries through the HDP.

The aim was to test the data collection process and analyse the utility of the shortlists for diagnoses and surgical procedures. The progress of the work has been slow, however. Until October 2004 only Canada has been able to provide data, now being merged with the HDP data set for common analysis. At WHO-FIC Reykjavik network meeting in October 2004 optimism was expressed for further submission of data from Australia within the coming weeks and the United States in the next few months.

The Workgroup discussed three tasks of importance to progress the work of the Workgroup:

- 1) Analysis of the pilot data of the 15 European with the addition of data from 2-3 non-European countries. This will be methodologically important work as people are not aware how national coding practice influences international comparability.
- 2) Representatives of the Working Group should participate in ongoing discussions of OECD, Eurostat, and the WHO Regional European Office. These



organisations are interested in continuing a dialogue with the Working Group on the refinement of the diagnostic short-list that has been developed as part of the work of the EU Hospital Data Project. The list now contains 130 items, and OECD has expressed a desire to shorten the list.

- 3) The EU HDP has identified 18 procedures for its short-list. While the EU report states that the list is not the final solution, it is a good way forward to test data collection process.

The next steps for the Working Group are to continue the dialogue with HDP and OECD, continue to evaluate data and convene its next meeting in Tokyo in October 2005.

6 Validity and Reliability

Unfortunately, coded data may be of low quality with poor validity and reliability, where the former measures correctness and the latter measures agreement. Validity as a property of measurements—here the outcome of coding—reflects the degree to which the measurement replicates the true status of what is being measured. It is also referred to as accuracy and trueness. Reliability, which reflects the degree to which repeated measurements on the same subject give the same result, is also known as precision, repeatability and reproducibility. It is usually studied as inter- and intra-rater reliability, which involves the fact that coders coding the same medical problem do not always use the same code, nor does the single coder coding identical problems. Furthermore, reliability is sometimes measured through its converse characteristic variability, that is, variation in the results of repeated measurements.

6.1 *Error Types in ICD Coding*

A study of the quality of ICD coded diagnoses [Surján 1999] highlighted the conceptual difference of clinical diagnosis (a very brief summary of physician's knowledge of a patient's health condition) the disease (a category of medical science) and conceptual units of a classification, like the ICD (classes of diseases comprising one or more diseases). There is a complex logical process leading from the clinical diagnosis to the ICD codes. It is important to see that due to the mentioned conceptual differences the number of relevant diagnoses of a health care episode does not agree necessarily with the number of appropriate ICD codes. Analysis of the mentioned logical process led to the understanding of some typical coding errors and their sources.

6.1.1 **Diagnosis Overlooking**

It may occur that some of clinical diagnoses of a patient do not appear in the discharge report. Physicians are highly specialised nowadays, they focus more and more on their own specialty (or subspecialty), while forgetting to document appropriately some comorbidities (typically chronic diseases) not belonging to their field. The entire patient record or the descriptive part of the discharge letter more likely will refer to such conditions than the explicit diagnostic statements.

6.1.2 **Errors of Induction**

The clinical diagnosis represents a lower level of abstraction than the disease, and the number of diagnoses does not necessarily match the number of diseases. An induction process is necessary to find all and only relevant diseases for a set of clinical diagnoses used to describe a certain clinical case.

In this induction phase the following errors may occur:

- The disease is confused with one of its more or less characteristic symptoms, e.g. 'meningeoma' might be described as 'serious headache'
- A single disease is described by more diagnoses. This may occur if some of the symptoms are more serious for the patient than the disease which causes it. By mistake such symptomatic diagnoses can be regarded as different diseases.

- Over-specification. The clinical diagnosis might be extremely specific. E.g. ‘atticitis’ is inflammation of the atticus, a minute compartment of the middle ear. Atticitis is a form of otitis media. Such very detailed concepts are not intelligible for most physicians, but for specialists. Such diagnoses are likely indexed to a generic class, such as ‘other disorders of ear’.

6.1.3 Indexing Errors

Indexing errors occur when all and only relevant diseases are identified, but their mapping to ICD categories is incorrect. Indexing errors also may occur in combination with induction errors, in which case it is difficult to say in which phase of the process went wrong. Indexing errors may happen due to disregarding the ICD indexing rules as well.

6.1.3.1 Pure Indexing Errors

If B is the set of the assigned ICD codes for a case (a given set of clinical diagnoses), and C is the set of the correct ICD codes for that case. An indexing error is present whenever $B \neq C$.

Three types of error are possible:

Type 1. Error of omission: B is a proper subgroup of C

Type 2. C is a proper subgroup of B

Type 3. There are codes $b \in B$ and $c \in C$ where $b \notin C$ and $c \notin B$

Type 1 error comes often from laziness or the lack of attention of the coder. There is a special case of type 1 error, where a disease that should be represented by two or more ICD categories but instead is assigned only to one. Type 2 error is accidental. Type 3 errors are very common. This is the case in which a ‘incorrect’ code appears instead of the ‘correct’ one. The most important cause of this error is that the semantic content of the categories is not defined exactly.

6.1.3.2 Indexing Errors in Combination with Induction Errors

Let us suppose a case where the patient has a single disease which has to be indexed to two ICD codes while the patient record mentions three different clinical diagnosis, all expressing a different aspect of the same disease. It is essential to understand, that in such cases none of the ICD categories must correspond directly to any of the given diagnoses.

For example consider a patient presenting himself for shortness of breath due to a recurrent laryngeal cancer with metastases in the regional lymph nodes. Suppose that he previously received radiotherapy. The following clinical diagnoses are appropriate:

Dg1. Recurrent laryngeal cancer with cervical metastases left side (T3N2M0).

Dg2. Dyspnoea

Dg3. Status post radiotherapy

We have to induce just to one disease, i.e. laryngeal cancer. The most specific ICD category, C32.9 (malignant tumour of the larynx), does not represent sufficiently the extension of the process, so C77.0 (secondary malignant tumour of lymph nodes of the head and neck) has to be added. These two ICD codes refer to one single pathological

process, i.e. to one disease. There is a code in ICD (R06.0) for dyspnoea, and also for history of irradiation (Z93.2). However, the indexing rules of ICD suggest using R06.0 only if the cause of dyspnoea is not known and Z93.2 only in cases where the cause of medical service was the history of irradiation. At least this is in our own interpretation of the ICD rules. Various interpretations of these rules are important causes of disagreement between different coders.

6.1.3.3 Coding Rules of ICD

ICD contains several coding rules which are described in the second volume of the printed version of ICD-10. Not being aware of these rules one can cause indexing errors, e.g. missing cross reference (the ‘dagger’ and ‘asterisk’ cross reference system describes the etiological and anatomical aspect of the diseases), or a generic category is used when there is a more specific category.

6.1.3.4 Labelling Errors

Violation of the indexing rules pertaining to the selection of the primary disease leads to labelling errors. By labelling it is meant that the relation of a condition to a given health care episode has to be expressed by labelling each code as primary condition, underlying disease, complication, co-morbidity etc. In addition to ICD rules, many countries have some special regulations for this labelling, and these rules regulate the indexing step.

Usually the regulations allow one and only one principal disease—in accordance with the original ICD philosophy. An improper selection of principal disease frequently occurs. Whenever one condition has to be expressed with more than one code, additional rules are required to decide which one of those codes should be labelled as the primary condition, and ICD rules regulate this question only partially.

6.2 *Potential Sources of Errors*

Figure 1 below illustrates the whole coding process as a series of subsequent abstractions, together with the actors who carry out the process and the known sources of the mistakes. Since coding is a human—or dominantly human—activity the sources of errors are also humans.

The first source of documentation errors is the physician him- or herself. Physicians are the primary users of medical documentation, and therefore one could think that they have the utmost interest in its quality. On the contrary, the fact is that many doctors are convinced that documentation is just a burden and that the quality of primary medical documentation is not always satisfactory. Especially, physicians have no interest in coding issues, so they likely disregard coding aspects from their documentation work.

The second source of error is the person doing the indexing. This person is either a physician or some specially trained health professional or a clerk. The personal errors in the classification phase highly depend on organisational and even on logistic issues. The most important cause here is the unawareness of the subsequent induction steps, skipping one or more intermediate levels.

From the non-human factors, the most important one is the ICD itself—more precisely the inconsistent structure of ICD, the lack of definition and the many deictic notions. A notion is deictic if its meaning can not be entirely derived from its description; the correct interpretation depends on other descriptions. The ‘not elsewhere classified’ categories of ICD are typical examples. To know what it means we have to know all the elsewhere classified notions.

Another important non-human factor is the language of ICD which is usually different from the clinical jargon. This makes difficult to map the clinical terms to names of the ICD categories.

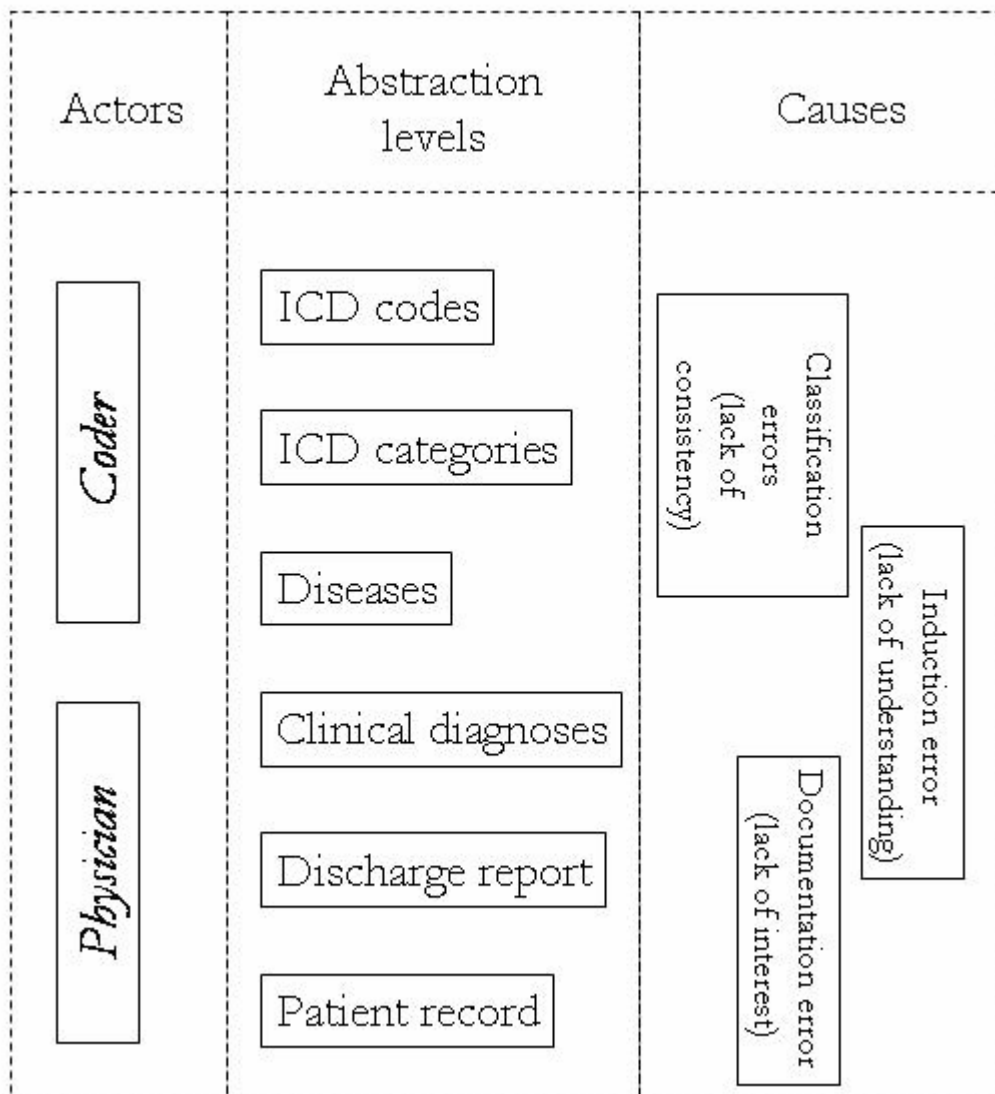


Figure 1. The coding process with actors, abstraction levels and error types.

6.3 *Validity in Mortality Coding*

One area that can serve as an example of how validity in coding is measured is the Swedish mortality coding. Death certificates are coded by trained coders who use computerised tools and can request supplementary information from the issuing unit as well as confer with a medical expert in unclear cases. Thus it can be seen as a kind of best practice. However, it should be observed that the major source of uncertainty is the establishment and reporting of the cause of death, which precede the coding.

In summary, the process can be described as follows. The death certificates are sent to a specialised unit at the National Board of Health where they are scanned and the information, which is written in plain language, is registered electronically. The registered diagnostic expressions are coded with the help of the software tool MIKADO (Multiple Coding of Causes of Death) developed at Statistics Sweden, and the underlying cause of death is chosen with the help of ACME (Automated Classification of Medical Entities) from the US National Center for Health Statistics. [SoS 2004]

The error rate of the coding of underlying cause of death is monitored. Every second year or so, a subset of about 2.5 % of the death certificates is randomly selected and independently re-coded for comparison and error tracking. Coding errors stem from registration, coding (MIKADO) and the selection of underlying cause of death (ACME). In 1998 the percentage coding error was estimated to be [SoS 2004]:

- 1.2 ± 0.3 on chapter level
- 2.2 ± 0.5 on block level
- 4.0 ± 0.6 on three-character level
- 6.2 ± 0.7 on four-character level
- 6.3 ± 0.7 on five-character level

6.4 *Reliability in Morbidity Diagnosis Coding*

Validity and reliability are interrelated. Reliability is not a sufficient condition of quality, but it is a necessary one; if repeated measures are uniform, there is a possibility that ratings are valid. This association is used when a true score is difficult or costly to obtain.

When comparing results from coding quality studies it is necessary to consider the nature of the material presented to the coders—where the coding starts—i.e. how many levels of abstraction in Figure 1 that must be passed before the codes are reached. High interpretive burden is associated with low reliability. For example, the validity numbers presented in section 6.3 refer to the top levels, while the reliability numbers in section 6.4.1 concern all levels.

6.4.1 *Swedish Primary Health Care*

In a Swedish study comparing three coding tools, six general practitioners each coded 152 medical problems presented as medical record extracts on paper [Nilsson 2000]. These were coded according to the primary care version of the Swedish translation of ICD-10. The reliability, which was measured through the percentage level of agreement

and Cohen's kappa, was studied on code level and on aggregated levels; presented here are the percentages.

The classification consists of 972 diagnoses and diagnosis groups brought together in 20 out of 21 ICD-10 chapters (that on external causes is excluded). The inter-rater agreement at code level was 54, 57 and 59 % depending on the tool used. At ICD chapter level, the agreement was 80, 81 and 84 %.

6.4.2 British Diagnosis and Procedure Coding

In a British study, the reproducibility of clinical coding in two National Health Service hospitals was evaluated [Dixon 1998]. In a random sample of 1,607 admissions, codes for main diagnosis (ICD-9) and procedure (OPCS-4) assigned by local staff were compared with those assigned independently by an external team. The main outcome measure was the level of exact agreement (complete code) and approximate agreement (first three positions of the code).

Concerning main diagnosis, hospital A had an exact agreement of 43 % and approximate agreement of 55 %. The figures for hospital B indicated better agreement: 60 % and 72 %, respectively. In both hospitals, agreement was higher for specific diseases. For example; the approximate agreement for asthma was 86 % and 91 %, and for fractured femur it was 84 % and 89 %. The reliability of procedure coding was higher than that of diagnosis coding. The exact agreement in hospital A was 58 %, while the approximate agreement was 70 %. Corresponding numbers for hospital B was 76 % and 83 %.

6.4.3 Case Histories Coded in Four European Countries

Steinum reports on a pilot study where 13 abstracted case histories were distributed to 180 physicians in clinical hospital practice in four European countries [Steinum 1997]. The histories appeared as they might have been presented in a daily care situation, with data of varying relevance to the case. A few of them were quite clear cut, but most posed certain problems of diagnosis choice. No cost information was given. The physicians were invited to assign diagnoses and corresponding ICD-codes as they would do in a real clinical discharge situation.

A total of 1,363 responses were given to the 13 case histories. A few respondents submitted procedure codes to certain cases, but these did not qualify as operation room procedures, and had no bearing on the grouping results. Diagnoses were checked for validity. In some cases only diagnoses in text were given, and in these cases the authors have submitted the corresponding ICD-9-codes. The cases were finally grouped by the author according to Medicare DRG (HCFA version 12).

The 1,363 responses were distributed into 58 DRGs. Overall there was a substantial variation in the choice of diagnosis, although the variation differed from case to case. This variation in diagnosis choice was reflected in DRG assignment and case mix index. Similarly, estimated cost per case varied substantially. The mean values of cost estimations, however, were quite stable.

For seven case histories, exemplified by case history 7 in Figure 2, DRG assignment was quite homogeneous, as more than 90 % of observations resulted in a DRG pair (with/without complication and comorbidity). The remaining six case histories showed a wide diversity in DRG distribution due to diversity in choice of principal diagnosis. In case history 2 (see Figure 3) the respondents presented 13 different principal diagnoses on three character category according to ICD-9. These 13 diagnoses represented five different diagnostic chapters, and corresponding DRGs belonged to four different major disease categories (MDCs).

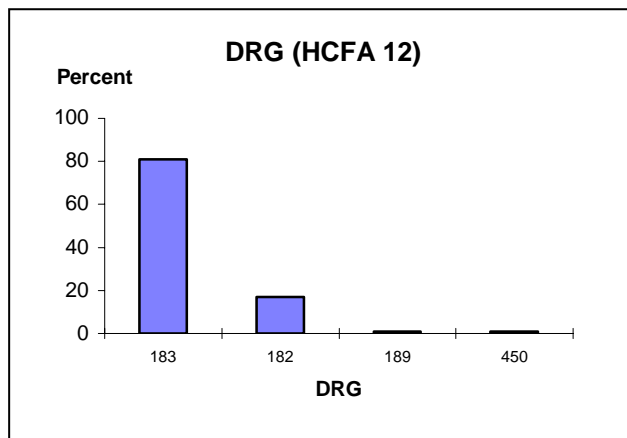


Figure 2. DRG distribution of case history 7.

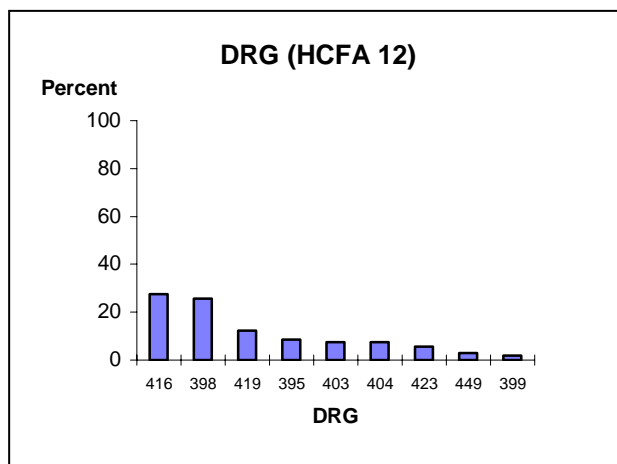


Figure 3. DRG distribution of case history 2.

6.5 Impact on Coding: 'Who Is Coding'

There are two main groups of professionals involved in morbidity coding. One group consists of clinical coders, administration officers, medical record managers or statistical assistants. The second group are medical doctors. According to the WHO survey described in section 3.1-3.3, the first group are coding between 29 and 53 percents and the second group about 40 percents (note that multiple answers are permitted).



In Australia, Great Britain and United States there are clinical coders who are coding the diagnoses for inpatients. With use of the physician's formulation of the diagnoses at discharge, they code out of the written documentation about the patients condition, findings and treatment during the hospital stay.

In the Nordic countries there is the physician treating the patient that have the responsibility to formulate the diagnoses, but also to code the diagnoses. The thinking behind this is that that the practitioner diagnosing and treating the patient also is the one that have the best opportunity to find the corresponding code in the classification.

However, the disease classifications have been changed a lot during the years with regular international revisions by the WHO. The classification is more complicated and a huge system of coding rules have been formalized in ICD-10, volume 2, but also a lot of notes, inclusions and exclusions in the tabular version of ICD-10 (volume 1) and to some extent also in the alphabetic version (volume 3). This implies that education is needed to learn about all the rules. Other staff than the physicians are more often educated and trained. For example, in Norway and Sweden medical doctors get almost no education at all in the use of ICD-10 for coding, despite their mandatory responsibility for coding of all discharges from hospitals, which has been effective for many years.

When the physician is diagnosing, he or she considers all the formalized information about the patient, but also the personal knowledge about the patient, which can include informal information (what the he or she knows, but is not documented in the medical record). The diagnosis is formulated with the use of linguistic traditions (Latin, English, national language) or local traditions (depending of speciality of the physician, 'we have always done it this way').

Using rheumatoid arthritis as an example, the classification gives the possibility to choose between M05 (Seropositive rheumatoid arthritis) and M06 (Other rheumatoid arthritis). The practitioner can have knowledge since years that a particular patient has a seropositive rheumatoid arthritis, but uses the common expression 'rheumatoid arthritis' in the documentation in the medical record. Consequently the he or she will choose the code M05 based on knowledge about the patient. A clinical coder would choose the code M06 when the documentation about the seropositive condition is lacking.

In another example a patient is treated with antibiotics because of pneumonia. The patient responds very well to the treatment and is discharged from the hospital. The physician is using the diagnostic concept 'pneumonia' and assumes that it was a bacterial pneumonia, as the patient recovered. The coding will be J18.9 (Pneumonia, organism unspecified), and perhaps he or she will not reflect on the fact that the 'specification' is about a microbial specification.

7 Enhancing Classification through Terminology

7.1 *A Finnish Service Classification*

The identification of the different units of social and health care and the services which they provide are essential to locate and order the services needed by customers. The most used tool today is a telephone directory, but a computerised environment that in addition to communication between the units allows data mining would be better alternative.

The social and health care units and their services are registered by the national statistical authorities who also gather and report the services provided by the units on a yearly basis. In Finland, Stakes is responsible for collecting and reporting national data on social and health care units, and e.g. in health care the mandatory dataset is the entity code, service branch, medical speciality and reason for the service and interventions.

In Finland, the project 'The National Registry for Social and Health Care Units' was launched in 2004 as a part of the national patient record project. The project has two subprojects: 'Identification of the Social and Healthcare Units' which is based on the ISO/OID¹- methodology and 'Service Classification'. During the pilot phase in 2005, data will be collected to create a national registry for social and health care units. A similar project (Basisenhetskodeverket, BEK) has been started in Norway in 2004.²

The Finnish service classification project has obtained the names of about 500 different services provided by the public, private or the third sector (non-governmental organisations, NGOs). During the pilot phase, about 80 % of the units providing social and health care services will be given their ISO/OID identification code as well the service data for the national registry. The registry will be stored and updated regularly on the National Codeserver, which is a database for collecting, storing and delivering national terminology systems (code-based information, e.g. classifications, terms and organisational data).

The language of the service classification has been based on the used names of the services in social and health care, which are compositions of terms, e.g. 'surgery polyclinics for children' and 'day care for mentally challenged children'. Moreover, the terminology of the services can be seen as a compilation of attributes specifying the service. However, the terminology has not been harmonized or systematized, and in a following phase, synonyms for the service names will be collected and added to the registry.

¹ ISO/OID (Object Identifier) is an ISO standard (ISO 8824:90) for a specification to have a systematic international coding to identify e.g. objects, persons and units.

² Basisenheter for statistis og økonomisk opfølging i helsetjesten: Definitioner – Egenskaper – Registering. KITH Rapport 24/03



There is a need in statistics to have added value information on the usage of the social and health care resources and the chains of care at local, regional and national levels. Harmonization of the terms describing different services is consequently essential not only for communication, e.g. in consultations, but also for the national statistics and local administration. Furthermore, if services are systematically defined according to an ontological specification, e.g. with a sub- and superordinate hierarchy, the attributes and relationships built into the new classification might make it a more flexible and sophisticated tool to classify, analyse and report the services produced. Compared to the traditional 'list of 500 service names', it is also evident that the attributed service classification will have advantages e.g. in updating and adding new service terms and codes to the coding system.

The pilot project is a collaboration between Stakes and Helsinki University Institute for Information Technology (HIIT), and technically it is a pilot to test if the national social and health care statistics will benefit from Semantic Web technologies in near future (2 years) or in a longer-term perspective (5-10 years). The intention is to use tools such as the Protégé³ editor and the OWL⁴ language to define the attributes and the relationships between services, service needs, individuals needing the services, service providers, financing mechanisms etc. The first phase will be restricted to elderly care, and the main reason to choose this service branch is that it combines many services provided in co-operation with the social care and health care system.

Key questions for the pilot are:

- Can services be better described with Semantic Web technologies than with standard methods used today in national statistics?
- What new possibilities does Semantic Web modelling provide for the input of the information in statistics compared to the current tools?
- What new possibilities does Semantic Web modelling offer in providing and distributing information?
- What new possibilities do Semantic Web technologies offer to master analysis and to report the data with attributed datasets?

Some provisional modelling has already been done with Protégé. The refinement of the generic service model and specification to the elderly care will be done in early 2005. The pilots to get the data in and out from the Codeserver as well the pilot with the search engine will follow in early spring 2005. The first pilot results will be presented during the SemanticMining summer school in 2005.

³ <http://protege.stanford.edu/>

⁴ Web Ontology Language, <http://www.w3.org/2004/OWL/>

7.2 Terminology Systems for Data Aggregation

During the last decades, several medical terminology systems have emerged, and SNOMED CT—the fusion of SNOMED RT and Clinical Terms (Read Codes) Version 3—is one of the latest instances.

Traditionally, classification systems e.g. the ICD classifications have been used as the basis for collecting and aggregating statistical data. These systems have fixed monohierarchical architectures with the indexes built into the codes. SNOMED CT, however, is intended to be used in connection with electronic health records (EHR), and its architecture is polyhierarchical. The structure is not built into the code, but the hierarchical relationships are represented along with other relationships separately. As a result the question arises whether data collected in EHR-systems can be aggregated using the relationships in the terminology.

In a hierarchical terminology system it should be possible to aggregate data to any top or intermediate level. I.e. it should be possible to use each entry as an aggregation point (see Figure 4).

In order to demonstrate that a terminology can be used as described, parts of ICD-10 were mapped code by code to SNOMED CT and data from the Danish National Patient Register from the year 2002 were used as source data. The tables contain ICD-10 coded data from 1.1 million discharged patients (hospital admissions) and 3.3 million outpatients. Examples of data aggregations from the registry using SNOMED CT as an aggregation tool is shown in Figure 5 and 6.

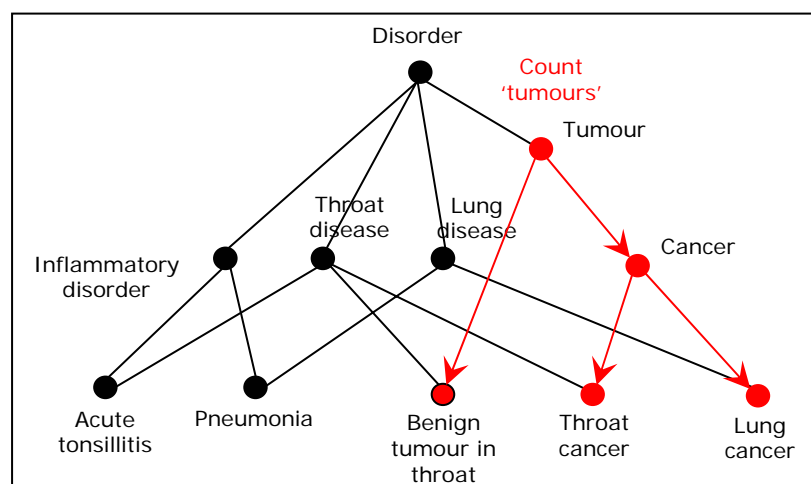


Figure 4. Schematic representation of a disorder axis in a terminology with polyhierarchical architecture. Any entry can be used as an aggregation point.

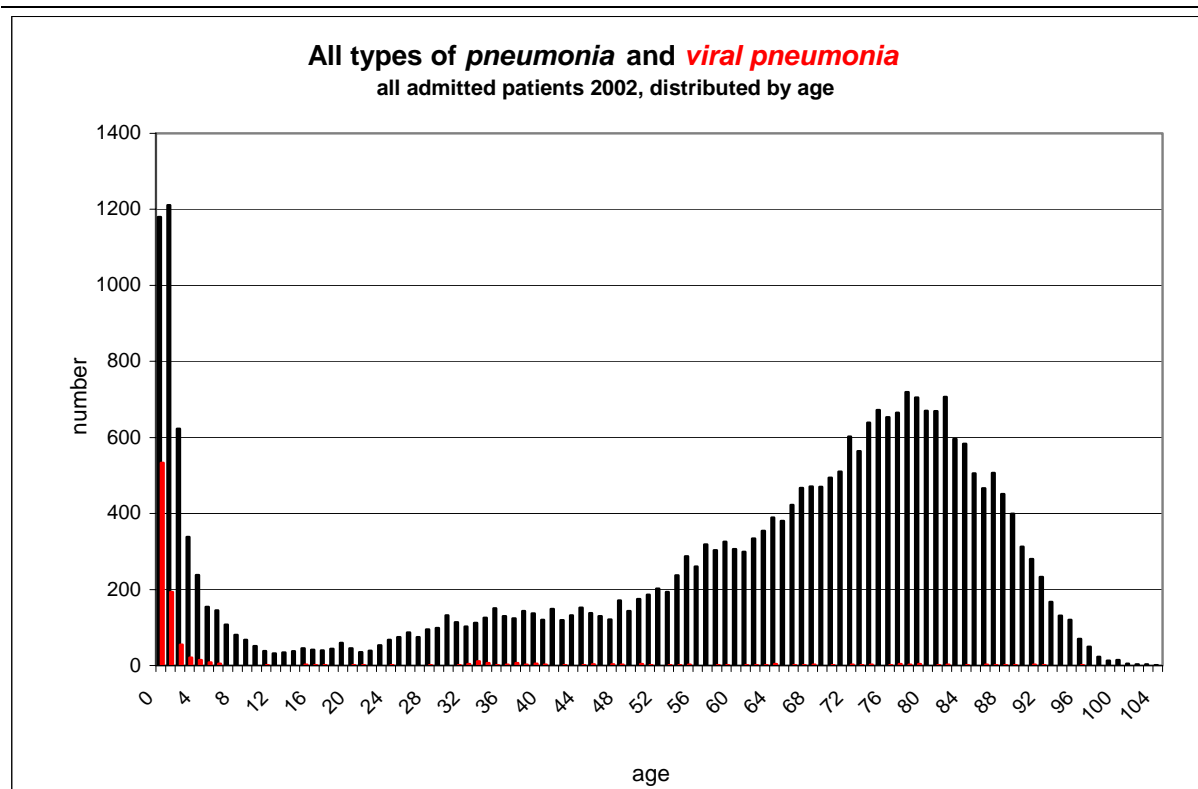


Figure 5. Age distributed data from the Danish National Patient Register aggregated with SNOMED CT.

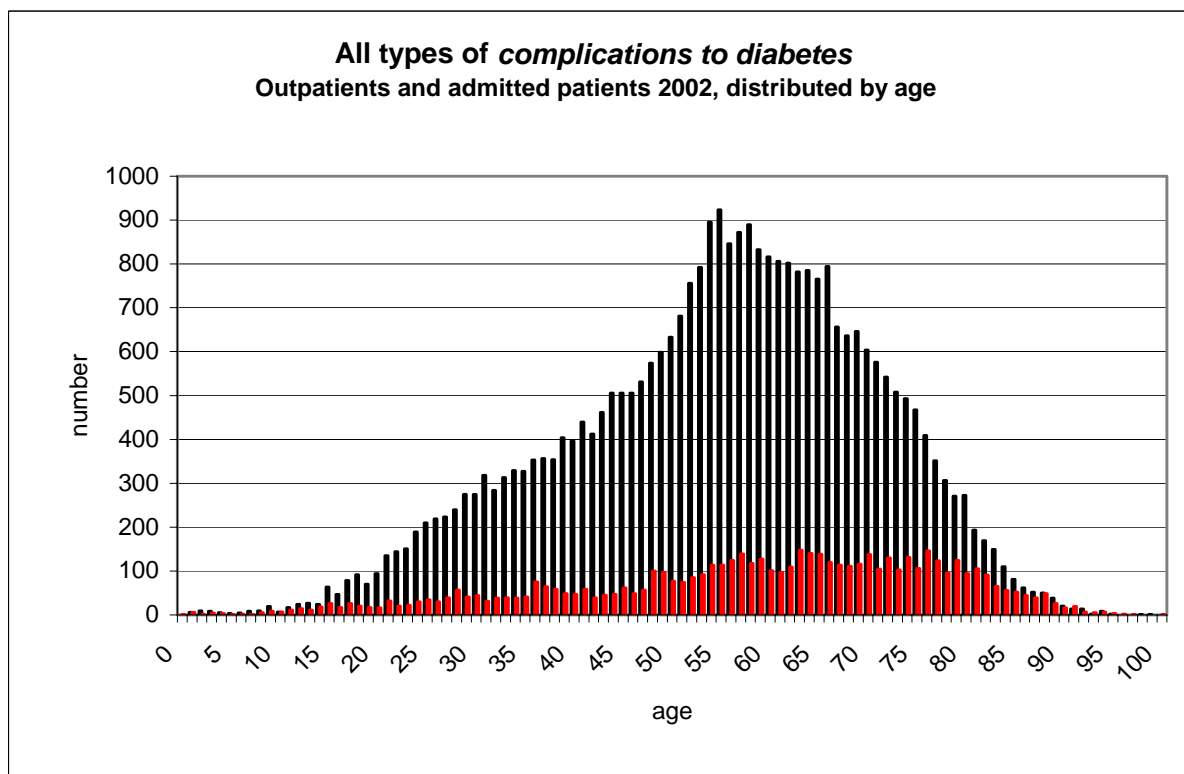


Figure 6. Age distributed data from the Danish National Patient Register aggregated with SNOMED CT. Note that the *complications to diabetes* do not exist in ICD-10.



After mapping classification entries to terminology system entries it seems to be possible to aggregate data from a patient register using the relationships represented in the terminology system. Therefore it should also be possible to use terminologies to aggregate data collected in EHR systems.

However, careful consideration must be given to the mapping between the classification and the terminology system. Conceptual mapping of terms rely on interpretation, and for example, due to classification rules that take context into account, 'pneumonia' in ICD-10 does not necessarily mean the same thing as 'pneumonia' in SNOMED CT.

Using the relationships represented in a terminology as an aggregation tool also poses strong demands of the correctness of the relationships in the particular terminology. However, since the relationships are parts of the definition of the particular concepts, they should also be 'correct' from a terminological point of view. Thus this demand is not different from the demand of correctness in the terminology per se.

References

- Cimino 1996 Cimino JJ. Review Paper: Coding Systems in Health Care. *Methods Inf Med*, 1996 Dec. 35(4-5): p. 273-284.
- Cimino 1998 Cimino JJ. Distributed Cognition and Knowledge-Based Controlled Medical Terminologies. *Artif Intell Med*, 1998 Feb. 12(2): p. 153-168.
- Dixon 1998 Dixon J, Sanderson C, Elliott 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 Mar;20(1):63-9.
- Gersenovic 1995 Gersenovic M. The ICD Family of Classifications. *Methods Inf Med*, 1995 Mar. 34(1-2): p. 172-175.
- Grimsmo 2001 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 Jun;19(2):76-82.
- HDP 2003 Hospital Data Project. Final Report, June 2003 (EU Agreement No. S12.304597).
http://europa.eu.int/comm/health/ph_projects/2000/monitoring/fp_monitoring_2000_frep_09_en.pdf
- Nilsson 200 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 Inf Med* 2000 Dec;39 (4-5):325-31.
- Nyström 1986 Nyström E. Den svenska dödsorsaksstatistikens framväxt och tidiga historia, In: Hälsa, sjukdom, dödsorsak, Nordenfelt L, ed. 1986, Liber: Malmö. p. 107-133. [in Swedish]
- SoS 2004 National Board of Health. Dödsorsaker 2002 [in Swedish with English summary]. Stockholm: Socialstyrelsen, 2004. Available from <http://www.socialstyrelsen.se/Publicerat/2004/8533/2004-42-5.htm>
- Steinum 1997 Steinum O. Variation in Diagnosis Choice on Standardized Case Histories. Paper for presentation at 9th Australian Casemix Conference 1997. Uddevalla, 1997.
- Surján 1999 Surján G. Questions on Validity of International Classification of Diseases-Coded Diagnoses. *Int J Med Inf* 1999;54:77-95.



- Walker 2004 Walker S, McKenzie K. Report on findings of the ICD-10 Coder Needs Assessment Surveys. WHO Family of International Classifications Network Meeting. WHOFIC/04.027, 2004.
- WHO 2004 World Health Organization. Draft WHO Business Plan for Classifications. Geneva, 2004.