Ann. Ist. Super. Sanità
Vol. 27, N. 2 (1991), pp. 195-200

THE LOGICAL STRUCTURE OF THE VIDEOFAR DRUG DATA BASE

B. CAFFARI and R. RASCHETTI

Laboratorio di Epidemiologia e Biostatistica, Istituto Superiore di Sanità, Rome, Italy

Summary. - *The quality of the analyses that can be carried out by a Drug Prescription Monitoring System depends on the completeness and accuracy of the information on drugs. Data quality depends also on the organization of the data base that must be designed to allow higher retrieval power with regard to the information level theoretically contained in stored data. Considering these requirements we developed inside the VIDEOFAR project, starting from the preceding experiences at the Istituto Superiore di Sanità, an activity of design and realization of a Drug Data Base whose structure, both logical and physical, is described.*

KEY WORDS: drug data base, drugs prescription.

Riassunto (La struttura logica del Data Base Farmaci VIDEOFAR). - *Il livello delle analisi che possono essere sviluppate in un Sistema di Monitoraggio delle Prescrizioni Famaceutiche dipende in gran parte dalla completezza e dalla precisione delle informazioni riguardanti i farmaci. La qualità dei dati dipende anche dall'organizzazione del Data Base che deve essere progettato per consentire il massimo potere di recupero rispetto al livello informativo teoricamente contenuto nei dati memorizzati. Tenendo conto di queste esigenze si è sviluppata, nell'ambito del progetto VIDEOFAR, a partire dalla precedente esperienza dell'Istituto Superiore di Sanità, una attività di progettazione e realizzazione di un Data Base Farmaci di cui viene descritta la struttura, sia logica sia fisica.*

PAROLE CHIAVE: basi dati farmaci, prescrizioni farmaceutiche.

Introduction

There is a close correlation between the capacity of a national health service to effectively promote citizens health and the availability of information systems controlling the main information flows (basic information for medical service and optimization of resources).

The pharmaceutical sector is one of the most important aspects characterizing, in terms of quality, the activity of a national health service.

The realization of a Drug Data Base, within the framework of a complex information system like VIDEOFAR, constitutes the basis to support the proper management and diffusion of information about drugs [1, 2].

The "drug prescription" event is a complex and dynamic process (Fig. 1) where it is possible to identify at least four stages:

- characterization of the problem (diagnosis);
- choice of intervention strategy (desired action);
- intervention (drug prescription);
- patient response (compliance).

On the whole the prescription of a drug should be assessed on the basis of the following criteria:

- adequacy: the prescription of the most suitable therapeutic treatment;
- efficacy: production of desired effects;
- safety: avoiding the onset of adverse reactions;
- cost-effectiveness: to comply with medical criteria without overlooking socio-economic parameters.

The above described features cannot be considered as independent from one another. For instance, the adequacy of the prescription basically depends on the adequate correlation between diagnosis and prescription, which obviously entails yet other factors, including training and reliability of scientific information provided to general practitioners [3].

The analysis of such processes presents difficulties caused by the usual lack of adequate information elements that allow, for each "prescription" event, the evaluation of the decision-making phases preceding it.

Nonetheless, the development of "Systems" whose aim is to observe the drug prescription process (Drug Prescription Monitoring Systems - DPMSs) may provide appropriate information to support planning activities; moreover the analysis of the prescription models may single out possible intervention areas [4-6].

Rough data contained in a prescription (physician code, patient code, drug code, dates, etc.) must go through a record-linkage phase to convert codes into information, that can then be subjected to analysis procedures.

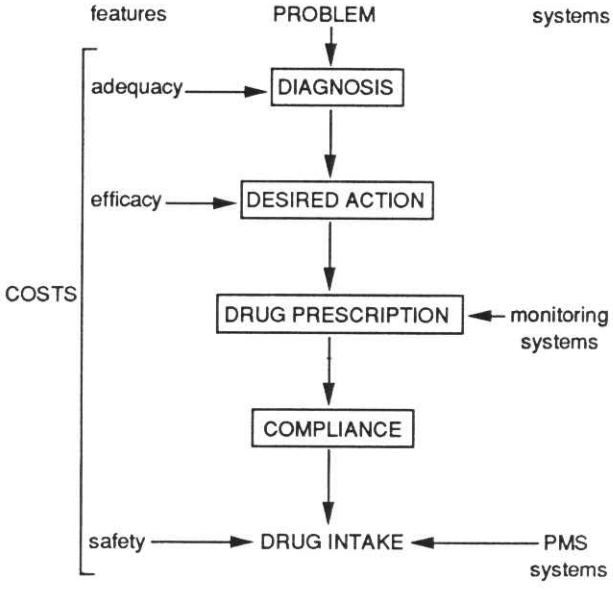

Fig. 1. - Events regulating the prescription process.

The record-linkage process associates each prescription with a highly complex set of information regarding variables such as: substances contained in a drug, therapeutic classification, daily doses, registration in particular formularies, etc. Such data are extremely important to select the analysis criteria.

The quality of the results that can be obtained using a DPMS, largely depends upon the thoroughness and accuracy of drug-related information.

It is moreover important to stress that the value of a Data Base does not only depend upon the information level of stored data. It also depends on the quality of its organization, that must be designed to allow maximum retrieval power with respect to the theoretically available information level in stored data.

The logical structure of the Data Base

The information stored in a Drug Data Base may be grouped according to three different sets:

- the pharmaceutical preparations;
- the classification;
- the substances (active principles) contained in the drugs, according to the general logical organization shown in Fig. 2.

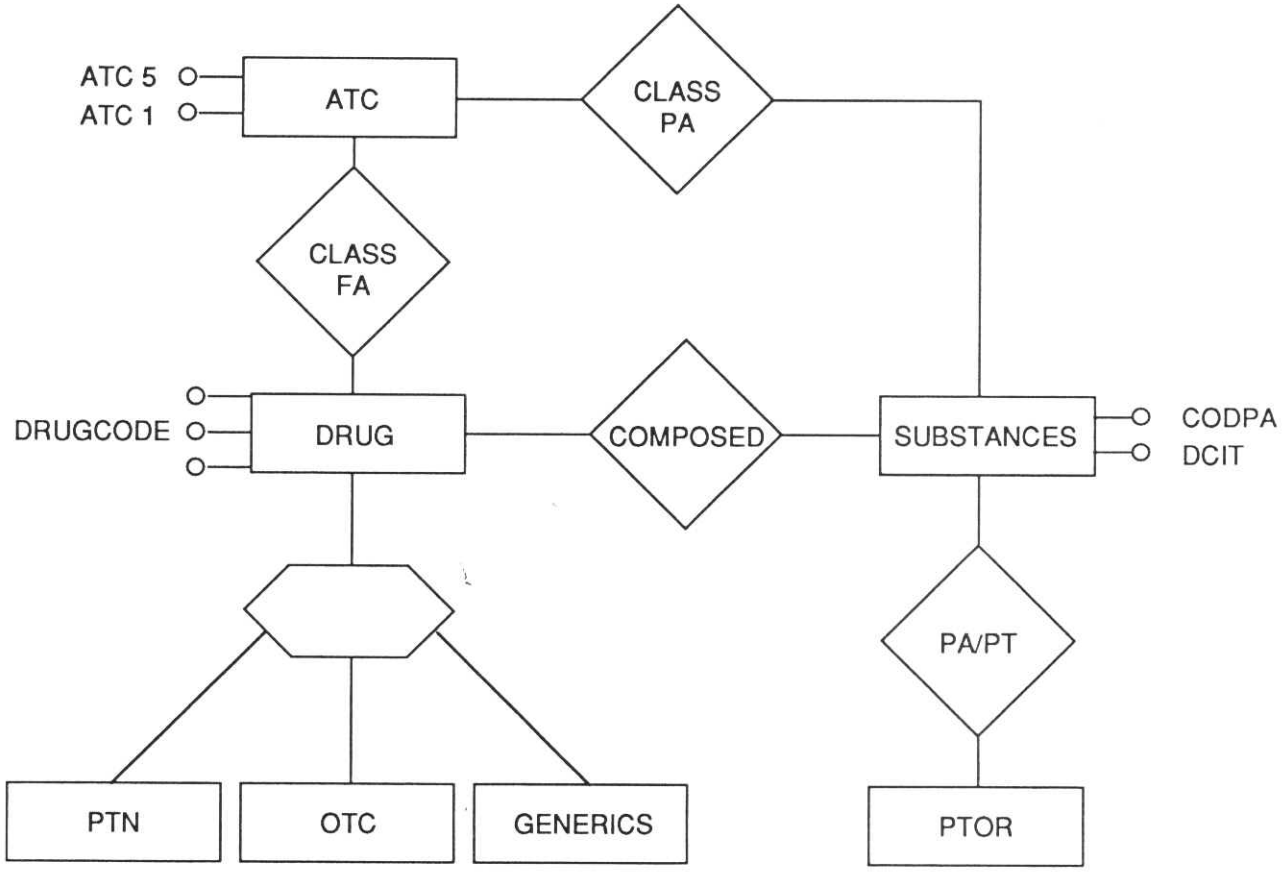

Fig. 2. - The conceptual structure of the data base according to the notation of Chen's entity-relationship model.

The figure should be read in the following way: the Entity "Drug" is a generalization of "PTN" (National Drugs List), "OTC" and "Generics"; it is classified by the "ATC" and is compound by substances that is classified by the "ATC" and is present in "PTOR" (Regional Drugs Formularies).

These initial schema led to the formulation of a logical model based on Codd's Relational model; the various attributes were so organized into a set of relation, the structure of which is shown in Fig. 3.

As may be noted, the different subsets (relations) are conceptually correlated by "common" key values, that define possible search paths, depicted in the figure by means of segments.

The main Relations are the following:

DRUGS: it contains all the information characterizing each individual medicine package. The attributes defined for this relation are as follows: identification attributes, such as code, trading name and description of the product; description attributes, which regard type of package, dosage, type of prescription, registration in the National Therapeutic Drugs list, substance code.

In particular, the type of package is described by a six-character code that hierarchically identifies:

- route of administration (OS, TP, RT, VG, PP);
- pharmaceutic form;
- type

for instance, according to the type parameters reported below:

ORAL	OS	
Capsules	CU	
hard	RG	
soft-pearls	ML	
amylaceous	AL	
micro	MC	
Tablets	CP	
simple		SL
sugar coated pills		RT
gastroresistant		GR
effervescent		EF
sublingual		SG
chewable		MA
delayed regulated release		CA
continued regulated release		CC
sustained regulated release		CO
planned regulated release		CG
discontinuous regulated release		CN
discoid		DS
Pills	PL	

and so on.

CPA: it contains the information regarding all the substances that may be present in medicines. Both single compounds and the associations of several substances are identified by a code (CODPA) matching that adopted in the Drug List. The first four digits of the code indicate the basic substance, thus making it possible to identify the active principles regardless of the different possible salifications. In the case of associations, the base code refers to the different components of the association.

The relation is structured (Fig. 4) to allow both queries of a code (or substance name), and of all the possible compounds in which a particular substance may be present.

The constraint that has to be complied with in the logical organization of the CPA relation, is that each single component of an association in any event must occur in the relation as an individual substance, even when it is not associated with any medicine in the DRUGS relation.

The Group and Subgroup attributes allow to describe, according to a two-tier hierarchical logic, whether the substances belong to particular chemical "families".

All of the substances are described in the CPA by the Italian Common Name (DCIT); next to the DCIT there is an item containing the first letter of the name of the substance followed by the first three consonants, so as to allow for an approximate query (a simplified best match query), when the name of the substance being sought is not known exactly.

Additional names (International Nonproperty Name, IUPAC, etc.) are reported in the IDENTITY relation, and are correlated to the CPA relation by the basic code.

Each association code has a COMP.NO. value >1, equivalent to the number of substances present in an association.

ATC SCHEME, CLASFAR, CLASPA contain the information about the ATC classification.

We have adopted the ATC System because this classification has proved to be oriented towards the characterization of possible drug use, giving priority to the diagno-

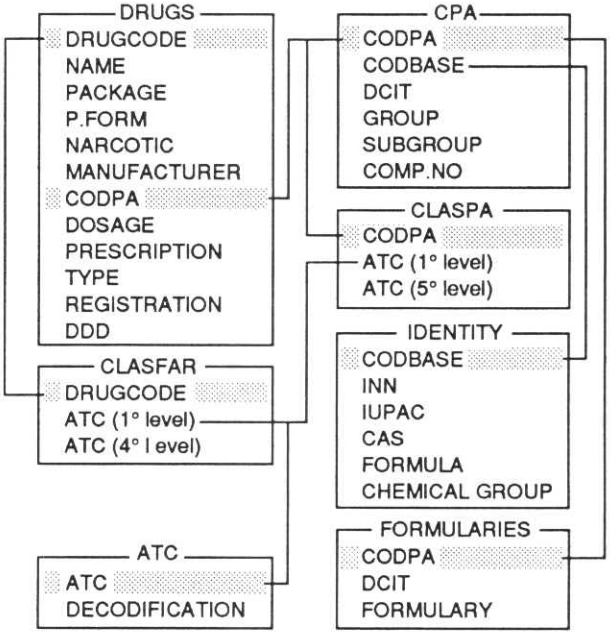

Fig. 3. - Data base logical paths.

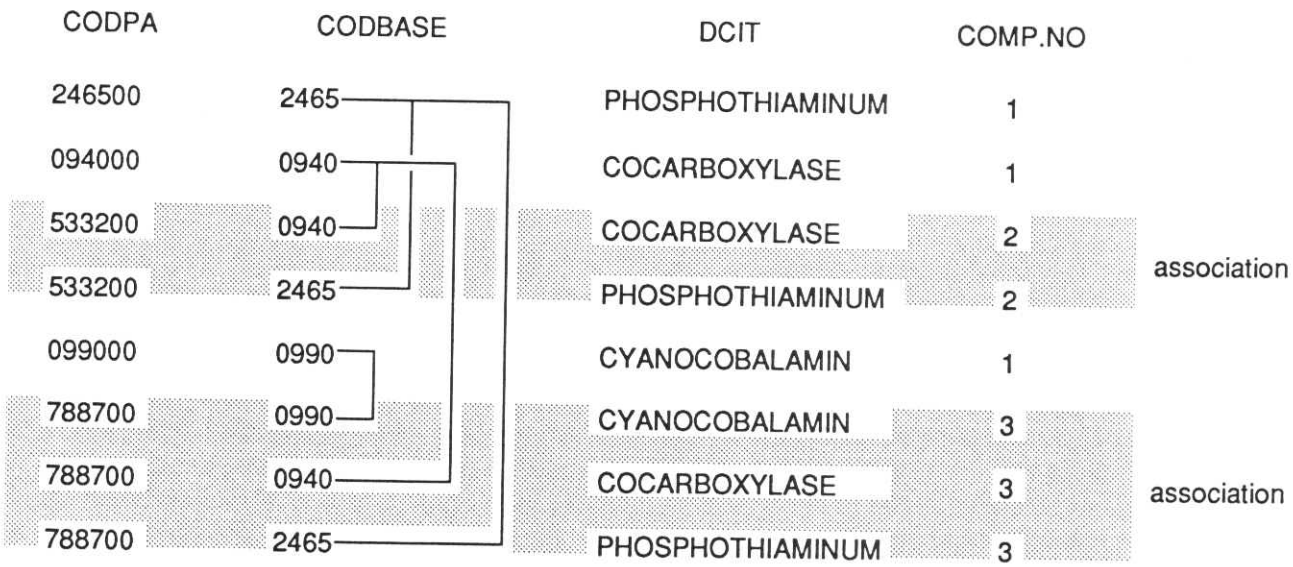

Fig. 4. - Logical organization of the CPA relation.

sis-related aspect of prescription. In this connection, when used for analysis purposes, it has been found to be more appropriate than other schema in the attempt to identify, through drug prescriptions, the type of problems normally faced by general practitioners.

It is possible that in some groups of substances (i.e. antibiotics, corticosteroids, etc.), drugs are "scattered" throughout several categories. This is due to their being differentiated according to possible routes of administration (topical, systemic, etc.).

An estimate of the population "exposed" to certain families of drugs must thus entail the merging of categories that differ from the anatomical level. For instance, to analyze "Antibiotics" consumption in a population we should combine data regarding drugs in the following ATC categories: General systemic antiinfectives, Genito-urinary system, Alimentary tract and metabolism, Dermatologicals, Respiratory system, Sensory organs.

Our solution consists in the definition of some grouping criteria (for instance: antibiotics, corticosteroids, vitamins, etc.) that, on the basis of certain families of active principles allow for "transversal" aggregations on the ATC scheme.

Here is reported a subset of these groups:

- Corticosteroids;
- Mao inhibitors;
- Antiparkinson;
- Vitamins.

Chemotherapeutic agents (antibiotics, antymicotics, cephalosporines, penicillins, sulphamides, tetracyclines).

The coexistence of different aggregation criteria guarantees the possibility to meet more complex needs in terms of information access, and the chance to develop drug prescription analysis procedures in general practice, according to several data organization layers (Fig. 5).

FORMULARIES: this relation reports data regarding the substances registration in particular lists such as the PTOR (Regional Hospital Therapeutic Registers), WHO essential drugs list, defined registers for medical practice, etc.

The name used for the substances is that assigned by each individual formulary, and may thus slightly differ from the DCIT assigned to the substance. It was decided to keep the original names also due to the fact that they actually consitute the basis for an archive of synonyms.

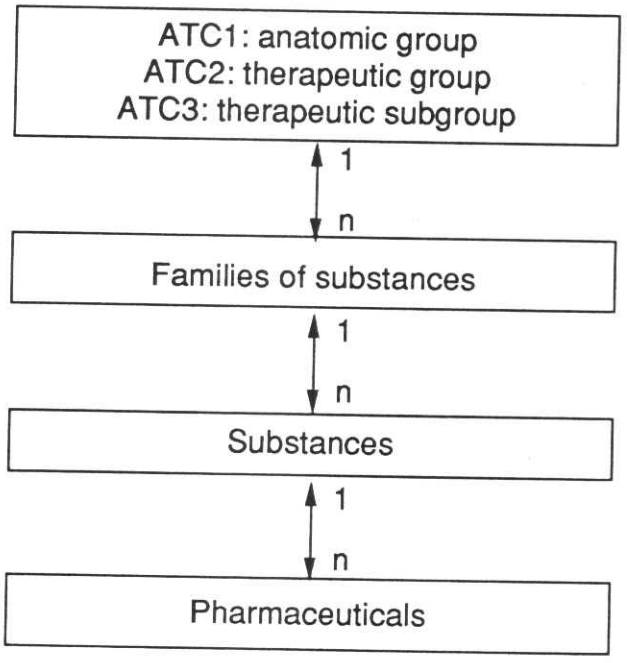

Fig. 5. - Data organization layers. Formularies: this relation reports data regarding the substances registration in particular lists such as the PTOR (Regional Hospital Therapeutic Registers), WHO essential drugs list, defined registers for medical practice, etc. The name used for the substances is that assigned by each individual formulaty, and may thus slightly differ from the DCIT assigned to the substance. It was decided to keep the original names also due to the fact that they actually constitute the basis for an archive of synonyms.

The physical structure of the data base

The Data Base originally implemented in the Hewlett-Packard IMAGE 3000 environment, is currently available in DBIII Plus too.

The physical structure of the different data sets is described in Table 1. The items glossary is reported in Table 2.

Data quality control

The quality control of a Data Base is generally based on two levels:

- the one relating to the logical structure and to the data model employed;
- the one relating to the significance of data stored in the Data Base.

As regards the first point, it is based upon the fact that in a data model there is always a collection of "integrity" rules that implicitely or explicitely define a set of significant states for a Data Base. In the case of the relational model, these rules fall into two categories:

- integrity of the entities;
- referential integrity;

that take concrete shape in primary key constraints (in other terms, no redundance must exist) and constraints of associations between entities. The rationale of the latter constraint derives from the fact that in "normalization" processes, a phenomenon occurs whereby key values are scattered to different relations, which are thus correlated by attributes that have the same significance.

Table 1. - *Physical structure of the different data sets*

Relations	Attributes	Index
DRUGS	DRUGCODE	Y
	NAME	Y
	PACKAGE	N
	P. FORM	Y
	NARCOTIC	N
	MANUFACTURER	Y
	CODPA	Y
	DOSAGE	N
	PRESCRIPTION TYPE	N
	REGISTRATION	N
	DDD	N
CLASFAR	DRUGCODE	Y
	ATC4	Y
	ATC1	Y
ATC	ATC	Y
	DECODIFICATION	N
CPA	CODPA	Y
	CODBASE	Y
	DCIT	Y
	GROUP	Y
	SUBGROUP	Y
	COMP. NO.	N
CLASPA	CODPA	Y
	ATC5	Y
	ATC1	Y
IDENTITY	CODBASE	Y
	INN	N
	IUPAC	N
	CAS	N
	FORMULA	N
	CHEMICAL GROUP	N
FORMULARIES	CODPA	Y
	DCIT	Y
	FORMULARY	Y

Table 2. - *Glossary of data base attributes*

1	ATC	Category in the decodification of the ATC classification scheme
2	ATC1	Level one of the ATC scheme; indicates general categories
3	ATC4	Level four of the ATC scheme (maximum level of drugs)
4	ATC5	Level five of the ATC scheme (identifies the individual substance)
5	CAS	Number of the Chemical Abstract Service Registry
6	MANUFACTURER	Code of the drug manufacturer
7	CODBASE	Code of substance excluding salifications
8	CODPA	Active principle code
9	DRUGCODE	Single Ministry Code of the drug
10	PACKAGE	Description of the type of package of the pharmaceutical
11	DCIT	Italian Common Name of the active principle
12	DDD	Daily doses contained in the package
13	DECODIFICATION	Description of the ATC in the scheme decodification
14	DOSAGE	Codification of particular data (pediatric, retard, strong, etc.)
15	REGISTRATION	Registration date of drug
16	FORMULA	Molecular formula of the substance
17	P. FORM	Codification of the pharmaceutical form (tablets, vials, etc.)
18	FORMULARY	Formulary in which a substance is contained
19	GROUP	Group to which the substance belongs, if any (vitamins, etc.)
20	CHEMICAL GROUP	Chemical group to which a substance belongs
21	INN	International Nonproprietary Name of the active principle
22	IUPAC	Description of the chemical formula
23	COMP. NO.	Number of substances that share the same composition
24	NAME	Trading name
25	SUBGROUP	Subgroup to which substance belongs, if any
26	NARCOTICS	Flag of registration into Narcotics Table
27	PRESCRIPTION TYPE	Particular type of prescription

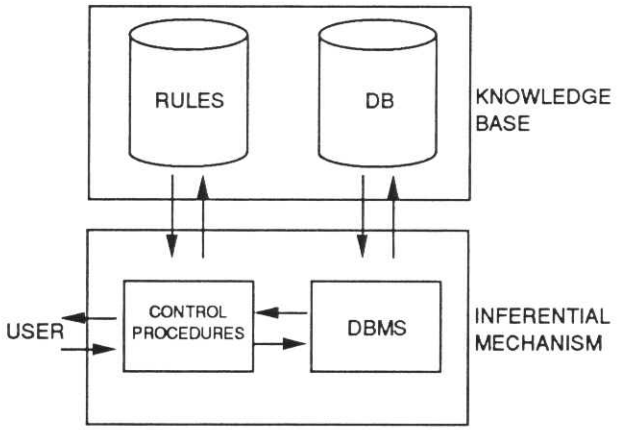

Fig. 6. - Logical schemes of the data base of the rules for data quality control.

For instance, in the DRUGS relation (Fig. 3), the "active principle code" (CODPA) attribute must be a subset of the domain of values that, for the same attribute, is defined in the CPA relation. In other terms, it is not possible to introduce a new element in DRUGS, unless its composition is already reported in CPA.

These constraints of integrity for the Drugs Data Base are guaranteed by a set of specific procedures.

The semantic level control in our case essentially refers to the ATC classification.

The "rules" regulating the classification (defined in the volume "Guidelines for ATC classification" Nordic Statistics on Medicines [7], were summarized according to two main criteria:

- administration route;
- composition (single compounds or associations).

Such criteria have generated about 100 rules, represented in a special Data Base [8], used according to the logic in Fig. 6 and to maintain the scope of the Data Base during classification procedures.

Review submitted on invitation by the Editorial Board of the *Annali*. Accepted for publication: 4 June, 1990.

REFERENCES

1. RASCHETTI, R., MELCHIORI, R., BARONCELLI, V. & CAVAZZUTTI, G. 1980. SIDAF: a drug information data base system. In: *Proceeding of the IV Annual Symposium on Computer Application in Medical Care*. Washington, November 1980.

2. RASCHETTI, R., BARONCELLI, V., FILIPPONI, M. & CARLUCCI, F. 1983. Progetto SIDAF: analisi e progettazione di un sistema informativo per il settore farmaceutico del Servizio Sanitario Nazionale. *Ann. Ist. Super. Sanità* **19** (2/3): 339-349.

3. WORLD HEALTH ORGANIZATION. 1985. Drugs in general practice. WHO, Ginevra. *EURO Reports and Studies*, **91**.

4. TOGNONI, G. & LAPORTE, J.R. 1984. Epidemiologia del farmaco. *Ter. Mod.* **111**(1).

5. STROM, B., CARSON, J., LEE MORSE, M. & LEROY, A. 1985. The computerized on-line Medicaid pharmaceutical analysis and surveillance system: a new resource for postmarketing drug surveillance. *Clin. Pharmacol. Ther.* **38** (4): 359-364.

6. STOLLEY, P. & STROM, B. 1986. Evaluating and monitoring the safety and efficacy of drug therapy and surgery. *J. Chron. Dis.* **39** (12): 1145-1155.

7. NORDIC COUNCIL ON MEDICINE. 1981-1983. Guidelines for ATC classification. (NLN publication, n. 16 - part 3).

8. RASCHETTI, R., MENNITI IPPOLITO, F., CAFFARI, B. & MAGGINI, M. 1988. Struttura concettuale di un sistema esperto per il controllo di qualità dei dati negli studi epidemiologici. In: *Atti del Convegno Attualità e Prospettive dell'Informatica in Medicina*. Firenze, 26-29 ottobre, 1988.