Model Formation ■

Building and Evaluation of a Structured Representation of Pharmacokinetics Information Presented in SPCs: From Existing Conceptual Views of Pharmacokinetics Associated with Natural Language Processing to Object-oriented Design

Catherine Duclos-Cartolano, PharmD, Alain Venot, MD, PhD

Abstract **Objective**: Develop a detailed representation of pharmacokinetics (PK), derived from the information in *Summaries of Product Characteristics* (SPCs), for use in computerized systems to help practitioners in pharmaco-therapeutic reasoning.

Methods: Available knowledge about PK was studied to identify main PK concepts and organize them in a preliminary generic model. The information from 1,950 PK SPC- texts in the French language was studied using a morpho-syntactic analyzer. It produced a list of candidate terms (CTs) from which those describing main PK concepts were selected. The contexts in which they occurred were explored to discover co-occurring CTs. The regrouping according to CT semantic types led to a detailed object-oriented model of PK. The model was evaluated. A random sample of 100 PK texts structured according to the model was judged for completeness and semantic accuracy by 8 experts who were blinded to other experts' responses.

Results: The PK text file contained about 300,000 words, and the morpho-syntactic analysis extracted 17,520 different CTs. The context of 592 CTs was studied and used to deduce the PK model. It consists of four entities: the information about the real PK process, the experimental protocol, the mathematical modeling, and the influence of factors causing variation. Experts judged that the PK model represented the information in 100 sample PK texts completely in 89% of cases and nearly completely in the other 11%. There was no distortion of meaning in 98% of cases and little distortion in the remainder.

Conclusion: The PK model seems to be applicable to all SPCs and can be used to retranscribe legal information from PK sections of SPCs into structured databases.

■ **J Am Med Inform Assoc.** 2003;10:271–280. DOI 10.1197/jamia.M1193.

Pharmacology and pharmacokinetics provide information that makes it possible to understand drug-body interactions. Such information is useful in the reasoning process (1) for choosing a drug to prescribe;[1–3] (2) for comparing drugs;[4] (3) for giving advice on drug administration; (4) for identifying contraindications, interactions, or therapeutic strategies in physiological conditions or particular diseases;[5] and (5) for listing drugs according to some of their properties .

Pharmacokinetics concerns the descriptive and quantitative study of what happens to a drug in the body to which it is administered.[7] Mathematical modeling is used for the study of the dynamic processes involved.

In compartmental modeling, the body is represented as a whole, made up of virtual compartments between which exchanges occur. The compartmental approach uses mathematical models comprising differential equation sets[8] to predict the serum level of a drug according to the dose, the intake interval, the number of administrations, and the mode of administration. These models have been integrated into software to assist clinicians in determining the optimal dosage schedule (e.g.,

Affiliation of the authors: Laboratoire d'Informatique Médicale et de Bioinformatique (LIM&BIO), UFR Santé, Médecine, Biologie Humaine, Université Paris 13, Bobigny, France

The authors thank P. Zweigenbaum for his help in the text processing with Lexter® software, M.C. Bonjean for supplying the Vidal® drug data base, and Drs. H. Dréau, A. Dubosq, P. Farmeur, M. Ghez, P. Letoumelin, L. Ridoux, and J.L. Risi, for evaluation of the model.

Correspondence and reprints: Alain Venot, LIM&BIO, UFR de Santé, Médecine et Biologie Humaine - Léonard de Vinci 74, rue Marcel Cachin 93017 Bobigny cedex, France; e-mail: <alain.venot@limbio-paris13.org>.

Received for publication: 08/01/02; accepted for publication: 12/16/02.

USC*PACK[9]). Initially, such software was only available for a few products with narrow therapeutic indices, but it now covers a broad range of products. It also allows the integration of particular population features.[10] In physiology-based pharmacokinetic modeling,[11] the compartments represent anatomical or physiological entities, and include specific parameters for metabolism, tissue binding and tissue reactivity. These models are useful in predicting tissue concentrations of a chemical agent and are primarily intended for toxicology studies.

Thus, modeling choices available do not allow direct handling of knowledge data by the clinician and need to be integrated into a software system for particular tasks such as dosage schedule optimization.

Most pharmacokinetics data are found in textual form in drug compendia and drug databases. Their sources are either summarized data produced by the drug regulatory agencies[12–14] or data from studies published during drug development.[15,16] In these databases, the information about pharmacokinetics is always expressed in natural language.[12,16,17] Sometimes information is organized in sections with labels representing generic concepts such as "absorption," "distribution," "metabolism," and "elimination,"[13,18,19] or narrower concepts such as bioavailability, bioavailability by oral route, bioavailability by oral route for tablets, bioavailability by intramuscular route, and effect of food on absorption.[15] Finally, the contents of the pharmacokinetic description are similar, whether the source is European (e.g., Vidal) or American (e.g., AHFS DI).

Pharmacokinetics data are written in natural language, which can be stored in electronic databases or drug compendia for consultation by practitioners[20] but is not well suited to the development of automated prescription assistance.[21] The conceptual and formal representation of the domain ontology is essential to handle the knowledge effectively in computerized systems.[22] The OpenGALEN Common Reference Model contains pharmacokinetics descriptions such as drug absorbing, excreting, metabolizing, and protein binding process,[23] but no current drug ontology uses it. The "Alcohol and Other Drugs Thesaurus"[24] describes an ontology of pharmacokinetics aimed at expressing pharmacokinetics processes quantitatively. We are not aware of a more finely detailed model of pharmacokinetics.

The objective of this work was to build a finely structured representation of pharmacokinetics that could be used in a computerized system to facilitate the use of pharmaco-therapeutic information currently contained only in electronic drug compendia. Prescribers could then use a computerized system to get answers to such questions as: (1) which anti-hypertensive drugs are mainly eliminated in urine, (2) which anti-infective drugs diffuse in bone, and (3) for which anti-asthmatic drugs are there pharmacokinetic data for children. It would also be easy to build comparative lists to give an overview of differential properties of drugs (for example: compare the metabolism of cardiac glycoside drugs). Our aim was to elaborate an object-oriented model that can represent in a structured and exhaustive way the information contained in pharmacokinetics sections of SPCs. Here we describe our methodology for knowledge acquisition based on domain analysis and natural language processing, and the method used to evaluate the content coverage and the accuracy of the developed model. We present both the object-oriented model for representation of pharmacokinetics information in UML formalism and an evaluation of the model.

Methods

We used a two-step approach to identify the information elements contained in the pharmacokinetics section of the SPCs. First, we analyzed the compartmental and physiology-based pharmacokinetics modeling by identifying the main concepts on which they rely. Second, we used natural language processing tools to analyze the contents of pharmacokinetics texts extracted from SPCs. We then built an object-oriented model of pharmacokinetics information, which was evaluated for its ability to represent the initial text information and maintain its sense.

Knowledge Acquisition Method

Establishing the Foundations of the Model

First, we studied established pharmacokinetics knowledge using educational pharmacokinetics sources.[7,8,11] We manually identified the fundamental concepts of pharmacokinetics and defined the experimental schedule required to produce real data for use in compartmental or physiology-based modeling.

This first stage of analysis led to a preliminary model of the domain. This model describes the categories of knowledge found in pharmacokinetics. These categories are considered the main concepts that should be found in pharmacokinetics texts and are used to group terms in the following analysis.

Terminological Analysis of Pharmacokinetics Section of SPCs

Data sources. The pharmacokinetics texts were taken from the Vidal drug database that contains the information specified by the French drug agency in the SPCs. 5,293 different drugs had SPCs with a pharmacokinetics section, and 1,950 unique texts were extracted (two or more different drugs can have the same pharmacokinetics text). We combined them in a single file.

Natural language processing tool. We used the robust syntactic analysis tool Lexter to analyze the text file.[25] It performs a morpho-syntactic analysis of the sentences of a French corpus, and yields a dependency network of candidate terms (CTs). CTs are presented in uninflected form (lemma form). They are either simple (single noun, adjective, verb, adverb) or complex (words arrangements in the phrase such as nominal syntagma or adjec-

tival syntagma (adverb + adjective)). Complex CTs are divided into a head and an expansion (e.g., the nominal syntagma "substantial protein binding" is separated into a head "protein binding" and an expansion "substantial"). Through the terminological network provided by Lexter, each CT is linked to all of the CTs for which it is the head or the expansion. Each CT is also linked to the textual units where it appears (e.g., the CT "plasmatic protein," a nominal syntagma, appears in the textual unit "the level of plasmatic protein binding is low [14 to 21%]").

Terminological exploration. The documentation analyzed was substantial and therefore the expected number of CTs was high.[26] The most common words occur frequently and constitute the majority of the word tokens in the text.[27] We ordered CTs according to their syntactic category and their frequency of occurrence and then examined the CTs that occurred more than 3 times.

- ***Selection of CTs***. Pharmacokinetics-specific CTs were manually assigned to a main concept defined in the preceding analysis. CTs with important meaning not directly related to pharmacokinetics were assigned to a "nonspecific" concept. This first analysis led to the identification of the main semantic groupings.

- ***Relation between CTs***. The lexical environment of words with similar meaning (such as binding, fixation) was explored. The terminological network of these CTs and the textual units in which they appeared helped us to find the co-occurring CTs. These CTs were listed according to their frequency of co-occurrence. We repeated this analysis for all the pharmacokinetics-specific CTs.

Object-oriented Model Building

For CTs with similar meanings we compared the lists of co-occurring CTs and deduced common underlying concepts (e.g., absorption is a reaction, metabolism is a reaction, absorption occurs in the stomach, metabolism is hepatic, so that we can deduce that a reaction occurs in a location). These concepts were then organized either as classes, attributes, or class relationships. They were then added to the basic model (class addition, class partition, class generalization, or class specialization). The representations of a sample of 10 randomly selected texts were assessed, and the findings were used to refine the model.

Object-oriented Representation

To represent the object-oriented model of pharmacokinetics, UML formalism[28] was chosen. The relationships are either:

- ***Inheritance relationship***: Class A inherits from Class B or Class B subsumes Class A (Class A _ ▶ Class B)

- ***Aggregation relationship***: Class A is part of Class B or Class B is composed of Class A (Class A _ ◇ Class B)

- ***Association relationship***: Class A is associated with Class B (Class A _Class B)

The cardinality constraints are noted as follows(:)0...1 = 0 to 1; 1...*= 1 to many, 0...* = 0 to many, 1 = 1, * = many.

Evaluation Method

Material

To perform the evaluation, we used 100 pharmacokinetic texts randomly extracted from the 1,950 French SPC texts available. The selected texts had between 1 and 26 sentences and 14 to 717 words. The sentences were manually converted by one of us (CD) in an Access database that reproduced the structure of the model. This took almost 3 hours per text although the time varied according to the length of text. A total of 5,225 class instances and 9,347 attribute instances were created. All of the attributes specified in the model were filled at least once. To facilitate assessment, we used colored forms that presented (1) the identification of the drug (name, INN, dosage, route), (2) the whole pharmacokinetics text, and (3) each sentence of the text in both natural language and structured format. We used a set of colors to distinguish class attributes, attribute values, and relationships (Fig. 1).

Evaluation Criteria

The following criteria were used to evaluate the model:

- ***Completeness***: measures the ability of the model to represent all information units contained in the initial text; it also shows if there are concepts missing from the pharmacokinetics model. This criterion has four values: completely, almost completely, little or not represented.

- ***Semantic accuracy***: measures the ability of the model to store the meaning of the initial whole text. This criterion shows the descriptive competence of the pharmacokinetics model and the potential ambiguity of some concepts. It has four values (entirely, substantially, little, or not distorted).

Method

For each structured text, eight experts working independently were asked to assess these two criteria. They compared the structured representation to the initial text for each sentence and assigned a value to each criterion. When they found a defect of representation or of meaning they attributed a cause: either a defect of the model or a defect of the textual source. For example: the sentence "in case of renal impairment it is not necessary to modify the dosing schedule" would not be represented by the model because it is not pharmacokinetics but dosing schedule.

In this first analysis, experts assigned a value to the whole text according to the model point of view.

The eight evaluators first worked on a preliminary sample of five texts to familiarize themselves with the model and the evaluation principle. For the evaluation, 25 structured texts (selected from the sample of 100 texts)

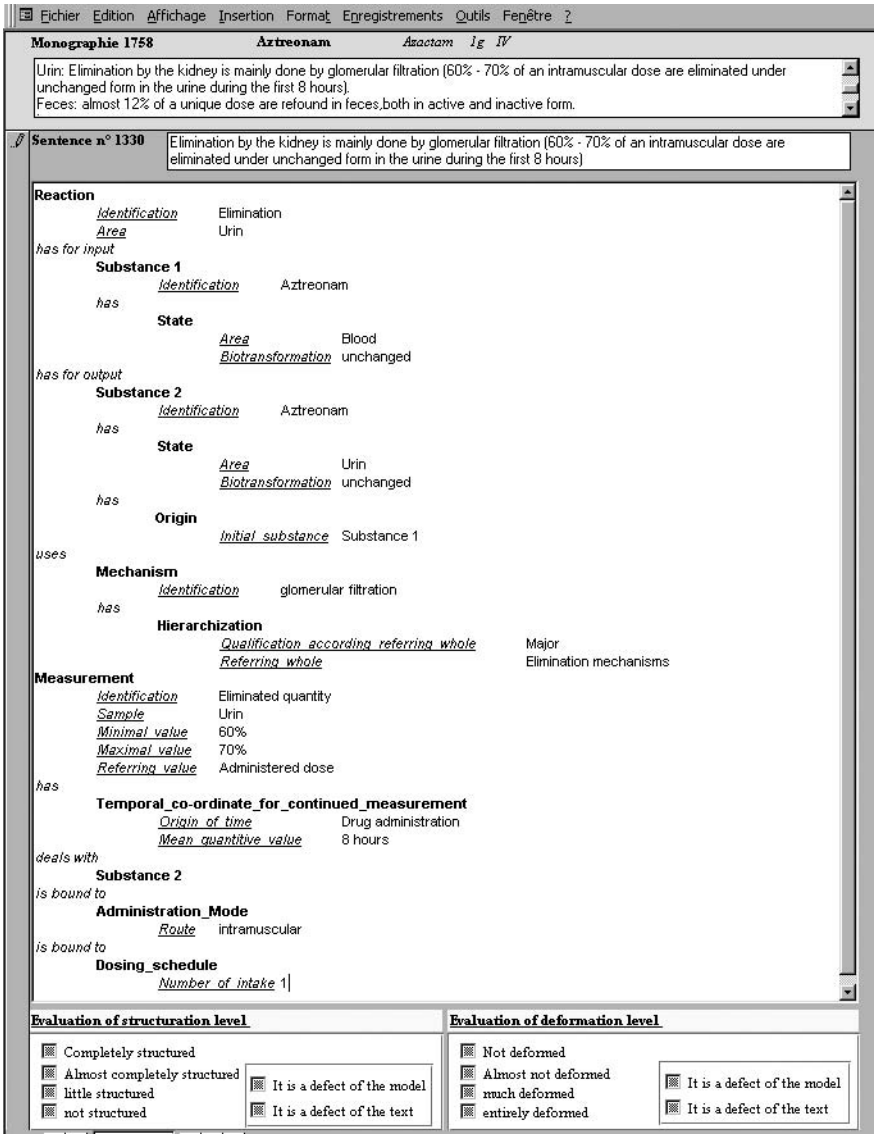

Figure 1. Example of both textual and structured representation of a sentence extracted from a monograph (Aztreonam monograph) presented to the experts for the evaluation of the model.

were assigned randomly to each evaluator. Each of the 100 texts was evaluated by two experts, each blind to the findings of the other. Any disagreement between evaluators was resolved using the Delphi method[29]: for each evaluator involved in the disagreement, results of the evaluation of the other evaluator were presented and then a new evaluation was asked.

Statistics

We determined the 95% confidence interval of the frequency of each criterion value. We performed ANOVA non-parametric tests to determine:

- If there was a significant difference between the corpus of texts assigned to each evaluator (number of sentences)
- If there was a significant difference between the evaluators' responses.

Results

Knowledge Acquisition Results

Generic Model of Pharmacokinetics

We constructed a preliminary global model of pharmacokinetics from the results of the domain study using UML formalism (Fig. 2). The model distinguishes three entities: the administration protocol, the real pharmacokinetic process description and the mathematical model building.

From an *administration protocol* in which the route and dosage regimen are specified (dose, administration frequency, number of intakes): the *real pharmacokinetics process* can be described as such:

- The parent compound or its metabolites undergo a series of reactions including absorption, distribution, metabolism and elimination processes.

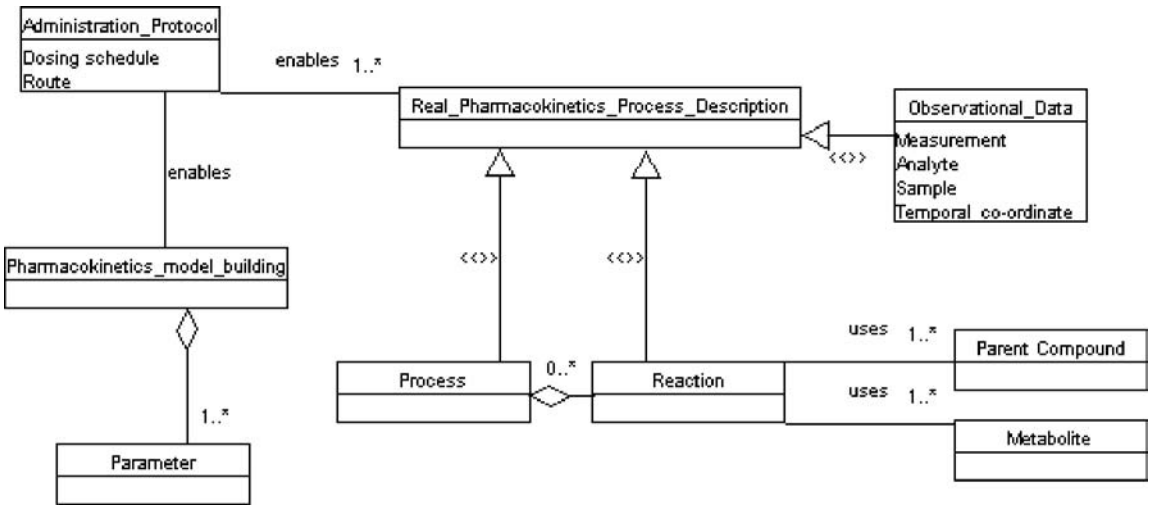

Figure 2. Preliminary global model of the pharmacokinetics domain according to UML formalism.

- It is studied by assays for the drug in samples (of, for example, blood or urine) collected at determined times.

A *mathematical pharmacokinetics model* can be elaborated using the experimental data and its parameters (for example half-life, clearance, distribution volume).

Lexico-semantic Analysis

The pharmacokinetics text file contained about 300,000 words and was composed of texts of 5 to 1,043 words.

The Lexter lexico-semantic extraction based on the 1,950 pharmacokinetics texts provided a lexicon of 206,500 entries (17,520 different CTs) and 20,761 textual units. Of the 17,520 different CTs, 3,132 occurred more than 3 times and only 222 occurred more than 100 times. Nouns and nominal syntagma were the most frequent syntactic units.

We selected CTs representing direct instances of major pharmacokinetics concepts described in the generic model, such as CTs related to an administration protocol, a process, an observational datum and a parameter. There were 592 such specific pharmacokinetics CTs. We also selected CTs representing useful concepts such as quantification, variation, type of disease, or a population. These nonspecific terms represented 535 CTs. Table 1 shows an example of how the more frequent CTs were grouped according to their similar meanings in pharmacokinetics-specific or nonspecific concepts.

Study of the lexical environment of CTs with similar meanings yielded the set of common CTs that frequently co-occur. For example, fixation was most frequently described with CTs related to a binding site, a numerical value, a substance name, a precision, an ordinal value, and a study type. These CTs also co-occurred with "binding." We could, therefore, deduce that the concepts underlying "binding" and "fixation" are common (Table 2).

Model Building

We compared the co-occurring concepts for CTs with near meanings to identify class attributes, class relationships, and new classes. For example, instances of a reaction co-occurred most frequently with other CTs that described where, when, why, and how the reaction occurred, on what the reaction acted, and how the reaction was assessed. To describe a reaction class, its characteristics need to be identified. The reaction class should be linked to classes describing (1) properties, (2) mechanisms, and (3) substances; a measurement class is needed to illustrate it.

The new classes, relationships, and attributes were then used to refine the generic model. The initial reaction class was then linked to a property class and a mechanism class. Parent compound and metabolite were subsumed by a substance class that was linked to an origin class and a state class. Process classes and reaction classes were subsumed by a more abstract class called "explanatory data." Observational data was replaced by a measurement class, which had links with temporal-coordinate class, value class and evolution class. The administration protocol was divided into four classes (administered product, administration mode, dosing schedule, and treatment schedule). The administration protocol class was aggregated with two new classes—"measurement protocol" and "population features"—to form a new abstract class called "experimental protocol."

Description of the Final Model of Pharmacokinetics

The pharmacokinetics model describes the reality of the processes obtained under experimental conditions, the compartmental pharmacokinetics model and its parameters, and the variations obtained under particular con-

Table 1 ■ Main Semantic Groupings for the Most Frequent Candidate Terms

Concept Grouping	Frequency	Examples of Candidate Terms
Pharmacokinetic specific groupings		
Grouping of candidate terms related to administration		
Route	2,610	route, oral, intramuscular, perfusion
Dose	1,717	dose
Form	179	tablet
Grouping of candidate terms related to process		
Release	133	release
Absorption	1,355	absorption, to absorb, resorption
Distribution	1,936	distribution, to bind, binding, fixing, accumulation, plasmatic protein binding, diffusion, to cross
Metabolism	2,352	metabolite, to metabolize, hepatic first-pass,
Elimination	2,854	elimination, to eliminate, to excrete, excretion
Grouping of candidate terms related to observational data		
Measurement	3,860	plasmatic level, plasmatic concentration, peak
Grouping of candidate terms related to parameter		
Parameter	2,858	half life, bioavailability, clearance
Non specific groupings		
Quantification	1,563	low, high, elevated, major
Variation	1,442	to increase, to modify, to decrease, to vary
Activity	727	active, inactive
Population	628	adult, elderly, infant, healthy volunteer
Speed	517	fast, slow
Pathology	202	renal insufficiency, hepatic insufficiency

ditions. The whole detailed model is described in Figure 3.

- ***Information about the real pharmacokinetics process knowledge***

Information about the real pharmacokinetics process can be separated into that which explains the behavior of the drug in the body through time, and illustrative information, which includes the results of the measurements made on samples collected at given times.

Explanatory data describe the drug-body interaction. The changes of a drug in the body can be described as a succession of four processes: absorption, distribution, metabolism, and elimination. Each can be broken up into one or more reactions with one or more mechanisms. The reactions involve a substance, which changes states because of alterations in its state of binding, activity, localization, and biotransformation. The whole process—reactions, substances, and mechanisms can have certain properties (e.g., intestinal absorption is slow) and be ordered (e.g., urinary elimination occurs mainly by glomerular filtration).

The real pharmacokinetic process can be investigated by measurement with results expressed quantitatively or qualitatively and associated with a temporal co-ordinate. These measurements can change in time with one or more phases.

- ***Experimental protocol***

The *experimental protocol* describes the conditions for obtaining data on the real process. It consists of an administration protocol, a measurement protocol and a population that has certain features.

The *administration protocol* is made up of information about the *administered drug* (e.g., tablet of drug X), the *mode of administration* (e.g., administration by the oral route on an empty stomach), the *dosing schedule* (e.g: repeated administration every 8 hours of 300 mg of X), and the *treatment schedule* (e.g.,an 8-day course).

The *measurement protocol* describes the conditions of measurements (e.g., by gas chromatography).

The *population features* concern the population sample used for experimentation. These features consist of general features (e.g., man or dog) and particular features that give details about the physiologic type (e.g., a child less than 12 years old), the genetic type (e.g., slow acetylator), or the pathologic type (e.g., renal insufficiency with a creatinine clearance lower than 30 ml/min).

Table 2 ■ Results of the Analysis of the 302 Contexts of Occurrence Linked to the Candidate Term "Binding" and of the 127 Contexts of Occurrence Linked to the Candidate Term "Fixing": Regrouping of the Candidate Terms Found Most Frequently in These Contexts of Occurrence (More than 10 times) with Frequency of Occurrence

Candidate Terms Found in the Studied Contexts of Occurrence	Number (Frequency) of Links with "Binding"	Number (Frequency of Links with "Fixing"	Deduced Co-occurring Concepts for Binding Reactions
Plasmatic protein	200 (0.66)	51 (0.40)	Binding site
Protein	53 (0.17)		
Proteic	16 (0.05)		
Albumin	11 (0.04)		
Numerical value expressed as a percentage	201 (0.66)	57 (0.45)	Numerical value
Substance name	115 (0.38)	61 (0.48)	Substance name
about	50 (0.16)	14 (0.11)	Precision
of about	27 (0.09)	14 (0.11)	
Low	45 (0.15)	13 (0.10)	Ordinal value
Important	20 (0.07)		
in vitro	10 (0.03)		Study type

- ***Mathematical model***

The information in the mathematical model describes the selected compartmental kinetic model applicable to the real processes observed. This information concerns either the model structure (e.g: monocompartmental model) or the pharmacokinetics model parameters (e.g., the half life is 3 hours).

- ***Information about the influence of factors causing variation***

This describes the changes in measurements, parameters, or processes related to the changes in experimental conditions (e.g., the passage through the hematoencephalic barrier is increased when there is meningeal inflammation; the drug half-life is higher for children; the unbound fraction is higher in patients with hypoalbuminemia).

Model Evaluation Results

We did not observe any significant difference between the corpus of texts (number of sentences) evaluated by each evaluator ($p > 0.657$) or between evaluator responses ($p > 0.591$ for completeness and $p > 0.456$ for semantic accuracy).

Evaluation indicated that the model gives a good representation of pharmacokinetics information: it was able to represent pharmacokinetics information completely in 89% of the cases (CI_{95} [83%–95%]), and in 11% of the cases in an "almost complete" way (CI_{95} [5%–17%]). For example, this sentence was judged as almost completely represented: "According to its main biliary excretion and to its important presystemic metabolism, an accumulation of fluvastatine is shown in patients having hepatic insufficiency." It was translated into the model into classes describing biliary excretion, metabolism and accumulation but the causal link between the various components could not be represented. There were no examples of significantly or entirely defective representation.

The meaning of the information was classified "not distorted" in 98% of the cases (CI_{95} [95%–100%]) and "almost not distorted" in 2% of the cases (CI_{95} [0%–5%]). The following sentence was judged "almost not distorted": "Metabolic transformation by hepatic microsome enzymes (inducible)" was translated into the model as: "Metabolism is a reaction, has hepatic area, has the generic site microsome, has the specific site enzyme, has the property inducible." The deformation of meaning is on the inducibility that is linked to metabolism instead of enzyme. There were no cases of significant or entire distortion.

The evaluators agreed on 93 texts and disagreed on only seven texts. The use of Delphi method to solve these cases was effective. A consensus was quickly obtained for all seven texts after a new evaluation. Each evaluator saw 25 texts and judged 21–25 times that the information was completely represented, 0–4 times that the information was almost completely represented, 23–25 times that the information was not distorted, and 0–2 times that the information was almost not distorted.

Discussion and Conclusion

We aimed to develop a pharmacokinetics model capable of representing all information contained in the pharmacokinetics section of SPCs. We used both general knowledge about pharmacokinetics and specific descriptions of drug pharmacokinetics available in SPCs. General knowledge helped us to define the generic structure of the pharmacokinetic model. This top-down approach, although performed manually, gave a quick organization for the main concepts, because the domain is delimited and is already the object of mathematical modeling. To refine this generic structure, we assessed the knowledge present in pharmacokinetics

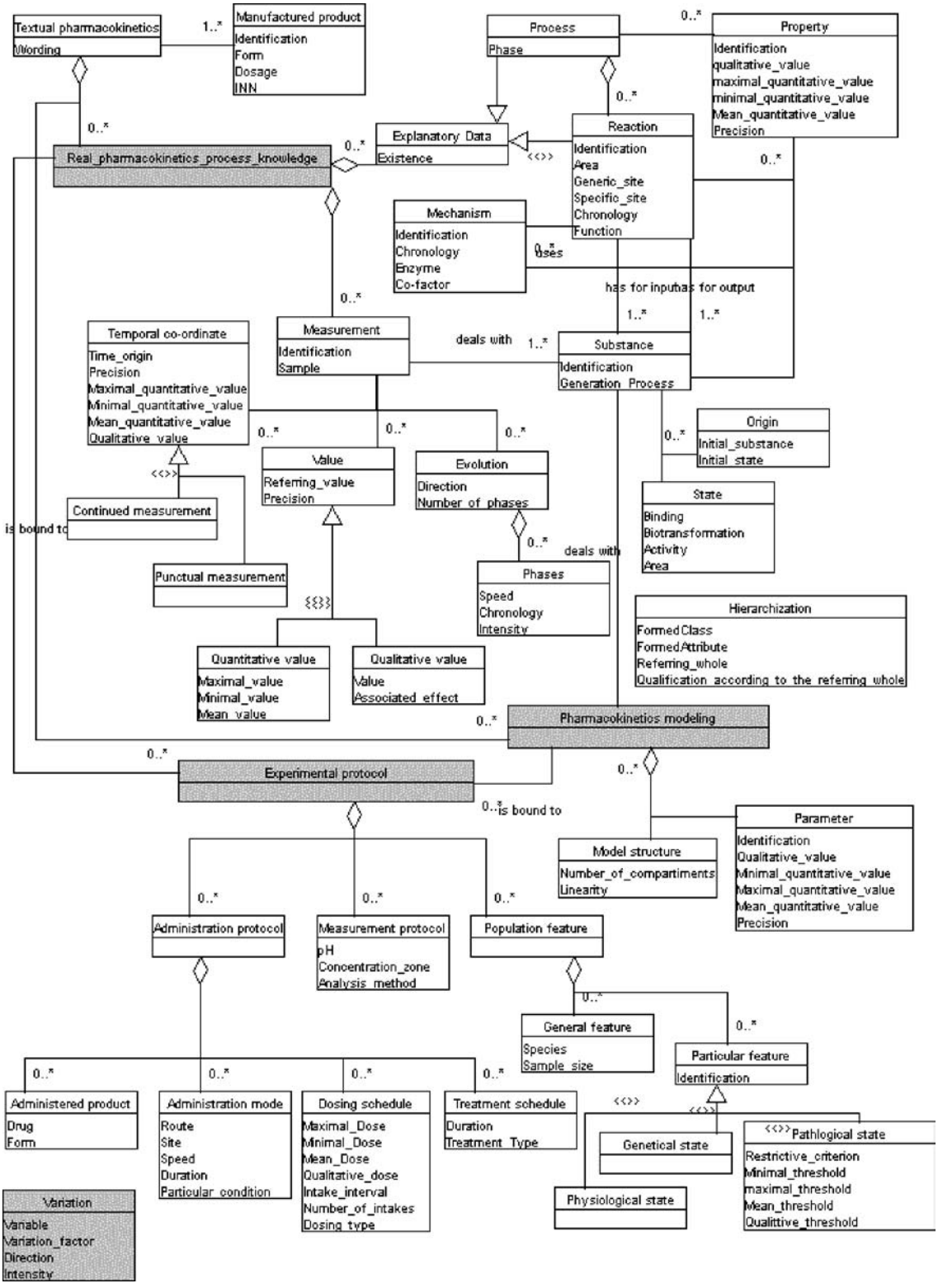

Figure 3. Object-oriented model of pharmacokinetics information found in SPCs according to UML formalism.

sections of SPCs. For this bottom-up approach, we combined natural language processing results using manual semantic analysis of the relevant pharmacokinetics candidate terms specifying generic concepts from top down analysis. The bottom-up and top-down approaches are standard methods in knowledge engineering and are used in the medical field,[30] but they are viewed as two separate ways of constructing a model.[31] Our approach can be viewed as a middle-out approach[32]: the most important concepts lead to generalization and special-

ization. Generic pharmacokinetics concepts were used to select domain-specific candidate terms (CTs) that were then used as reference points to navigate in the initial text and to discover other CTs that often co-occur with them. The study, focused on specific CTs with similar and near meanings, allowed exploration of the common lexical environment of pharmacokinetic-specific CTs. By grouping all the co-occurring CTs, it was possible to indicate the links between them and to generate concepts that are new and more precise that would be introduced into the initial model to extend it.

We initially selected very few CTs from the wide-ranging lexicon produced by the natural language processing tool. The selective exploration of the contexts of occurrence related to these selected CTs optimized the semantic analysis: the co-occurring CTs were mainly describers of the CTs selected.

The results of the lexico-semantic analysis were then easily transposed into an object oriented representation: co-occurring CTs were either attributes, new classes or class relationships. A large number of sentences was analyzed, and consequently the model was very rich but kept its initial generic structure.

The evaluation method used to validate the model had its strengths and weaknesses. As a detailed description of pharmacokinetics, the model has to be able to describe all information contained in SPCs. We therefore chose to represent the entirety of the textual information and not select elements from these texts as in some other evaluation studies.[33–37] According to the criteria of Friedman et al.[38] the model was frozen before the evaluation, the reference standard was established (sample of 100 randomized texts), but, unfortunately, the developer of the model participated in the evaluation. We chose one of the model developers to convert the texts into the model format because the task was very time-consuming, even for someone already familiar with the structured representation of pharmacokinetics (the task took almost 3 hours per document). We believe that it did not alter the objectivity of the evaluation because each structured representation was twice evaluated by independent evaluators. The model is not designed to evolve with time; therefore we did not need a methodology that could be reused for further evaluations, as used by Rocha et al.[37] or Zweigenbaum et al.[40] However, we compared the evaluators to judges[41] who assess how well the structured representation corresponds to the initial text.[42,43]

The evaluation of the model suggested that it adequately represents any information contained in the pharmacokinetics section in SPC. The completeness criteria indicate that some concepts are missing (such as causal relation), but the defects of representation related to the pharmacokinetics model were minor. The major defects in representing texts were related to the content of the initial text, which sometimes was vague and inconsistent (missing data or comparison between values which have not been quantified) or did not deal with pharmacokinetics (pharmacodynamics data, drug interactions). The "semantic accuracy" criterion showed that the information was mostly not distorted and the few distortions were judged to be minor. Lastly, the pharmacokinetics model satisfies the criteria of conciseness[44]: the model does not store useless definitions (none of the classes or attributes were never filled). We conclude that the model developed satisfies the desiderata of completeness and coherence of coverage described by the Canon Group.[30]

Our model is not unique but is of value because it has been evaluated and is apparently applicable to all SPCs. To avoid redundancy of attributes, we chose to create abstract classes that offer a high level of generalization, increasing the complexity of the model.

The model stores descriptive knowledge about pharmacokinetics, whereas drug databases focus on numerical information. The fine structure of the model covers all the information contained in SPCs but could also store information found in bibliographic reviews of drug pharmacokinetics. The production of structured data is unfortunately time consuming, but this could be improved by developing a knowledge editor that could recognize and extract automatically the frequent patterns of information.

However, simplified versions of the model could be created according to projected applications. This requires specifying classes and their attributes according to their frequency of use and their functional interest. For example each instance of a reaction could be a new class (absorption class or protein-binding class) and their specific properties added as attributes. The property class would then disappear. Such simplified versions could be used to structure certain elements of pharmacokinetics, but cannot represent all of the information contained in all pharmacokinetics texts.

As apparent during evaluation of the model, the quality of the textual source of information (SPCs) is often insufficient. Using the model to enter legal information about pharmacokinetics into drug databases may lead to an information gain and help clarify the content. This information could then be used by a computerized system to select drugs for a given indication according to pharmacokinetics selection criteria or to compare drug pharmacokinetic behavior in a given therapeutic class.

References ■

1. Leonetti G, Cuspidi C. Choosing the right ACE inhibitor. A guide to selection. Drugs 1995;49(4):516–535.
2. Sanchez-Navarro A, Sanchez Recio M. Basis of anti-infective therapy: Pharmacokinetic-pharmacodynamic criteria and methodology for dual dosage individualisation. Clin Pharmacokinet 1999;37(4):289–304.
3. Brodie M. Monostars: an aid to choosing an antiepileptic drug as monotherapy. Epilepsia 1999;40(Suppl 6):S17–22.
4. Wroe CJ, Solomon WD, Rector AL, Rogers JE. DOPAMINE—A tool for visualizing clinical properties of generic drugs. In The 5th Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP), 2000.

5. Van Hyfte D, Van Der Maas A, Tjandra-Maga T, De Vries Robbé P. A formal framework of knowledge to support rational psychoactive drug selection. Artif Intell Med 2001;22: 261–275.
6. Solomon WD, Wroe CJ, Rector AL, Rector AL, Fistein JL A reference terminology for drugs. Proc AMIA Symp 1999;152–6.
7. Benet L, Sheiner L. Pharmacokinetics: Absorption, distribution and excretion. In Hardman J, Limbird L (eds): Goodman and Gilman's The Pharmacological Basis of Therapeutics, 10th ed. New York, McGraw-Hill, 2001.
8. Wagner J. Pharmacokinetics for the Pharmaceutical Scientist. Basel, Technomic Publishing Pompany; 1993.
9. Jelliffe R. The USC*PACK PC programs for population pharmacokinetic modeling, modeling of large kinetic/dynamic systems, and adaptive control of drug dosage regimens. Proc Annu Symp Comput Appl Med Care 1991:922–924.
10. Keller F, Frankewitsch T, Zellner D, Simon S, Czock D, Giehl M. Standardized structure and modular design of a pharmacokinetic database. Comp Methods Programs Biomed 1998; 55:107–15.
11. Bischoff K. Physiologically based pharmacokinetic modeling. In Pharmacokinetics in Risk Assessment. Washington, DC, National Academy Press, 1987.
12. British National Formulary: available at <http://bnf.vhn.net/bnf. 2002>, 2002.
13. USP DI: available at <http://www.usp.org>, 2002.
14. Vidal pro: available at http://www.vidalpro.net. 2002.
15. DRUGDEX: available at <http://www.micromedex.com/products/drugdex/>, 2002.
16. Parfit K (ed): The Complete Drug Reference, 32nd ed. London, The Pharmaceutical Press, 1999.
17. Thériaque: available at <http://www.theriaque.org>, 2002.
18. BIAM: available at <http://www.biam2.org>, 2002.
19. American Society of Health-Systems Pharmacists. AHFSfirst: available at <http://www.ashp.org/public/pubs/ahfs/index.html>, 2002.
20. Franke L, Avery AJ, Groom L, Hosfield P. Is there a role for computerized decision support for drug dosing in general practice? A questionnaire survey. J Clin Pharm Therap 2000; 25:373–7.
21. Milstein C, de Zegher I, Venot A, Séné B, Pietri P, Dahlberg B. Modeling drug information for a prescription-oriented knowledge base on drugs. Methods Inform Med 1995;34:318–327.
22. Zweigenbaum P, Bachimont B, Bouaud J, Charlet J, Boisvieux J. Issues in the structuring and acquisition of an ontology for medical language understanding. Methods Inform Med 1995;34(1/2):15–24.
23. OPEN GALEN: Open sources available at <http://www.opengalen.org/resources.html>, 2002.
24. Alcohol and other drug thesaurus: available at <http://etoh.niaaa.nih.gov/AODVol1/aodhqe.htm#EE>, 2002.
25. Bourigault D. LEXTER, a terminology extraction software for knowledge acquisition from texts. In 9th Knowledge Acquisition for Knowledge Based System Workshop (KAW'95), Banff, Canada, 1995.
26. Assadi H, Bourigault D. Analyses syntaxiques et statistiques pour la construction d'ontologies à partir de textes. In Charlet J, Kassel G, Bourigault D (eds): Ingénierie des Connaissances, évolutions récentes et nouveaux défis. Paris, Eyrolles, 2000:325-336.
27. Manning C, Schütze H (eds): Foundations of Statistical Natural Language Processing. London, The MIT Press, 2000.
28. Unified Modeling Language (UML), version 1.4: available at <http://www.omg.org/technology/documents/formal/uml.htm>, 2002.
29. Kors J, Sittig A, van Bemmel J. The Delphi method to validate diagnostic knowledge in computerized ECG interpretation. Methods Inform Med 1990;29(1):44–50.
30. Evans D, Cimino J, Hersh W, Huff S, Bell D. Toward a medical-concept representation language. The Canon group. J Am Med Inform Assoc 1994;1(3):207–217.
31. Rocha R, Huff S. Development of a template model to represent the information content of chest radiology reports. MEDINFO 2001:251-255.
32. Uschold M. Creating, integrating and maintaining local and global ontologies. In First Workshop on Ontology Learning (OL 2000) in Conjunction with the 14th European Conference on Artificial Intelligence (ECAI 2000). Berlin, 2000.
33. Moorman P, Van Gineken A, Siersema P, Van der Lei J, Van Bemmel J. Evaluation of reporting based on descriptional language. J Am Med Inform Assoc 1995;2(6):365–373.
34. Gouveia-Oliveira A, Raposo V, Salgado N, Almeida I, Nobre-Leitao C, de Melo F. Longitudinal comparative study on the influence of computers on reporting of clinical data. Endoscopy 1991;23(6):334–337.
35. Kuhn K, Gaus W, Weschler J, Janowitz P, et al. Structured reporting of medical findings: Evaluation of a system in gastroenterology. Methods Inform Med 1992;31(4):268–274.
36. Delvaux M, Crespi M, Armengol-Miro J, Hagenmuller F, et al. Minimal standard terminology for digestive endoscopy: Results of prospective testing and validation in the GASTER project. Endoscopy 2000;32(4):345–55.
37. Rocha R, Huff S, Haug P, Evans D, Bray E. Evaluation of a semantic data model for chest radiology: application of a new methodology. Methods Inform Med 1998;37(4–5):477–490.
38. Friedman C, Hripcsak G. Evaluating natural language processors in the clinical domain. Methods Inform Med 1998;37(4–5): 334–344.
39. Bell D, Greenes R. Evaluation of UltraSTAR: performance of a collaborative structured data entry system. Proc Annu Symp Comput Appl Med Care 1994:216–222.
40. Zweigenbaum P, Bouaud J, Bachimont B, Charlet J, Boisvieux J. Evaluation of a normalized conceptual representation produced from natural language patient discharge summaries. Proc AMIA Annu Fall Symp 1997: 590–4.
41. Hripcsak G, Wilcox A. Reference Standards, Judges, and Comparison Subjects: Roles for Experts in Evaluating System Performance. J Am Med Inform Assoc 2002;9(1):1–15.
42. Chute C, Cohn S, Campbell K, Oliver D, Campbell J. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. J Am Med Inform Assoc 1996;3(3):224– 3.
43. Duclos C, Venot A. Structured representation of drug indications: lexical and semantic analysis of drug indications. Meth Inform Med 2000;39:83–7.
44. Gomez-Perez A. Evaluation of ontologies. International Journal of Intelligent Systems 2001;16:391–409.