

Model Formulation ■

Structured Representation of the Pharmacodynamics Section of the Summary of Product Characteristics for Antibiotics: Application for Automated Extraction and Visualization of Their Antimicrobial Activity Spectra

CATHERINE DUCLOS, PHARM.D, PH.D, GIAN LUIGI CARTOLANO, MD, MICHAEL GHEZ, MD, ALAIN VENOT, MD, PH.D

Abstract Objective: The aim of this study was to construct automatically a knowledge base concerning the pharmacodynamic properties of antibiotics and a visualization tool.

Design: The authors studied the various guidelines used to write the pharmacodynamics section of the Summary of Product Characteristics (SPC) for antibiotics and constructed a conceptual model of the information. Particular words, syntagms, and punctuation elements were marked in the SPC texts, and automatic extraction was then used to build a knowledge base. This base was used to create dynamic HTML tables displaying the activity spectra of the antibiotics.

Measurements: The authors analyzed the performances of automatic extraction (recall and precision).

Results: The conceptual pharmacodynamics model dealt with antibiotics, pathogens, susceptibility tests, and the prevalence of resistance. Automatic extraction had a recall rate of 97.9% and a precision of 96.2%. The tool displaying antibiotic spectra and resistance prevalences used color codes to identify differences in susceptibility.

Conclusion: This tool can provide an overview of the prevalence of resistance as expressed in SPC in primary care settings. Its potential impact should be evaluated.

■ *J Am Med Inform Assoc.* 2004;11:285–293. DOI 10.1197/jamia.M1425.

Bacterial resistance to antibiotics is becoming a major problem in health care. Misuse of antibiotics by the physician in terms of molecule choice^{1–5} is a frequently cited reason for the increase in the resistance of microorganisms.^{6–8}

To know which types of antibiotics are appropriate to treat the infection of a given patient, the results of antibiograms should be considered. However, there are various and frequent situations in which the physician must select the treatment without antibiogram data.

In hospitals, antibiograms are easily available, but there is an uncompressible 48-hour delay before identifying the pathogens and their resistance to antibiotics. Therefore, the physician starts the treatment without antibiogram results and may

then modify treatment subsequently in the light of antibiogram data.

The situation is different in primary care settings. When the physician suspects an infection, he or she can ask for antibiogram as in hospitals; however, treatment is very frequently empirical (e.g., in the case of otitis or bronchitis). The physician tries to guess the types of agent responsible and chooses the antibiotics active against these germs. In both cases, the physician needs to take into account the prevalence of the resistance of bacteria to particular antibiotics. The activity spectrum of antibiotics is therefore one of the elements of therapeutic choice.⁹ It gives information about susceptibility or natural resistance of bacteria to particular antibiotics and their acquired resistance prevalence. The prevalence of acquired resistance changes through time and space and also according to patient characteristics (age, disease, style of life). Misinformation and unknown changes about resistance prevalence can lead to the selection of inappropriate antibiotics.^{10–12}

In some hospitals, computerized decision support systems have been developed to help physicians choose empirical antibiotic therapy.^{13–18} Such systems are not available everywhere; in primary care settings, physicians are unlikely to have ready access to computerized tools that could assist them in selecting the most appropriate antibiotic. Physicians should have access to the prevalence of resistance to specific antibiotics.¹⁹ However, although they can consult Web sites devoted to reporting bacterial resistance,²⁰ this information,

Affiliations of the authors: Laboratoire d'informatique médicale et de bioinformatique (LIM&BIO), UFR Santé, Médecine, Biologie Humaine, Université Paris 13, Bobigny, France (CD, MG, AV); Laboratoire de microbiologie, CHI Saint Germain/Poissy, Saint Germain en Laye, France (GLC).

The authors thank Jean François Forget for making the Vidal Drug Database available for this study.

Correspondence and reprints: Catherine Duclos, PharmD, PhD, LIM&BIO, UFR de Santé, Médecine et Biologie Humaine Léonard de Vinci, 74 rue Marcel Cachin 93017 Bobigny cedex, France; e-mail: <catherine.duclos@avc.ap-hop-paris.fr>.

Received for publication: 07/15/03; accepted for publication: 02/04/04.

presented in free text, is not available for all antibiotics. Large observatories for the surveillance of antibiotics resistance are planned at the national and international levels^{21–23}; they give structured information on bacterial resistance, but for only a few bacteria and a few antibiotics.

A main source of information about these activity spectra is the pharmacodynamics section of the Summary of Product Characteristics (SPC) of antibiotics. In this SPC section, the activity spectrum is described, and the national prevalence is given and reevaluated every five years.²⁴ These SPCs are available to physicians either in drug dictionaries or in electronic compendia.^{25,26} However, the consultation of these documents remains very time-consuming, mainly because this information is purely textual. For example, it would take a long time to generate a list of all the antibiotics active against *Streptococcus pneumoniae* or to compare bacterial spectra of various antibiotics.

The characteristics of the pharmacodynamics section of SPCs include a standardized structure and a controlled terminology checked by drug agencies such as the Food and Drug Administration (FDA) and the European Agency for the Evaluation of Medicinal Products (EMA).^{27,28} Therefore, it should be possible to extract this information automatically, to structure and code it, and to store it in a knowledge base. Appropriate presentation of such data would enable the physician to compare at a glance the resistance profiles of one or many bacteria for various antibiotics.

Here, we describe a tool providing visual information about antibiotic activity spectra, which should be useful to physicians prescribing antimicrobial therapy without antibiogram results.

We start by presenting our analysis of the drug regulatory authorities' expectations with respect to the content of the pharmacodynamic section of antibiotics SPCs. This analysis is used to establish a conceptual model of all the pharmacodynamic information contained in the antibiotics SPC. We then present the method for the automatic extraction of knowledge and its evaluation criteria. Finally, we show how the content of the knowledge base can be presented visually to the user by means of a computerized tool. This required the creation of an ontology for bacteria to solve the problem of variability in the names used. We finally discuss the strengths and weaknesses of our approach.

Materials and Methods

Building the Conceptual Model of Pharmacodynamic Information Contained in Antibiotics SPCs

To build the conceptual model of pharmacodynamic information contained in SPCs, we first studied European drug regulatory authorities' expectations with respect to the format and the content of the pharmacodynamic information in the SPC section for antibiotics.²⁸ These expectations are described explicitly in four chapters, the titles of which are General Properties, Breakpoints, Susceptibility, and Other Information. For each chapter, a definition explains what information is expected (e.g., "General properties of the drug should be stated in this section, e.g., classification and mode of action") and how it must be drawn up (e.g., "microorganisms should be listed alphabetically in the following order: aerobic gram-

positive microorganisms, aerobic gram-negative microorganisms, anaerobic microorganisms, and "other" microorganisms"). We considered that these expectations could be used as a "gold standard" to describe the antibiotics' pharmacodynamic properties in SPCs. All the types of information contained in the gold standard were identified to specify an initial set of concepts that should be present in the model built to represent the pharmacodynamic information contained in antibiotics SPC section. For each antibiotic, the pharmacodynamic section must be written according to these recommendations; therefore, all of the types of information found in the gold standard are likely to be found in SPC texts.

Second, we studied American drug regulatory authorities' expectations with respect to the format and the content of the pharmacodynamic information in the SPC section for antibiotics.²⁷ This was done to enrich the initial set of concepts with new types of information not present in European recommendations and to make the conceptual model of pharmacodynamics more general.

Third, we studied the way the identified concepts were incorporated in SPCs. This was done to verify the effective use of concepts found in gold standards and to identify relationships between these concepts.

All the concepts identified in the European and American recommendations and SPCs to develop a conceptual model of pharmacodynamic information as found in the SPCs of antibiotics were finally organized according to classes, attributes, generalization, composition, or association relationships into an object-oriented model representation using Unified Modeling Language (UML) formalism.²⁹ This was done by the author the most directly involved in previous modeling work,^{30–32} and the coherence of the model was checked by the microbiologist author.

Automatic Extraction of Information to Build a Knowledge Base Concerning Antibiotic Pharmacodynamics

Materials

We identified 103 types of systemic antibiotics available in France. A type of systemic antibiotic was defined by an international nonproprietary name (INN) and a route if the route used affects pharmacological properties (e.g., *Haemophilus influenzae* has intermediate susceptibility to cephalexin administered orally and is fully susceptible to cephalexin administered intravenously). All manufactured drugs available in France belonging to any one type of antibiotic (as defined above) must have the same SPC³³; therefore, we extracted a total of 103 SPCs from the Vidal drug database (Vidal, Issy Les Moulineaux, France). This database provides in an electronic format the SPC of each drug available in France. Figure 1 shows an example of SPC text.

Extraction Algorithm

When reading SPC, we identified some characteristic key words or phrases or punctuation elements that delimited information sections. Thus, susceptible species were always listed in the part of text between the elements delimiting the start of the section ("susceptible species" or "usually susceptible species") and those delimiting the end of the

<p>Breakpoints The recommended MIC breakpoints for ceftriaxone, separating susceptible organisms from intermediately susceptible organisms and intermediately susceptible organisms from resistant organisms, are: susceptible ≤ 4 mg/l, resistant >32 mg/l MIC for <i>Streptococcus pneumoniae</i>: susceptible ≤ 0.5 mg/l, resistant >2 mg/l</p> <p>Antibacterial spectrum The prevalence of resistance may vary geographically and with time for selected species and local information on resistance is desirable, particularly when treating severe infections. This information provides only an approximate guidance on probabilities as to whether microorganisms will be susceptible to ceftriaxone. When resistance patterns for particular species are known to vary in France, this is shown below.</p> <p>Susceptible Aerobic Gram-positive bacteria: <i>Staphylococcus</i> methicillin susceptible <i>Streptococcus pneumoniae</i> (15–35%)... Aerobic Gram-negative bacteria: <i>Borrelia burgdorferi</i>, <i>Branhamella catarrhalis</i>, <i>Citrobacter freundii</i> (20–30%), <i>Citrobacter koseri</i>, <i>Enterobacter</i> (20–40%), <i>Escherichia coli</i>, ... Anaerobic bacteria: <i>Clostridium perfringens</i>, <i>Fusobacterium</i> (15–20%), <i>Peptostreptococcus</i>, <i>Prevotella</i> (15–20%)</p> <p>Resistant Aerobic Gram-positive bacteria: <i>Listeria</i>, <i>Staphylococcus</i> methicillin resistant... Aerobic Gram-negative bacteria: <i>Acinetobacter baumannii</i>, <i>Burkholderia cepacia</i>... Anaerobic bacteria: <i>Bacteroides fragilis</i>, <i>Clostridium difficile</i></p>
--

Figure 1. Extract of the pharmacodynamics section of the French Summary of Product Characteristics (SPC) for ceftriaxone translated in English. MIC = minimal inhibitory concentration.

section (“moderately susceptible species,” “inconstantly susceptible species,” “resistant species,” or “usually resistant species”).

Once the characteristic tags were marked, a sequential extraction algorithm was developed. This algorithm was designed to extract information concerning the antibacterial spectrum, which is the most structured section. A heading was extracted (e.g., the “susceptible species” heading) and divided into a number of subheadings (e.g., “gram-positive aerobes,” “gram-negative aerobes,” “anaerobes,” “others”). All the bacteria under a given heading were then extracted, together with their prevalences of resistance and clinical efficiencies. The extracted data were stored in a relational database, the structure of which corresponded to the pharmacodynamic conceptual model previously developed.

Extraction Quality Evaluation

We assessed the performance of the extraction algorithm by comparing automatically extracted spectra with the spectra initially reported in the 103 studied SPCs.

A spectrum extracted from an SPC consists of the information about the pathogen (name, mode of respiration, Gram staining) and that about the effect of the antibiotics against this pathogen (susceptibility, prevalence of resistance).

Let A be the number of spectra automatically extracted using the extraction algorithm, and B the number of spectra that were manually extracted by the microbiologist author reading the SPCs. Let C be the number of spectra automatically extracted that were considered to match exactly with the spectra manually extracted, and D the number of spectra automatically extracted that did not match exactly with the spectra manually extracted.

We calculated four indicators:

- Recall = C/B
- Silence = $(B-C)/B$
- Precision = C/A
- Noise = D/A

Building an Application to Visualize Antimicrobial Spectrum of Antibiotics

The application developed was designed to compare antibiotic activity against that of one or many selected pathogens. The program uses Hypertext PreProcessor (PHP) scripts to treat extracted data from SPC and store the data in a relational database. The following PHP script shows how the parameter “selected bacteria” is passed to a query that returns a list of antibiotics associated with susceptibility:

```
$germe = $_GET['selectedbacteria'];
$query = « SELECT * FROM Extraction WHERE
          Extraction. Bacteria = $selectedbacteria; »;
$result = mysql_query($result);
```

The result is displayed in Hypertext Markup Language (HTML) tables, which are created dynamically.

When selecting the pathogen (as manually extracted from SPCs), some difficulties appeared: pathogens could be defined by species (e.g., *Escherichia coli*), genus (e.g., *Acinetobacter*), group of bacteria (e.g., coagulase-negative staphylococci, gram-negative cocci), or group of pathogens after the exclusion of various bacteria (e.g., gram-negative aerobic bacilli other than *Aeromonas hydrophila*).

To correctly answer to the query “which antimicrobial agents are active against *E. coli*?,” the software, thus, must look for all the possible names and ways of describing this pathogen in the SPCs (*E. coli*, *Escherichia coli*, enterobacteria, gram-negative aerobic bacillus, aerobic bacteria). The program, thus, needs to know that “*E. coli*” is equivalent to “*Escherichia coli*” and that “*E. coli*” is an “aerobe,” for example.

The input of the system (selected bacteria) should then be associated with an inference system able to explore synonymy and hierarchical relationships between bacterial names extracted from the SPCs to identify all the pathogen names associated with that selected. Thus, if *Escherichia coli* is initially selected, the inference system will propose as an input of the query “*Escherichia coli* or *E. coli* or enterobacteria

or gram-negative aerobic bacillus or gram-negative bacillus or aerobes" as shown by the following PHP script:

```
$selected_bacteria = $_GET['selectedbacteria'];
//terminologic_search() returns an array of all terminologic
  variants for the selected bacteria
$germ = terminologic_search($selected_bacteria);
for($i = 0 ;$i < count($germ)-1 ;$i++)
{
$condition.= 'extraction.bacteria='.$germ[$i]. ' OR ' ;
}
$condition.= 'extraction.bacteria='.$germ[count($germ)];
$query = « SELECT * FROM Extraction WHERE
  $condition; »;
$result=mysql_query($result);
```

Terminology used to describe pathogens found in SPCs should, therefore, be organized on the basis of synonymy or hierarchical relationships.

To classify the pathogens manually extracted from SPC, the following principles were used:

- The classification must be based on morphologic and respiratory characteristics
- All the pathogen names found in SPCs must be described in the classification

We, therefore, analyzed the various electronic microbiological nomenclatures available: the *Deutsche Sammlung von Mikroorganismen und Zellkulturen* (DSMZ bacterial nomenclature)³⁴ and the List of Bacterial Names with Standing in Nomenclature.³⁵ We also analyzed the terminological resources Galen,³⁶ Unified Medical Language System (UMLS),³⁷ and Medical Subjects Headings (MeSH).³⁸

None of these classifications fully met our needs. We, therefore, had to build a hierarchical classification of pathogens strictly limited to the bacteria named in the SPCs. This work was done by the microbiologist author adhering to the rules of bacterial taxonomy.³⁹

Results

Concepts Identified from European and American Recommendations and SPCs

The concepts identified in the European (EMA) and American (FDA) recommendations and in SPCs are shown in Table 1. The European and American recommendations are quite similar. The EMA is more focused on reporting the prevalence of resistance, whereas the FDA has more specific recommendations about the methods used to measure breakpoints. The concepts found in SPCs are those found in European recommendations. Some discrepancies appear in the identification of the bacterial species according to their respiration and Gram staining: for one antibiotic the nonrecommended concepts of "gram-negative bacteria" and "gram-positive bacteria" are used, and for nine antibiotics, some bacteria are displayed without type of respiration or Gram-staining characteristics.

Table 1 ■ Concepts Found in European or American Recommendations or Summary of Product Characteristics (SPCs) to Describe the Pharmacodynamic Section of SPCs for Antibiotics

General Concepts	Specific Concepts	European Recommendations	FDA Recommendations	SPC
General properties	Classification	Yes	Yes	Yes
	Mode of action	Yes	Yes	Yes
	ATC Code	No	No	Yes
Breakpoint	Susceptible breakpoint	Yes	Yes	Yes
	Resistant breakpoint	Yes	Yes	Yes
	Inhibition diameter	No	Yes	No
	Minimal inhibitory concentration	Yes	Yes	Yes
	Panel of tested strain	No	Yes	No
Susceptibility	Susceptible bacterial species	Yes	Yes	Yes
	Intermediate susceptibility species	Yes	Yes	Yes
	Unsusceptible species	Yes	Yes	Yes
	Aerobic gram-positive microorganisms	Yes	Yes	Yes
	Aerobic gram-negative microorganisms	Yes	Yes	Yes
	Anaerobic microorganisms	Yes	Yes	Yes
	"Other" microorganisms	Yes	Yes	Yes
	Prevalence of acquired resistance	Yes	No	Yes
	Gram-negative bacteria	No	No	Yes
	Gram-positive bacteria	No	No	Yes
	Unqualified Gram staining and respiration	No	No	Yes
	Clinical efficiency	Yes	Yes	Yes
	Other information	Occurrence of cross-resistance	Yes	No
Antimicrobial agents for which cross-resistance occurs		Yes	No	Yes
Mechanism of resistance		Yes	No	Yes
Level of resistance		No	No	Yes
Frequency of mutation		No	No	Yes
Condition for emergence		No	No	Yes

FDA = Food and Drug Administration; ATC = Anatomical Therapeutic Classification.

The reading of the SPC gives a more specific idea of what is expected in the section dealing with resistance.

Conceptual Model of Pharmacodynamic Information Contained in SPCs

The conceptual model of antibiotic pharmacodynamics was directly deduced from the conceptual analysis described in the previous paragraph according to UML formalism. In this formalism:

- generalization relations are shown by a solid lined path from the child to the parent with a large hollow triangle at the end of the path where it meets the more general element
- composition relations are shown by a solid filled diamond as an association end adornment
- associations are shown as lines

In Figure 2, the classes of the model of antibiotic pharmacodynamics are represented.

The entity *antibiotic* is defined by its *international nonproprietary name* (e.g., ceftriaxone), its *code in the Anatomical Therapeutical Classification (ATC; e.g., J01)*, its *route* (e.g., per

os, systemic, intravenous), and its *mode of action* (e.g., inhibiting protein synthesis). The pharmacodynamic properties of this antibiotic determine its action against a set of pathogenic bacteria.

The entity *organism* is described in terms of *name* (e.g., *Streptococcus pneumoniae*, *E. coli*, enterobacteria), *Gram staining* (e.g., gram-positive, gram-negative), *mode of respiration* (e.g., aerobic, anaerobic), and *resistance profile* (e.g., resistant to methicillin).

The activity of the antimicrobial agent against the pathogen can be measured by means of in vitro susceptibility tests.

The entity *susceptibility test* is defined by *minimal inhibitory concentrations (MICs)*, *inhibition diameters (IDs)*, a *susceptibility* (susceptible, intermediate, resistant) established according to the *breakpoints of the MIC* or of the *ID*. Susceptibility in vitro may have been confirmed in vivo (clinical trials confirming *clinical efficiency*).

The entity *resistance prevalence* is defined in terms of *minimal* and *maximal* values, which vary according to *geographic situation* (e.g., France), *origin of the strain* (e.g., nosocomial or community strain), *physiological condition of the host* (e.g.,

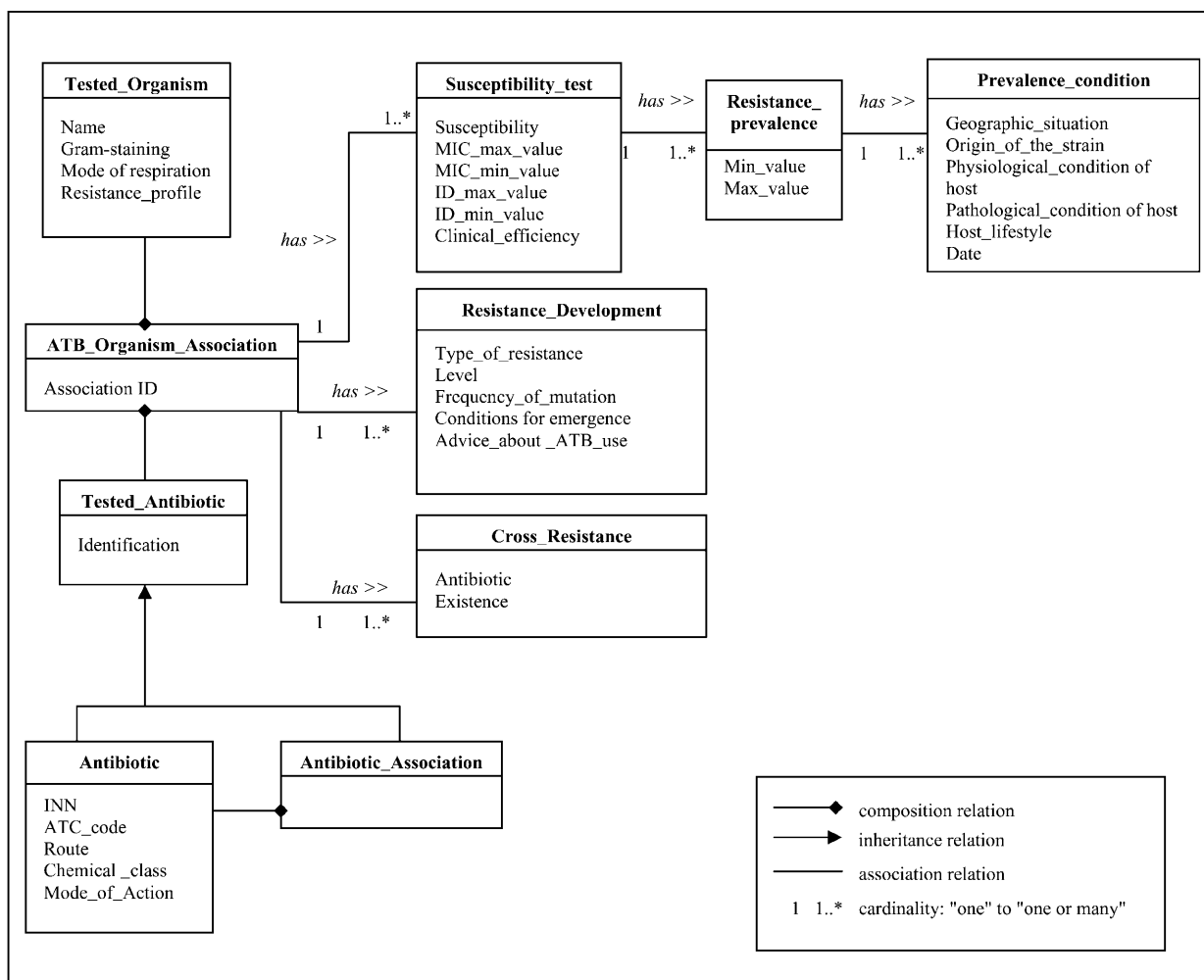


Figure 2. Antibiotic pharmacodynamics conceptual model according to Unified Modeling Language (UML) formalism. MIC = minimal inhibitory concentration; ID = inhibition diameter; ATB = antibiotic; INN = international nonproprietary name; ATC = anatomical therapeutical classification.

pregnancy), *pathological condition of the host* (e.g., immunodeficiency), *host lifestyle* (e.g., schoolchild), and *time*.

The entity *resistance development* is defined by a *type of resistance* (e.g., chromosomal or plasmid-borne), together with a *level* (e.g., low or high), a *frequency of mutation*, and the *conditions* required for the development of resistance. These data may lead to issuance of *advice about the use* of the antimicrobial drug (e.g., alone or in association with cases of severe infection).

The entity *cross-resistance* is defined by its existence and the list of *antibiotics* for which cross-resistance exists. An example of instantiation of the model is shown in Figure 3.

Automatic Extraction

Manual analysis of the 103 SPC identified 2,914 spectra. The automatic extraction process yielded 2,966 spectra: 2,852 spectra were correctly extracted, and 114 spectra were not accurately extracted. Recall was 97.9%, silence was 2.1%, accuracy was 96.2%, and noise was 3.8%.

Incorrect extraction can be due to several causes, including limitations of the extraction algorithm and poor quality of the text. The following errors due to the algorithm were identified:

- a string between parentheses was considered to be prevalence information, whereas in some cases it was details concerning the pathogen (e.g., *Staphylococcus aureus* [methicillin-resistant])
- decomposition failure, for example, the group of words "*streptococcus* susceptible or resistant to penicillin G" was considered as a single pathogen, whereas the expert considered two pathogens: "*streptococcus* susceptible to penicillin G" and "*streptococcus* resistant to penicillin G"

Examples of extraction errors due to the quality of the text included:

Tested antibiotic
<u>INN</u> = ceftriaxone
<u>Route</u> = intravenous
<u>Chemical class</u> = betalactam
Tested organism
<u>Name</u> = <i>Citrobacter freundii</i>
<u>Gram-staining</u> = negative
<u>Mode of Respiration</u> = aerobic
ATB_Organism_Association
has <u>Sensitivity_Test</u>
<u>Susceptibility</u> =Susceptible
<u>MIC_max_value</u> =4 mg/l
has <u>Resistance_Prevalence</u>
<u>Min_Value</u> = 20%
<u>Max_Value</u> =30%
has <u>Prevalence_Condition</u>
<u>Geographic_Condition</u> =France

Figure 3. Example of instantiation of the pharmacodynamic model for the tested association: ceftriaxone–*Citrobacter freundii*. *Citrobacter freundii*, which is an aerobic gram-negative bacteria, is susceptible to intravenous ceftriaxone (MIC \pm 4 mg/L), with a prevalence of resistance in France of 20% to 30%. INN = international nonproprietary name; ATB = antibiotic; MIC = minimal inhibitory concentration.

- misinterpretation of abbreviations, for example, a pathogen listed as *C. freundii* was decomposed into two pathogens "C" and "freundii" instead of "*Citrobacter freundii*"
- inability to deal with inaccurate punctuation in the text, for example, "*Actinobacillus; actinomycetemcomitans*" was found instead of "*Actinobacillus actinomycetemcomitans*"
- some resistance chapters were not preceded by a specific heading and were computed as antibacterial spectrum information

Computerized Tool for Spectrum Visualization

Bacteria Classification Systems

The existing electronic DSMZ bacterial nomenclature and List of Bacterial Names with Standing in Nomenclature are nomenclatures and not classifications. They cover all microorganisms, regardless of whether they are pathogenic and organize them according to family, genus, and species, but no information about the mode of respiration or Gram staining of the bacteria is available.

In the UMLS metathesaurus, concepts indicating pathogens are all of the semantic type "bacterium." Moreover, all of these concepts are on the same level (e.g., "cocci" is not a parent of "streptococcus"). In MeSH, bacteria are organized hierarchically, according to morphology and Gram staining, but the mode of respiration is not considered. In Galen, bacteria are organized hierarchically according to morphology (bacilli, cocci) and species. Their properties in terms of respiration profile and Gram staining are also described, but not all the pathogens listed in SPCs are included (e.g., methicillin-resistant staphylococcus).

The manual extraction of spectra identified 279 terms naming pathogens. Eighty-seven synonyms were identified, and the remaining 192 terms could be classified. Four new pathogen terms were introduced to ensure the coherence of the classification: "gram-negative anaerobic bacilli," "gram-positive anaerobic bacilli," "gram-negative anaerobic cocci," and "non fermentative gram-negative aerobic coccobacilli."

Spectrum Visualization

The system displays an HTML table (Fig. 4) with the selected bacteria in the columns and the types of antibiotic, grouped according to chemical class (e.g., ceftriaxone IV, which is a third-generation cephalosporin, a type of betalactam) in the rows.

In each cell, the susceptibility of the bacterium is given, taking into account the minimal and maximal values for the prevalence of resistance. To make it easier to read the susceptibility data, color coding or gray scaling (green or pale gray: susceptible; orange or gray: intermediate; red or dark gray: resistant) has been used. The corresponding cell in the table is divided into "susceptible," "intermediate," and "resistant" zones, the areas of which are proportional to prevalence.

A request based on "*Escherichia coli*" was made to illustrate the value of the inference system based on our classification of pathogens. Sixty-three antibiotics were returned when the request was limited to the selected bacterial name, and 84

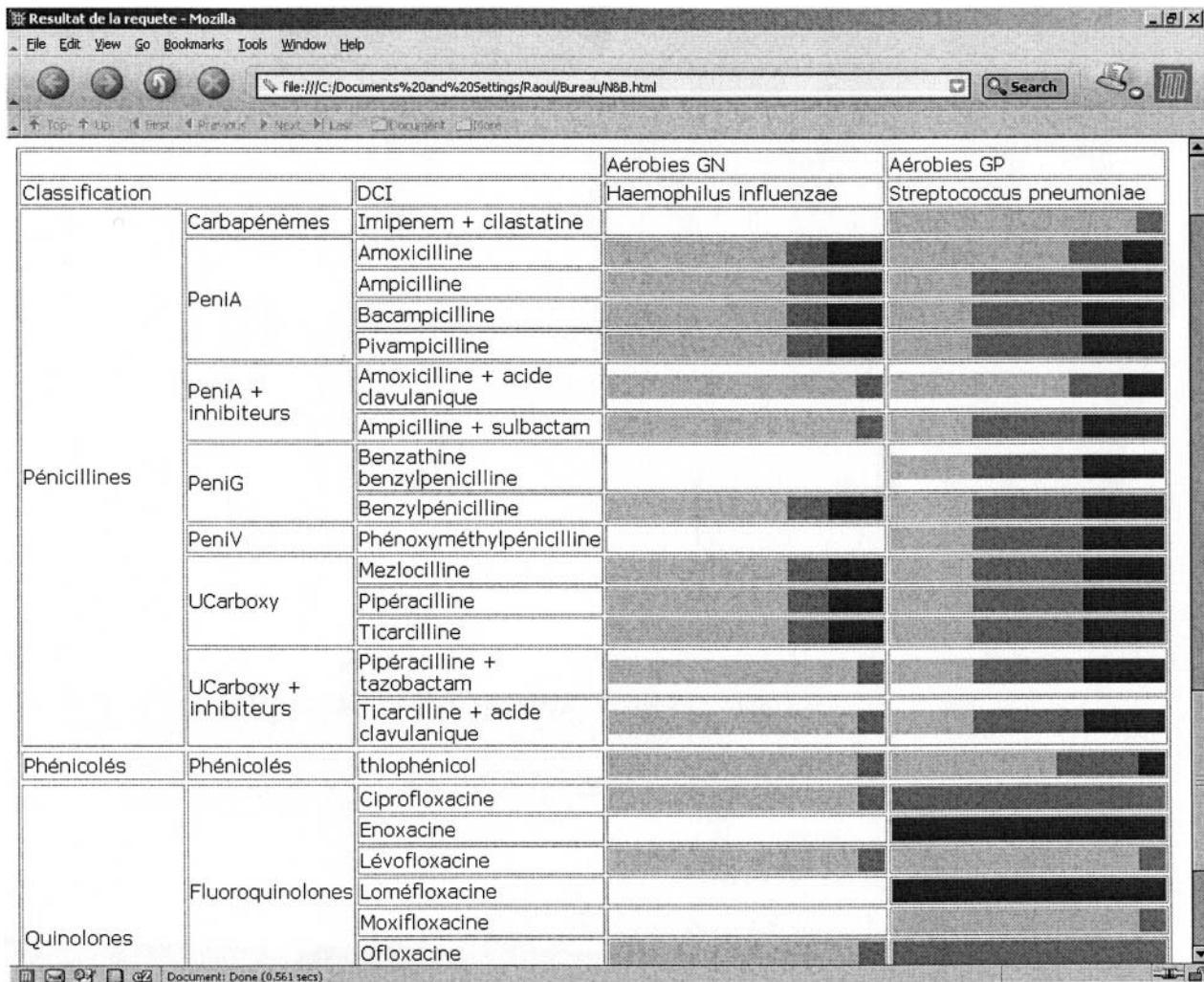


Figure 4. Antibiotic pharmacodynamics visualization tool. The activity spectra against *Haemophilus influenzae* and *Streptococcus pneumoniae* are displayed. The physician can see the antibiotics likely to cure otitis in a child. Legend of gray scaling: pale gray: susceptible bacteria; medium gray: intermediate-susceptibility bacteria; dark gray: resistant bacteria.

antibiotics were returned when the request was enhanced to all terminological variants of the selected bacterial name.

Discussion and Conclusion

We aimed to develop a tool to improve the access to existing official information about the susceptibility of bacteria to various antibiotics so as to help practitioners prescribe when they do not have antibiogram data available. This official information can be found in the pharmacodynamic section of SPCs.

The SPC is a validated document describing the properties of a given drug and is widely available electronically from many drug databases. SPCs have already been successfully used to develop models of information concerning indications³⁰ and pharmacokinetics.³¹

The methods used to model information can be based on a bottom-up approach, lexical units describing the domain being used to identify concepts.^{32,40,41} The method can also be based on a top-down method.⁴¹ This is the approach we used to develop the model representing the information contained in the pharmacodynamic section of SPCs. The pharmaco-

dynamic section has the unusual characteristic of having both structure and terminology standardized. This is because the section is written according to recommendations of drug regulatory agencies; these recommendations give a precise definition of the different chapters the section must contain and the way information must be displayed. These recommendations are considered to be the gold standard for SPCs. All the concepts identified in the gold standard are inevitably present in SPCs. The inverse, however, is not true: not all of the concepts found in SPCs are found in the gold standard. These discrepancies are due to nonrespect of recommendations (e.g., using new terms to classify pathogens) but also to a lack of precision in the expected content of some chapters (e.g., the resistance chapter does not deal with frequency of mutation).

The model that we propose is not new but has not previously been formalized. The same concepts are found in computerized empiric therapy advisors^{14-16,18,42,43} and susceptibility reporting systems⁴⁴⁻⁴⁶ indicating the coherence of the model.

The completeness of the model was not evaluated. Concepts outside of the gold standard may be poorly defined because

we did not explore those parts of SPC texts dealing with these concepts at lexical and semantic levels. We integrated a temporal qualifier to fix the date of the reporting of prevalence of resistance of a bacterium because this information can change with time, and it may be interesting to keep track of this evolution. We also integrated a class concept called the "prevalence condition" to allow inclusion of data coming from other sources (such as laboratory results).

The model is limited to the description of an antibiotic-bacterium pair. In practice, the practitioner knows a disease but not necessarily all the bacteria that can be responsible for the disease.¹² We will have to integrate a link between a given disease and all the possibly causative bacteria.

Standardization of the writing of the pharmacodynamics section of antibiotics SPCs allows the automatic extraction of information to be fed into a knowledge base. The extraction was limited to activity spectra of antibiotics, because it was the most structured and interesting information for the development of our application. As the level of syntactic variability in these parts of texts is low, extraction rules based on syntactic criteria could be established. The extraction method does not apply to other chapters of the pharmacodynamic section (such as the resistance chapter) because they are expressed in natural language. An extraction method coupled to natural language processing techniques may allow automatic handling of these parts of texts.

The efficiency of the extraction algorithm was validated on the basis of classical criteria according to Friedman and Hripcsak⁴⁷: the model was frozen before the extraction, the reference standard was established by the microbiologist on the basis of an exhaustive set of SPCs, and the microbiologist was not directly involved in the development of the system. We did not distinguish a test set and an evaluation set among the SPC texts because there were too few. The reference standard was made, blindly, by a single expert; thus, we cannot display variability in standard reference.

The extraction of antimicrobial spectrum was evaluated for correctness. According to our criteria, the extraction method was satisfactory but not perfect. Correctness might be improved with the modification of the extraction algorithm (such as new treatment of strings between parentheses), but other factors cannot be modified, such as the quality of source text writing. This insufficient quality of the text source (SPCs) has been noted elsewhere.³¹ These imperfect results imply that an extracted data correction step must be included in the system.

The use of pharmacodynamic knowledge bases in computerized systems for comparing antibiotics raises the problem of heterogeneity in the terminology used to describe bacteria. Variability in terminology is also observed in the decision support systems used for antibiotic therapy¹⁶ and in systems of susceptibility reporting.⁴⁵ The ICONS system⁴² uses a hierarchical classification built locally by the microbiology laboratory. We constructed a hierarchy of the terms used to refer to bacteria in SPCs. This makes it possible to optimize searches for antibiotics active against a particular bacterium. Inference systems based on terminological models have previously been demonstrated to be of value in searches for information.⁴⁸ We will need to adapt our classification in the future in response to inevitable changes in bacterial nomen-

clature (e.g., *Moraxella catarrhalis* is now known as *Branhamella catarrhalis*).

We chose to develop an application for comparing activity spectra of antibiotics against selected bacteria. According to Schneiderman's approach, we believe that this visual representation of the data makes it possible to present a large volume of complex data in a very intuitive, easy-to-understand, and easy-to-learn format.⁵⁰ Nevertheless, the application will be further evaluated among a user group. The approach we chose to present information visually is limited to the extract and overview tasks (gain an overview of the entire set of data).⁴⁹ Other elements could be added to the tool such as zooming (e.g., when an antibiotic is selected, its dosage can be displayed), filtering (e.g., the selection can be limited to the antibiotics for which bacteria are fully susceptible), and historical considerations (e.g., display the evolution of the prevalence of resistance of a bacteria for one or many antibiotics).

Our activity spectrum visualization tool aims to fill a gap¹⁹ in the provision of prevalence resistance information in primary care settings. The potential impact of this tool on the routine practice of doctors requires further evaluation. Does it modify the doctor's knowledge? Does it improve the quality of antibiotic prescription? We are currently designing an impact study in a primary care setting to answer these questions.

References ■

1. Maki D, Schuna A. A study of antibiotic misuse in a university hospital. *Am J Med Sci.* 1978;275:271-82.
2. Dunagan W, Woodward R, Medoff G, Gray J, Casabar E, Lawrenz C. Antibiotic misuse in two clinical situations: positive blood cultures and administration of aminoglycosides. *Rev Infect Dis.* 1991;13:405-12.
3. Leibovici L, Konigsberger H, Pitlik S, Samra Z, Drucker M. Patients at risk for inappropriate treatment of bacteremia. *J Intern Med.* 1992;231:371-4.
4. Macaraeg P, Lasagna L, Bianchine J. A study of hospital staff attitudes concerning the comparative merits of antibiotics. *Clin Pharm Ther.* 1971;12:1-12.
5. Edwards L, Levin S, Lepper M. A comprehensive surveillance system of infections and antimicrobials used at Presbyterian-St. Luke's Hospital. *Am J Public Health.* 1972;62:1053-5.
6. Goldman D, Weinstein R, Wenzel R. Strategies to prevent and control the emergence and spread of antimicrobial resistant microorganisms in hospital. Consensus statement. *JAMA.* 1996; 275:234-40.
7. Magee J, Pritchard E, Fitzgerald K, Dunstan F, Howard A. Antibiotic prescribing and antibiotic resistance in community practice: retrospective study 1996-1998. *BMJ.* 1999;319:1239-40.
8. Seppala H, Klaukka T, Lehtonen R, Nenonen E, Huovinen P. Outpatient use of erythromycin: link to increased erythromycin resistance in group A streptococci. *Clin Infect Dis.* 1995;21: 1378-85.
9. Mantous C, Amoateng-Adjepong Y. Empiric antibiotic use and resistant microbes: a "Catch-22" for the 21st Century. *Chest.* 2000;118(1):9-11.
10. Moss F, McSwiggan D, McNicol M, Miler D. Survey of antibiotic prescribing in a district general hospital. I. Pattern of use. *Lancet.* 1981;2:349-52.
11. McGowan J. Antimicrobial resistance in hospital organisms and its relation to antibiotic use. *Rev Infect Dis.* 1983;5:1033-48.
12. Yu V, Stoehr G, Starling R, Shogan J. Empiric antibiotic selection by physician: evaluation of reasoning strategies. *Am J Med Sci.* 1991;301:165-72.

13. Shortliffe E, Axline S, Buchanan B, Merigan T, Cohen S. An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Comput Biomed Res.* 1973;6:544–60.
14. Kristensen B, Andreassen S, Leibovici L, Riekehr C, Kjaer A, Schonheyder H. Empirical treatment of bacteraemic urinary tract infection. Evaluation of a decision support system. *Dan Med Bull.* 1999;46:349–53.
15. Evans R, Pestotnik S, Classen D, Burke J. Development of an automated antibiotic consultant. *MD Comput.* 1993;10(1):17–22.
16. Leibovici L, Gitelman V, Yehezkeli Y, et al. Improving empirical antibiotic treatment: prospective, nonintervention testing of a decision support system. *J Intern Med.* 1997;242:395–400.
17. Barrufet P, Puyol-Gruart C. Terap-IA, a knowledge-based system for pneumonia treatment. Proceedings of the international ICSC symposium on engineering of intelligent systems 1998;1: 176–82.
18. Warner H, Blue S, Sorenson D, et al. New computer-based tools for empiric antibiotic decision support. *Proc AMIA Annu Fall Symp.* 1997:238–42.
19. Sexton D. The impact of antimicrobial resistance on empiric antibiotic selection and antimicrobial use in clinical practice. *J Med Liban.* 2000;48:215–20.
20. Schindler J, Schindler Z, Schindler J. A WWW-based information system on resistance of bacteria to antibiotics. *Med Inform.* 1998; 23:179–85.
21. Grüneberg R, Felmingham D, Harding I, Shrimpton S, Nathwani A. The Nearchus project: antibiotic susceptibility of respiratory pathogens and clinical outcome in lower respiratory tract infection at 27 centres in the UK. *Int J Antimicrob Agents.* 1998;10:127–33.
22. Inoue M. What can PROTEKT tell us at local level? *J Chemother.* 2002;14(suppl 3):17–24.
23. Observatoire national de l'épidémiologie de la résistance bactérienne aux antibiotiques (ONERBA). *Res-ONERBA.* Available at: <http://www.onerba.org/fr/res-onerba/>. 2003. Accessed May 3, 2004.
24. Powell M. Licensing new antibacterial agents—a European perspective. *Int J Antimicrob Agents.* 2000;16:199–203.
25. RxList. Available at: <http://www.rxlist.com>. Accessed May 3, 2004.
26. OVP-Vidal. Vidal pro. Available at: <http://www.vidalpro.net>. Accessed May 3, 2004.
27. Food and Drug Administration. Point to consider: clinical development and labeling of anti-infective drug product. Available at: <http://www.fda.gov/cder/guidance/ptc.htm>. Accessed May 3, 2004.
28. European Agency for the Evaluation of Medicinal Products. Note for guidance on the pharmacodynamic section of the SPC for anti-bacterial medicinal products (CPMP/EWP/520/96). Available at: <http://www.emea.eu.int/pdfs/human/ewp/052096en.pdf>. Accessed May 3, 2004.
29. Unified Modeling Language (UML), version 1.4. Available at: <http://www.omg.org/technology/documents/formal/uml.htm>. Accessed May 3, 2004.
30. Duclos C, Venot A. Structured representation of drug indications: lexical and semantic analysis of drug indications. *Methods Inf Med.* 2000;39(1):83–7.
31. Duclos-Cartolano C, Venot A. 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. *J Am Med Inform Assoc.* 2003;10:271–80.
32. Venot A, Duclos C. Methodology for the analysis and representation of the medical information about drugs in the Summary of Product Characteristics (SPC). *Proc AMIA Annu Fall Symp.* 1999:166–70.
33. AFSSAPS. Spectre d'activité antibactérienne. Available at: <http://agmed.sante.gouv.fr/pdf/5/atb.pdf>. Accessed May 3, 2004.
34. Deutsche Sammlung von Mikroorganismen und Zellkulturen. Bacterial Nomenclature. Available at: <http://www.dsmz.de/bactnom/bactname.htm>. Accessed May 3, 2004.
35. List of Bacterial Names with Standing in Nomenclature. Available at: <http://www.bacterio.cict.fr>. Accessed May 3, 2004.
36. OPEN GALEN Open sources. Available at: <http://www.opengalen.org/resources.html>. Accessed May 3, 2004.
37. UMLS: unified medical language system. Available at: <http://www.nlm.nih.gov/research/umls/>. Accessed May 3, 2004.
38. MeSH: Medical Subject Heading. Available at: <http://www.nlm.nih.gov/mesh/meshhome.html>. Accessed May 3, 2004.
39. Garrity G, Bell J, Lilburn T. Taxonomic outline of the pro-caryotes. In: *Bergey's Manual of Systematic Bacteriology* (ed 2). New York: Springer-Verlag, 2003.
40. Rocha R, Huff S. Development of a template model to represent the information content of chest radiology reports. *Medinfo.* 2001:251–5.
41. 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;1:207–17.
42. Heindl B, Schmidt R, Schmid G, et al. A case-based consiliarius for therapy recommendation (ICONS): computer-based advice for calculated antibiotic therapy in intensive care medicine. *Comput Meth Programs Biomed.* 1997;52:117–27.
43. Gierl L, Pollwein B, Wolter M, Schmidt R. Focusing on resistance development in a case based teleconsultation system for antibiotics therapy advice. *Medinfo.* 1998;9(pt1):549–53.
44. Schmidt R, Gierl L. Case-based reasoning for antibiotics therapy advice: an investigation of retrieval algorithms and prototypes. *Artif Intell Med.* 2001;23:171–86.
45. Willard K, Connelly D, Johnson J. Radical improvements in the display of clinical microbiology results: a Web based clinical information system. *Am J Med.* 1996;101:541–9.
46. Sharp S. Effective reporting of susceptibility test results. *Diagn Microbiol Infect Dis.* 1993;16:251–4.
47. Friedman C, Hripcsak G. Evaluating natural language processors in the clinical domain. *Methods Inf Med.* 1998;37:334–44.
48. Brown J, Sonksen P. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. *J Am Med Inform Assoc.* 2000;7:392–403.
49. Chittaro L. Information visualization and its application to medicine. *Artif Intell Med.* 2001;22:81–8.
50. Schneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. *Proc IEEE Symp Visual Language.* 1996:336–43.