Une expérience originale de test automatisé d'un système d'aide à la décision clinique basé sur les guides de bonnes pratiques

Anis Ellini, Vahid Ebrahiminia, Alain Venot, Jérôme Nobécourt

▶ **To cite this version:**

Anis Ellini, Vahid Ebrahiminia, Alain Venot, Jérôme Nobécourt. Une expérience originale de test automatisé d'un système d'aide à la décision clinique basé sur les guides de bonnes pratiques. 18es Journées Francophones d'Ingénierie des Connaissances, Jul 2007, Grenoble, France. not specified. hal-00509866

HAL Id: hal-00509866
https://hal.science/hal-00509866

Submitted on 16 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une expérience originale de test automatisé d'un système d'aide à la décision clinique basé sur les guides de bonnes pratiques

Anis Ellini[1,2], Vahid Ebrahiminia[1], Alain Venot[1] et Jérôme Nobécourt[1]

[1] Laboratoire d'Informatique Médicale et Bioinformatique (EA 3969-LIM&BIO), UFR SMBH, Université Paris 13
{j.nobecourt, avenot}@smbh.univ-paris13.fr, vebrahiminia@yahoo.com

[2] Département santé-publique. Hôpital Avicenne. AP-HP. Bobigny
anis.ellini@avc.aphp.fr

Résumé : Le test rigoureux des systèmes informatisés d'aide à la décision avant leur diffusion pour utilisation par les médecins est primordial. Ce test doit être systématique et exhaustif pour assurer la fiabilité et l'innocuité des recommandations produites. Dans ce travail, nous proposons une méthode de test pour le « module critique » du système ASTI, un système informatisé d'aide à la stratégie thérapeutique basée sur les guides de bonnes pratiques développé par le Lim&Bio. À partir des « spécifications papiers » du module critique, nous avons créé des organigrammes représentant les algorithmes d'exécution du moteur et une table de décision pour visualiser les règles de la base de connaissances du système. En nous basant sur ces organigrammes et cette table de décision, nous avons ensuite cherché à diviser l'ensemble de tous les cas cliniques possibles (les vecteurs d'entrée) en classes déclenchant des réponses proches. Nous avons réussi à identifier onze classes de cas cliniques que nous avons associés aux recommandations de traitements attendues. Les critères de classifications nous ont alors permis de générer automatiquement les cas cliniques nécessaires au test automatique du « module critique ».

Mots-clés : GBP, tests de modules d'un SBC, modélisation de règles, classes de cas cliniques, SADM

1 Introduction

Le médecin complète sa formation initiale par une formation continue tout au long de sa carrière. Pour l'aider à maintenir à jour son savoir faire, il dispose de Guides de Bonnes Pratiques Cliniques (GBP). Ces guides sont rédigés en France par des groupes d'experts agissant pour le compte de la Haute Autorité de Santé, organisme public indépendant. Chaque guide contient les recommandations de prise en charge pour une pathologie donnée. Cette prise en charge peut concerner de façon non exhaustive les moyens de dépistage, de prévention ou de traitements. Pour formuler ces recommandations, le comité se base en priorité sur les derniers résultats des recherches publiées et sur des avis d'experts.

Un médecin généraliste est confronté successivement au cours d'une même journée de travail, à des pathologies très variées (cardiaques, digestives, ORL,…). De plus, pour une même pathologie il doit résoudre différentes situations de prise en charge. L'éventail des situations va de l'urgence à poser un diagnostic chez un patient qui va mal, au renouvellement d'un traitement chez un patient dont la pathologie est stable, en passant par l'initiation ou la modification d'un traitement.

Se maintenir à jour et intégrer dans sa pratique quotidienne toutes les recommandations disponibles est une tâche très ardue quand on doit prendre en charge autant de situations cliniques et de pathologies différentes. L'évaluation des stratégies thérapeutiques des médecins montre d'ailleurs que celles-ci sont loin d'être optimales (Coste & Venot, 1998 ; Buetow et al, 1996).

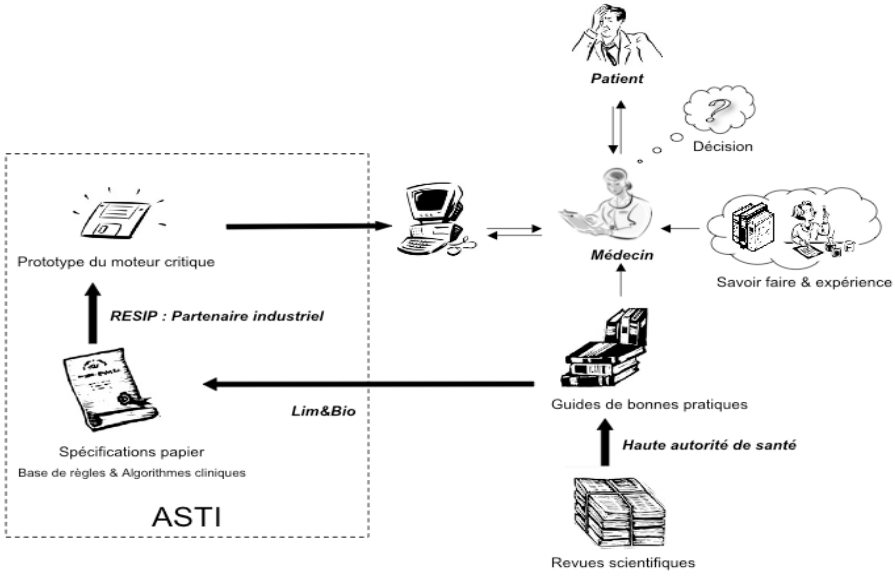

Fig. 1 – Où se trouve la connaissance qui aide le médecin à prendre une décision pour son patient ?

Les premiers guides de bonnes pratiques mis à la disposition des médecins étaient sous format « papier ». Les évaluations ont très vite montré que ce type de support n'était pas adapté à la consultation du médecin généraliste.

Plusieurs types de systèmes informatiques d'Aide à la Décision Médicale (SADM) ont été alors développés et ce d'autant plus qu'il a été démontré qu'ils pouvaient favoriser le respect des guides de bonnes pratiques par les médecins dans leur exercice quotidien (Rood et al, 2005) et qu'ils pouvaient donc améliorer la pratique médicale (Garg et al, 2005 ; Kawamoto et al, 2005).

Ces systèmes ont exploré principalement deux thématiques de recherche. Une première s'intéresse à définir des normes et des méthodes de développement de guides de bonnes pratiques qui soient directement implémentables. La deuxième s'intéresse à la création de systèmes à base de connaissances. Le projet ASTI (Aide à la Stratégie Thérapeutique informatisée) fait partie de cette deuxième catégorie.

Il a pour but le développement pour les médecins généralistes français d'un système informatisé d'aide à la décision améliorant la prise en compte et l'application des recommandations issues des guides de bonnes pratiques (Séroussi et al, 2004)

Le « mode critique » développé par le Lim&Bio est un des composants de ce système. Au cours de la consultation, le médecin saisit les données relatives à son patient (données cliniques, résultats de biologies, efficacité et tolérances du traitement en cours) ainsi que sa nouvelle proposition de traitement dans le dossier médical électronique. Avant l'impression de l'ordonnance, le moteur critique d'ASTI cherche à formuler une critique de la proposition du médecin. Le système émet aussi une proposition de traitement si celle du médecin n'est pas jugée optimale.

Le moteur critique d'ASTI construit cette critique de la prescription du médecin en comparant celle-ci au traitement « optimal » établi par le moteur. Ce traitement « optimal » est obtenu par un moteur qui tient compte des données cliniques du patient, de son historique thérapeutique et des données de la consultation concernant l'efficacité et la tolérance du traitement en cours. Dans ce moteur sont couplés des algorithmes cliniques qui viennent compléter des règles contenues dans une base de connaissances et qui sont extraites du GBP (Ebrahiminia et al, 2006).

Les algorithmes cliniques représentent des stratégies thérapeutiques générales employées dans le raisonnement médical, méta-connaissances développées par le consortium ASTI. Les stratégies sont génériques et donc applicables à toutes les pathologies chroniques. Leur emploi permet aussi de simplifier la base de connaissances en réduisant le nombre de règles qu'elle contient. Ces stratégies comprennent : la stratégie S0 qui est activée si le traitement en cours est efficace et bien toléré ; la stratégie S1 correspondant au cas où le traitement en cours est mal toléré et la stratégie S2 pour un traitement bien toléré mais dont l'efficacité n'est pas optimale. Les règles contenues dans la base de connaissances sont des règles de production de type « si-alors ». Elles sont formulées à partir des recommandations contenues dans le guide de bonnes pratiques et permettent par exemple de lister par ordre de priorité les traitements à proposer pour un certain profil de patient. Elles ont été regroupées en quatre types (R1 à R4), selon leur fonction. Les règles de type R1, par exemple, permettent de produire à partir des valeurs de variables cliniques et biologiques prédéfinies, une liste des traitements appropriés et classés par ordre de priorité. Une description détaillée des règles est dans (Ebrahiminia et al, 2006).

La première application de la démarche de test présentée ici concernait la prise en charge de l'hypertension artérielle non compliquée de l'adulte. Tester le prototype correspondant revenait à priori à répondre à la question « Est-ce que les traitements proposés par le système correspondent aux traitements attendus, au vu des spécifications énoncées, du contenu de la base de connaissances et ce pour toutes les caractéristiques physiques et biologiques des patients, des traitements passés, de l'évaluation de l'efficacité et de la tolérance des divers traitements en cours possibles ? ».

Dans ce qui suit nous allons voir qu'un test exhaustif de ce système avant sa diffusion est primordial mais irréalisable manuellement d'où le souhait d'un test automatisé. Nous commencerons par un état de l'art sur les tests des systèmes

informatisés d'aide à la décision. Nous présenterons ensuite notre méthode pour résoudre notre problématique. Cette méthode repose sur un modèle conceptuel à base d'organigrammes de séquences d'exécution que nous avons construit pour mieux appréhender le fonctionnement du moteur critique, ainsi qu'une table de décision figurant les règles contenues dans la base de connaissances. Nous verrons que ces deux outils nous ferons diviser l'ensemble des cas à tester en classes, les critères de classification nous permettant alors d'automatiser leur test.

2 La méthodologie de test des systèmes d'aide à la décision médicaux

En 2004, De Clercq a entrepris une étude comparative des principales approches pour la conception de systèmes d'aide à la décision informatisée incorporant des guides de bonnes pratiques cliniques (De Clercq et al, 2004). Cette étude s'est intéressée à la vérification et au test de ces approches. Il en est ressortit qu'une seule, *Proforma*, basée sur un langage formel, avait développé des outils dans ce sens. Les autres approches rapportaient seulement des travaux en cours sur le sujet qui n'ont pas été publiés depuis.

Concernant les autres types de systèmes informatisés d'aide à la décision clinique, seules quelques équipes en ont parlé (Shwe et al, 1989 ; Lucas, 1994 ; Sailors et al, 1996). En effet la plupart des auteurs y consacrent habituellement, tout au plus un petit paragraphe au sein de leurs publications. Le test de la plupart de ces systèmes consiste à travailler sur un échantillon de cas cliniques « réels et représentatifs » du domaine concerné et à vérifier si la réponse fournie était bien celle qui était attendue. Cette évaluation se situe en fait, en aval de la phase de test en «laboratoire ». Quand elle a été réalisée, cette évaluation est principalement passée par des études d'impact sur le « terrain » (Roderer, 2004).

Dans les domaines autres que médicaux, les publications sur le test des systèmes d'aide à la décision sont nombreuses. En 1997, Plant identifia 40 outils de test et de vérification des systèmes d'aide à la décision grâce à une revue de la littérature allant des années 1985 à 1995 (Murrell & Plant, 1997). Avant lui, en 1994, Preece réalisa un état de l'art sur la vérification et le test des systèmes à base de connaissances en Amérique du Nord (Preece, 1994). Tous deux insistaient sur l'importance de l'évaluation et du test de ces systèmes avant leur large diffusion. Preece, constatait que les théories sous-tendant les méthodologies et outils de test développés pour les systèmes à base de connaissances étaient principalement issues du domaine de la vérification et validation de l'ingénierie logicielle. La problématique principale commune est qu'on ne peut tester le système sur tous les vecteurs d'entrée, leur nombre étant trop important. Il faut alors choisir un « échantillon » qui soit raisonnable et qui permette le maximum de fiabilité. Une des classifications proposée pour ces méthodes est de les répartir en méthodes statiques et dynamiques.

Les méthodes statiques sont celles qui ne requièrent pas l'exécution du système. La méthode la plus utilisée est l'inspection. Elle consiste à faire vérifier le système par un expert du domaine, à travers la lecture d'un modèle conceptuel du système étudié. D'autres méthodes de « vérifications statiques » ont pour but la détection d'anomalies logiques au sein de la base de connaissances, comme la redondance,

l'ambiguïté, la circularité ou la déficience. Ceci s'applique bien au cas des systèmes dont la base de connaissances est constituée de règles (Preece & al, 1992).

Les méthodes dynamiques sont celles qui requièrent l'exécution du système. Elles permettent de réaliser le test selon deux principales approches. La première consiste à répartir les vecteurs d'entrée afin d'explorer les différentes structures du système : ce sont les « tests structurels ». La deuxième permet d'explorer les différentes fonctions que permet le système en repartitionnant les vecteurs d'entrée en conséquence, les tests sont alors dits : « tests fonctionnels ».

Néanmoins, aucun outil de test n'a connu une large diffusion. Les auteurs attribuent cela au fait que les outils développés demandaient trop d'efforts pour pouvoir être adaptés aux systèmes, et ont ainsi découragé les équipes qui pouvaient être initialement motivées. D'autre part, peu d'études se sont intéressées à évaluer ces outils. Ainsi leur pertinence et fiabilité sont restées inconnues.

3 Notre expérience sur ASTI

A partir de cette analyse, nous avons compris que nous ne disposions pas d'un outil directement applicable au prototype du module critique d'ASTI pour réaliser son test. Nous avons alors imaginé une méthode de test inspirée par les différentes théories précédemment présentées. Nous avons choisi d'élaborer une méthode de test « dynamique » permettant un test plus fiable. Nous en avons alors élaboré un modèle conceptuel en nous basant sur les spécifications papiers fournies au partenaire industriel du projet. Ce modèle conceptuel comprend des organigrammes et une table de décision. Le modèle conceptuel élaboré nous a conduit à proposer une méthode de test inspirée des modes de test de type parallèle (Preece, 1994) en répartissant les vecteurs d'entrée selon le type de réponses obtenues.

Tester le système de façon exhaustive revient à répondre à la question « est-ce que pour chaque patient entré dans la machine, la réponse fournie par le système est bien celle attendue ? ». C'est à dire est-ce que pour chaque vecteur d'entrée, le vecteur de sortie est bien celui attendu ?

Pour chaque patient, il faut saisir les données concernant ses variables cliniques et biologique, son historique thérapeutique et son traitement en cours. Les variables cliniques et biologiques peuvent prendre les valeurs oui/non. Leur nombre est variable selon la pathologie concernée. Pour le traitement en cours il faut préciser la classe thérapeutique, la tolérance et l'efficacité : la tolérance peut être bonne ou mauvaise, l'efficacité bonne, partielle ou nulle. L'historique thérapeutique regroupe l'ensemble de tous les traitements pris par le patient. Pour chaque traitement, il faut préciser l'efficacité et la tolérance (respectivement bonne, partielle ou nulle ; et bonne ou mauvaise). Un traitement peut être composé par une ou plusieurs classes thérapeutiques (association), on parle alors de monothérapie ou bithérapie, trithérapie etc. Une classe thérapeutique représente une famille de molécule ayant un même mécanisme d'action. Pour une pathologie chronique comme l'hypertension artérielle non compliquée de l'adulte, les variables cliniques et biologiques, pouvant prendre les valeurs (oui/non), sont au nombre de 22. Ainsi, Le nombre de vecteurs d'entrée possibles est très important.

Les réponses peuvent être classées en plusieurs catégories : (I) « Pas de réponses », le système n'a pas de recommandation à proposer. (II) Une liste de traitements qu'il considère comme optimale. (III) La modification de la posologie du traitement en cours. (IV) Aucune modification du traitement, ceci correspondant à continuer le même traitement à la même dose. Le système peut aussi proposer une recommandation combinant deux des catégories précédentes : une liste de traitements et modification de la posologie du traitement en cours par exemple.

Nous nous sommes donc rendus compte que le test exhaustif de tous ces cas, bien que souhaitable, est tout à fait impossible à réaliser manuellement. Il faut en effet saisir les vecteurs un par un, puis comparer pour chacun la réponse obtenue à celle attendue. Néanmoins, le nombre de catégories de réponses est limité comparativement au nombre de vecteurs d'entrée. Ainsi, une même catégorie de réponse peut être déclenchée par des vecteurs d'entrée différents. On peut alors diviser l'ensemble des vecteurs d'entrée en famille ou « classes ». Tous les vecteurs d'une même classe déclenchent la même catégorie de réponse. Il faut alors identifier des critères communs aux vecteurs d'entrée d'une même classe. Ces critères nous permettront de générer automatiquement tous les vecteurs de cette classe et de les relier à la réponse attendue. Le test de tous les vecteurs d'une même classe devient donc automatisable.

Nous avons analysé les spécifications papiers du système et nous nous sommes rendu compte que nous n'arrivions pas à obtenir une visualisation globale de l'ensemble du fonctionnement du module. Nous avons alors cherché à créer une modélisation qui nous permette, grâce à une meilleure visualisation de l'ensemble des interactions, d'en avoir une meilleure appréhension. Nous avons choisi de modéliser les différents algorithmes cliniques du module critique sous forme d'organigrammes de séquences d'exécution. La *figure 2* présente, pour l'exemple, l'organigramme des séquences d'exécution déclenchées lors de la stratégie S2 : un traitement bien toléré mais dont l'efficacité n'est pas bonne.

L'analyse comparative de l'ensemble des organigrammes construits montre qu'il existe des séquences communes aux différentes stratégies (A sur *figure 2*). Ainsi, pour les stratégies d'initialisation, S1 et S2, le système déclenche les règles de types R1 du même type que le traitement en cours, pour réaliser ensuite l'intersection entre les traitements proposés. C'est à partir de cette liste que le système va ensuite, en fonction de la stratégie choisie, apporter des modifications comme l'élimination de certains traitements déjà essayés et mal tolérés ou inefficaces (présents dans l'historique thérapeutique). Il existe donc une première phase où l'on enrichit la liste des traitements possibles et une seconde où l'on en élimine. Il nous a alors semblé intéressant d'analyser plus précisément cette première phase d'enrichissement et de vérifier s'il existait des cas où elle aboutit à une liste vide. En effet, rien ne sert alors d'aller plus loin dans la séquence d'exécution, le module ne faisant pas alors de proposition de stratégie thérapeutique. D'autre part, cette première représentation, strictement fidèle aux spécifications papiers, ne nous a pas permis d'expliciter de façon assez détaillée certaines séquences. Celles-ci ont été représentées par des boucles (B *sur figure 2*).

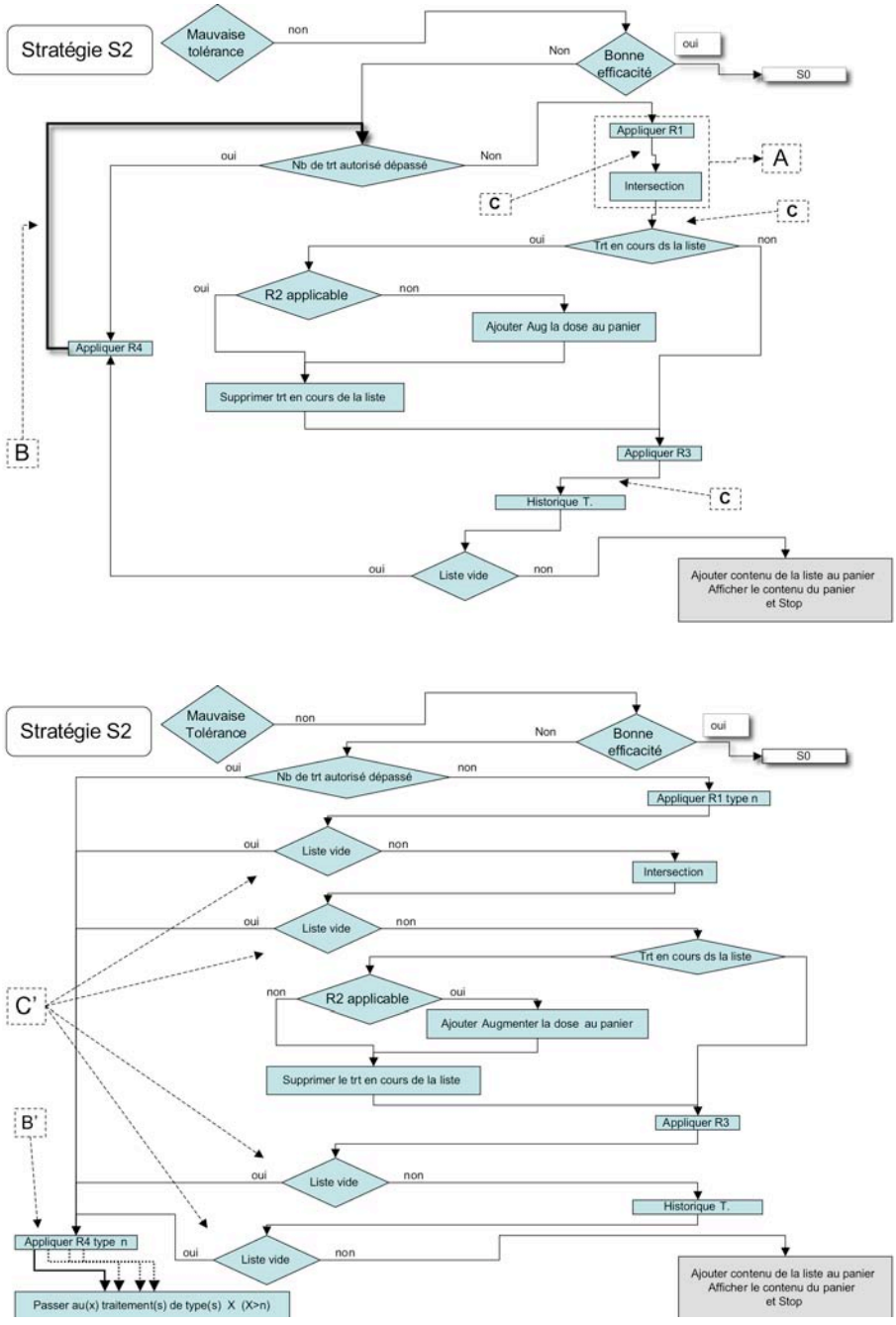

Fig. 2 – Organigramme de séquences d'exécution au cours de la stratégie S2. En haut est présentée la 1ère version et au-dessous la version enrichie et explicitée.

Nous avons alors repris ces organigrammes pour en créer une deuxième version. Dans cette nouvelle version, nous avons enrichi les organigrammes de toutes les ramifications pouvant aboutir à une liste vide (C puis C' sur la *figure 2*). D'autre part, nous avons explicité les séquences contenant des boucles (B sur la *figure 2*), en les déroulant (B' sur *la figure 2*).

Profil N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Variables clin et Bio														
insRénale modérée		N											N	N
InsRénale Sévère		N											N	N
InsCar grade III/IV		N												
Sujet agé			O											
Atcd Acc Vasculaire				O										
M Coro Stable					O									
Post Idm						O								
InsCar							O							
MicroAlbuminurie								**N**	O	O				O
Diabete Type 2								**O**		O			N	N
Diabete Type 1									O				N	N
Diabete 2+ Fdr risq											O			
Hypertro Vent Gche												O		
Proteinurie >=300mg													O	
Hyperkaliémie														
Sténose artère rénale														
Cl. Thérapeutiques														
MHD	O	O O O O O	O O	O	O O	O O	O O O O O	**O O O O**	O O O	O O O	O	O O	O O O	O O O
D_Thiazidique	N	O N N N N	O N	O	N N	N N	N O N N N	**O N N N N**	N N O	N N O	N	O N	N N O	N N O
D_anse	N	N N N N N	N N	N	N N	N N	N N O N N	**N N N N N**	N N N	N N N	N	N N	N N N	N N N
B_Bloqueurs	N	N O N N N	N N	N	O N	O N	N N N O N	**N O N N N**	N N N	N N N	N	N N	N N N	N N N
IEC	N	N N O N N	N N	N	N N	N O	O N N N N	**N N O N N**	O N N	N O N	O	N N	O N N	O N N
ARAII	N	N N N O N	N N	N	N N	N N	N N N N O	**N N N O N**	N O N	O N N	N	N O	N O N	N O N
InhCa++	N	N N N N O	N O	N	N O	N N	N N N N N	**N N N N O**	N N N	N N N	N	N N	N N N	N N N
AntiAldo_f_dose	N	N N N N N	N N	N	N N	N N	N N N N N	**N N N N N**	N N N	N N N	N	N N	N N N	N N N
Rang	1	1 1 1 1 1	1 1	1	1 1	1 1	1 2 2 2 2	**1 1 1 1 1**	1 2 3	1 2 3	1	1 1	1 2 3	1 2 3
Type de Trt	1	2	2	2	2	2	2	2	2	2	2	2	2	2
No règle	1	1	2	3	4	5	6	7	8	9	10	11	12	12

Fig. 3 – Table de décision pour les règles de type R1, traitements de type 1 et 2. La 7ème règle du traitement de type 2 se lit comme suit : En cas de présence de Diabète de type 2 *(Diabète type 2 = O)* et absence de micro-albuminurie *(MicroAlbuminurie = N)*, on peut prescrire en association avec les mesures hygièno-diététiques *(MHD = O*) toutes les classes thérapeutiques hormis les diurétiques de l'anse *(D_Anse = N)* et les anti-aldostérones à faible dose *(AntiAldo_f_dose = N).*

De même que pour les algorithmes cliniques, nous avons analysé l'ensemble des règles énumérées dans les spécifications papiers contenues dans la base de connaissances pour de la prise en charge de l'hypertension artérielle de l'adulte. Pour faciliter la visualisation de ces règles, nous les avons représentées sous forme d'une table de décision (figure 3).

A partir de cette table de décision, nous avons identifié les différents profils de variables cliniques et biologiques et les suggestions de traitements qui leur étaient reliées. En effet, il suffit de regarder si la combinaison de valeurs des variables en question est impliquée dans le déclenchement d'une ou plusieurs règles. En cas de déclenchement de plusieurs règles, le traitement à proposer est le résultat de

l'intersection entre les réponses fournies par ces règles. Sur la table, ceci se traduit par rechercher le ou les profils communs aux règles concernées.

Nous disposions finalement d'une visualisation détaillée des séquences d'exécution mises en jeu par les algorithmes cliniques et des règles contenues dans la base de connaissances du moteur critique de ASTI. Cette visualisation nous permet aussi de suivre la cascade des événements déclenchés par un vecteur d'entrée (représentant un patient) jusqu'au vecteur de sortie associé (la proposition du moteur critique d'ASTI).

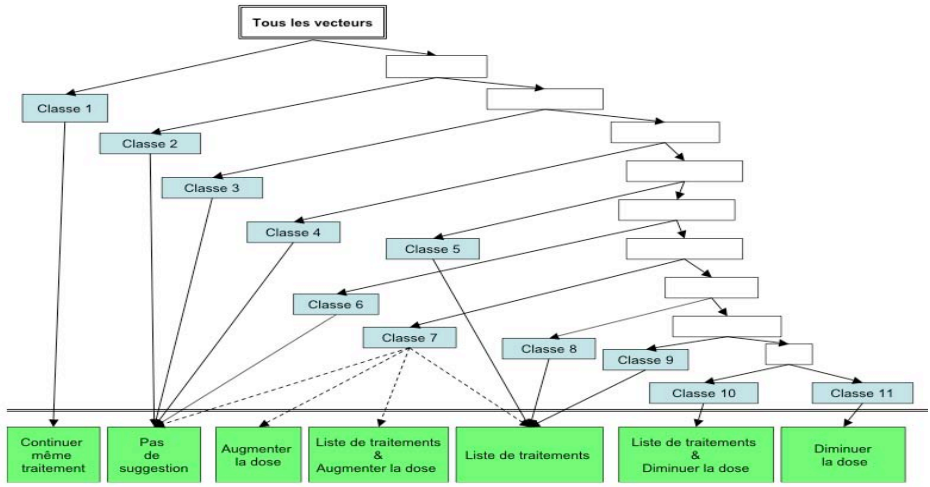

Fig.4 – Division des vecteurs d'entrée en classes déclenchants les mêmes réponses.

L'étude conjointe des organigrammes d'exécution et des vecteurs d'entrée nous a permis d'identifier plusieurs familles de vecteurs en fonction des sorties proposées et d'en hiérarchiser le test. Nous avons fait l'hypothèse qu'un certain nombre de vecteurs d'entrée pouvaient déclencher la même liste de traitement, ou, si l'on se place dans un niveaux de description plus bas, le même *type* de réponse (figure4).

La première classe (classe 1) contient tous les vecteurs d'entrée pour lesquels l'efficacité et la tolérance sont « bonnes » (la variable « Efficacité du traitement en cours » a pour valeur 2 et la variable « Tolérance du traitement en cours » a pour valeur 1). Chaque vecteur de cette classe doit produire la réponse « Continuer le même traitement ». L'automatisation du test de cette classe est donc possible. Il faut alors fixer ces deux variables aux valeurs « 2 » et « 1 », les autres variables du vecteur d'entrée prenant successivement toutes les combinaisons possibles. Pour chaque vecteur ainsi obtenu, il faut récupérer la réponse fournie par le système et vérifier que celle-ci est bien égale à « Continuer le même traitement ». Nous avons décomposé les vecteurs restants en deux sous-classes. La première (classe 2) contient tous les vecteurs d'entrée ne déclenchant aucune règle de type R1 d'aucun type de traitement (ces vecteurs sont ceux dont les valeurs des variables cliniques et biologiques ont été décrits plus haut) et la deuxième les vecteurs restants. Tous les vecteurs de la première doivent déclencher la réponse « pas de suggestion ». En effet,

si aucune règle de type R1 n'est déclenchée alors la liste de traitement possible est vide et doit le rester.

Nous avons ainsi réussi à classer l'ensemble des vecteurs d'entrée en 11 classes : (I) tous les vecteurs dont les valeurs des variables « efficacité du traitement en cours » et « tolérance du traitement en cours » sont bonnes (prenant respectivement les valeurs 2 et 1) ; (II) tous le vecteurs dont les valeurs des variables cliniques et biologiques ne déclenchent aucune règle de type R1 ; (III) tous les vecteurs dont les valeurs des variables cliniques et biologiques déclenchent des règles R1 dont l'intersection est vide ; (IV) tous les vecteurs dont la liste de traitements déclenchée par les règles de type R1 est inclus dans l'historique thérapeutique ; (V) tous les vecteurs dont le traitement en cours est « vide » ; (VI) tous les vecteurs dont la valeur de la variable « type du traitement en cours » est strictement inférieur à tous les types de traitements déclenchés par les règles de type R1 ; (VII) tous les vecteurs dont les valeurs de la variable « traitement en cours » sont différentes de 0 ; (VIII) tous les vecteurs dont les valeurs de la variable « traitement en cours » appartiennent à l'intersection de l'historique thérapeutique et de la liste des traitements déclenchés par les règles de type R1 ; (IX) tous les vecteurs dont l'efficacité du traitement en cours n'est pas bonne (soit égale à 1 ou 0) ; (X) tous les vecteurs dont l'ensemble d'intersection de l'historique thérapeutique et la liste des traitements déclenchés par les règles R1contient au moins un traitement de plus que le traitement en cours ; (XI) tous les vecteurs restants (soit ceux dont l'intersection de l'historique thérapeutique et la liste des traitements déclenchés par les règles R1 ne contient qu'un seul traitement, le traitement en cours). Les critères d'inclusions pour chaque classe peuvent alors servir pour générer automatiquement l'ensemble des vecteurs de cette même classe.

Pour certaines de ces classes (classe 1, 2, 3, 4 et 10) nous avons réussi à définir une même « réponse » en sortie pour tous les vecteurs. Pour d'autres (5, 7, 8 et 9), chaque vecteur produit une réponse (une liste de traitement par exemple) qui lui est propre, néanmoins ces réponses sont du même type pour tous les vecteurs d'une même classe (soit une liste de traitement, soit une liste de traitement et augmentation de dose, etc…). Enfin pour la classe 7 nous n'avons pas, en l'état actuel, encore réussi à définir les critères permettant de subdiviser les vecteurs la composant.

Lors de la première mise en oeuvre de notre méthode, nous avons réussi à mettre en évidence des erreurs qui nous ont conduit à demander une révision du prototype. La nouvelle version -en cours de développement- pourra être testé automatiquement.

4 Discussion et Conclusion

La modélisation des algorithmes cliniques sous formes d'organigrammes de séquences d'exécution, nous a permis de mieux en saisir les interactions mutuelles. En effet, il est plus aisé d'avoir une vision globale en parcourant un organigramme qu'en lisant du texte libre. D'autre part, l'exploration systématique de toutes les ramifications possibles exigée lors de la construction de ces organigrammes nous a aussi permis de détecter des omissions et d'éclaircir certaines ambiguïtés présentes dans le texte original (l'existence des boucles que nous avons décrites et déroulées par exemple ou une condition manquante dans une règle de type R4, etc…). Toujours et

grâce à la nouvelle vision plus globale apportée par cette représentation, la construction de ces organigrammes a été l'occasion de comparer et de rediscuter la validité des enchaînements entre les différentes stratégies représentées (absence de passage par les règles R1 dans la stratégie S0 par exemple).

Nous pensons ainsi que ces organigrammes devraient faire partie et compléter les spécifications papier à transmettre au partenaire industriel pour les prochains prototypes à développer.

La modélisation des règles de type R1 sous la forme d'une table de décision nous a permis d'identifier plus rapidement qu'en parcourant les règles écrites en texte libre les profils cliniques dont nous avions besoin. Il nous semble que cette synthèse est bien adaptée à la visualisation, à la vérification et à l'extraction d'information à partir d'une base de connaissances à nombre réduit de règles (22 règles dans le cas de la base d'hypertension artérielle). Shiffman a même proposé un modèle de table de décision « enrichie » pour la modélisation et la vérification des règles contenues dans les guides de bonnes pratiques cliniques (Shiffman, 1997). Aussi, nous pensons que cette table de décision pourrait être amélioré et enrichi par l'automatisation de certaines tâches, comme la réalisation d'intersection entre les traitements ou l'identification des profils ne déclenchant aucune règle. Il faut toutefois souligner qu'il existe d'autres méthodes et outils pour la vérification statique des bases de connaissances à base de règles. Elles permettent principalement de rechercher la présence d'anomalies logiques comme la circularité, la redondance ou l'ambiguïté parmi les règles constituant ces bases (Preece, 1992 ; Duftschmid et Miksch, 2001).

Pour le test du prototype « actuel » du module critique, nous avons proposé le regroupement dans une même classe des vecteurs d'entrée déclenchant des réponses « proches » et avons ainsi obtenu onze classes dont nous avons hiérarchisé le test. Pour certaines classes, tous les vecteurs déclenchaient une réponse « identique ». Pour d'autres, les vecteurs déclenchaient des réponses différentes mais appartenant à un même type. Enfin pour une dernière, nous n'avons pas réussi à associer « un vecteur » à «la réponse attendue ». Pour faire une correspondance unitaire « un vecteur – la réponse attendue», il faudrait pousser plus loin la subdivision des classes ; cependant cela impliquerait des critères de plus en plus compliqués pour réaliser cette classification et le développement d'un système mettant en œuvre ce test serait alors d'autant plus complexe. Cela reviendrait pratiquement à développer un « nouveau module critique » fonctionnant en « parallèle » à celui-ci. En plus de la complexité de sa mise en œuvre, il faudra aussi prévoir une méthodologie pour valider ce nouveau « système de test ». Il faut alors trouver le juste compromis entre la précision souhaitée pour le test et le coût pour le mettre en œuvre (Preece, 1994).

Nous pensons que l'approche présentée ici reste applicable indépendamment de la pathologie. Cependant, le rapport coût/précision du test est lui dépendant de la pathologie, ou plus précisément de la richesse du guide de bonnes pratiques. Le coût nous semble principalement dépendant de la complexité de la base de connaissances extraite à partir du guide de bonnes pratiques (nombre de règles, etc…). Nous allons évaluer donc cette « généricité » en l'appliquant au test du prototype d'une autre pathologie chronique : le diabète de type 2. Nous projetons à cette occasion de calculer puis de comparer le coût de la mise en oeuvre du test des 2 prototypes.

Références

BUETOW S.A., SIBBALD B., CANTRILL J.A. & HALLIWELL S. (1996). Prevalence of potentially inappropriate long term prescribing in general practice in the United Kingdom, 1980-95: systematic literature review. *BMJ*. 313, p. 1371-4.

COSTE J. & VENOT A. (1999). An epidemiologic approach to drug prescribing quality assessment: a study in primary care practice in France. *Medical Care*. 37, p 1294-307.

DE CLERCQ P.A. & al. (2004). Approches for creating computer-interpretable guidelines that facilitates decision support. *Artificial intelligence in medicine*.31, p. 1-27.

DUFTSCHMID G. & MIKSCH S. (2001). Knowledge-based verification of clinical guidelines by detection of anomalies. *Artificial intelligence in medicine*. 22, p. 23-41.

EBRAHIMINIA V., RIOU C., SEROUSSI B., BOUAUD J., DUBOIS S., FALCOFF H. & VENOT A. (2006). Design of a Decision Support System for Chronic Diseases Coupling Generic Therapeutic Algorithms with Guideline-Based Specific Rules. *Studies in health technology and informatics*. 124, p. 483-8.

GARG A.X., ADHIKARI N.K., MCDONALD H., ROSAS-ARELLANO M.P., DEVEREAUX P.J., BEYENE J., SAM J. & HAYNES R.B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes : a systematic review. *JAMA*. 293, p.1223-38

KAWAMOTO K., HOULIHAN C.A., BALAS E.A. & LOBACH D.F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 330, p.765.

Lucas P. (1994). Refinement of the HEPAR expert system : Tools and techniques. *Artificial intelligence in medicine*. 6, p. 175-88.

MURRELL S. & PLANT R.T. (1997). A Survey of Tools for the Validation and Verification of Knowledge-based Systems, 1985-1995. *Decision Support Systems*. 21, p. 307-23.

PREECE A., BATAREKH A. & SHINGHAL R. (1992). Verifying rule-based systems. *Knowledge Enginnering Review*. 7, p. 115-41.

PREECE A.D. (1994). Validation of knowledge-Based Systems : The state-of-the-Art in North America. *Journal of Communication and Cognition - Artificial Intelligence*. 11, p. 381-413.

PURCELL G.P. (2005). What makes a good clinical decision support system. *BMJ*. 330, p. 740-1

Roderer N.K. (2004). Outcome measures in clinical information systems evaluation. *Medinfo*. 11, p. 1096-100.

ROOD E., BOSMAN R.J., VAN DER SPOEL J.I., TAYLOR P. & ZANDSTRA D.F. (2005). Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation. *Journal of the American Medical Informatics Association : JAMIA*.12, p. 172-80.

SAILORS R.M. & al. (1996). Testing and validation of computerized decision support systems. Proceedings : a conference of the American Medical Informatics Association. p. 234-8.

SEROUSSI B., BOUAUD J., DREAU H., FALCOFF H. & VENOT A. (2004). Modalité d'interaction avec des systèmes d'aide à la decision médicale par alerte ou à la demande pour délivrer des recommandations : une étude préliminaire dans le cadre de la prise en charge de l'hypertension. IC. p. 65-76.

SHIFFMAN R.N. (1997). Representation of clinical practice guidelines in conventional and augmented decision tables. *Journal of the American Medical Informatics Association : JAMIA*. 4, p. 382-93.

SHWE M.A. & al. (1989). Validating the knowledge base of a therapy planning system. *Methods of information in medicine*. 28, p. 36-50.