Patient clinical profiles associated with physician non-compliance despite the use of a guideline-based decision support system: a case study with OncoDoc2 using data mining techniques

Brigitte Séroussi, MD, PhD[1], Arnaud Soulet, PhD[2], Nizar Messai, PhD[2], Cédric Laouénan, MD, PhD[3], France Mentré, MD, PhD[3], Jacques Bouaud, PhD[4]
[1] **UPMC, UFR de Médecine, Paris, France; AP-HP, Hôpital Tenon, Département de Santé Publique, Paris, France; Université Paris 13, Sorbonne Paris Cité, LIM&BIO (EA3969), Bobigny, France; APREC, Paris, France;** [2] **Université François Rabelais de Tours, Laboratoire Informatique (EA 6300), Tours, France;** [3] **Université Paris Diderot, Sorbonne Paris Cité, UMR 738, Paris, France; INSERM, UMR 738, Paris, France; AP-HP, Hôpital Bichat, Service de Biostatistique, Paris, France;** [4] **AP-HP, STIM, Paris, France; INSERM, UMR_S 872, eq. 20, Paris, France**

Abstract

OncoDoc2 is a guideline-based clinical decision support system (CDSS) applied to the management of breast cancer patients. OncoDoc2 has been routinely used during multidisciplinary staff meetings at the Tenon Hospital (Paris, France) for nearly 3 years. Despite the use of the CDSS that reminds physicians of the recommended treatments, the compliance rate of decisions is not 100%. We have used pattern mining techniques in order to elicit patient clinical profiles associated with non-compliance. We quantified each extracted pattern by three measures (support, growth rate, and unexpected rate) and we introduced a score to prune relevant emerging patterns. Non-compliance has concerned elderly patients in pre-surgery decisions, patients with micro invasive tumor in re-excision decisions, and patients HR+ and Her2+ in adjuvant decisions. In all cases, physician non-compliance with guidelines occurs when scientific evidence is lacking.

Introduction

Clinical practice guidelines (CPGs) are recommendations for clinicians about the care of patients with specific conditions. Based on the best available research evidence, they are currently developed by health professional societies and national health agencies to improve the quality of clinical care and decrease health care costs. However, despite the development and dissemination of CPGs, there continues to be variation in the effectiveness of guidelines to change the behavior of clinicians. Barriers to physician adherence to CPGs have indeed been identified[1]. Some barriers are physicians-centered. For instance, physicians may not implement CPGs because they are not aware or not familiar with recommendations, they may also not agree with CPGs contents, *e.g.* they do not believe in the effectiveness of recommendations in terms of clinical outcomes for patients, or they simply have difficulties to break the habit of old practices. Other barriers are "external barriers" and may be either environmental[2], guideline-related (many CPGs are considered to be oversimplified, rigid, and biased, *i.e.* evidence is often missing and expert opinions are the majority), or patient-related.

Simply providing CPGs, in their original format of narrative texts, either as paper-based or electronic documents, has had limited effect in changing physician behavior. Several reviews[3,4] suggest that clinical decision support systems (CDSSs) may be efficient tools to promote the adoption of CPGs by physicians. By providing patient-specific guideline-based recommendations, CDSSs should theoretically answer the question of physician awareness by informing or reminding her, according to her knowledge of CPGs contents, of state of the art decisions. However, by proposing the treatment recommended by CPGs, CDSSs do not solve the problem of physician agreement with CPGs contents. This could explain why reviews of computer-based guideline intervention strategies report mixed conclusions about the actual effectiveness of CDSSs to improve physician compliance with CPGs. Many studies have indeed showed positive effects, but others found only a limited impact of these systems upon physician practices. Delivering patient-specific recommendations at the point of care appears to be "neither necessary nor sufficient" to ensure compliance[5]. Research is thus currently carried out to assess which factors are responsible of the success or the failure of CDSSs. Beyond variations in clinical setting, culture, training, and organisation, research is mainly being conducted to

analyze the CDSSs used in order to elicit the technical features, *e.g.* design, implementation, and level of description, that would predict their effectiveness to increase clinician compliance with CPGs. Some authors of this article already studied the patient effect on non-compliance with the ASTI system[6] concluding that for "more complex" cases, general practitioners (GPs) accept to be helped and on-demand guidance-based systems are recommended, whereas for "simple" patient cases, GPs think they do not need to be helped, and alert-based CDSSs are both efficient and mandatory since GPs would not spontaneously seek for information. However, few studies attempt to assess the impact of patient clinical profiles on physician non-compliance with CPGs.

According to a long term political action, known as "Cancer Plan", initiated in France in 2003, therapeutic decisions concerning cancer patients should now be made by multidisciplinary staff meetings (MSMs) and implement CPGs. We have developed OncoDoc2[7], a guideline-based CDSS providing patient-specific recommendations in the management of non-metastatic invasive breast cancer according to local guidelines (CancerEst). OncoDoc2 has been evaluated in a before/after study[8] where the compliance with CancerEst CPGs of the breast cancer MSMs decisions of the Tenon Hospital, Paris, France, was improved. Following this study, OncoDoc2 has been routinely used for nearly 3 years with a compliance rate of 91.7%. The objective of this work is to take advantage of the resulting sample of non-compliant decisions, made by the same physicians, using the same CDSS, embedding the same CPGs, to study patient clinical profiles associated with non-compliance.

Knowledge Discovery in Databases (KDD) is an interdisciplinary field that cuts across the areas of databases, statistics, and artificial intelligence. The goal is to automatically discover information that may be generalized as new knowledge by experts. Data mining techniques have been applied with success in the analysis of guideline compliance[9,10]. Introduced by Agrawal and Srikant[11], pattern mining is an important tool for KDD and has been used in a wide range of applications and domains such as bioinformatics[12], or chemoinformatics[13]. The aim is to extract all the multi-criteria regularities, or patterns, satisfying some constraints that specify their relevance. Emerging patterns[14] (EPs) are patterns which frequency is significantly different between two datasets (*i.e.*, two classes).

Guidelines are suggestions for care, not rules. There are always individual patients who should be managed differently. In this paper, we have used pattern mining techniques in order to elicit patient clinical profiles considered as EPs of non-compliant breast cancer decisions. The objective is to characterize who are the patients for which physicians decide not to follow CPGs despite the use of the CDSS OncoDoc2 that reminds them of the recommended treatments, *i.e.* what are the patient criteria which make physicians deviate from CPGs.

Material

The CDSS OncoDoc2.   OncoDoc2[7] is a guideline-based CDSS applied to the management of breast cancer patients. The system relies on a formalized knowledge base (KB) structured as a decision tree. OncoDoc2 has been developed according to the documentary paradigm of decision support which allows for contextual interpretation of patient data and guidelines knowledge. Although the system can be automatically run from patient data recorded in an electronic health record, the KB is preferably browsed by the physician user. At each depth level, a question is displayed in a closed-ended form to document a clinical criterion that may either concern patient information (*e.g.* menopausal status, general condition, contraindication to surgery, etc.), therapeutic history (*e.g.* prior neoadjuvant treatment), or tumor properties (*e.g.* presence of microinvasion, HER2 or hormone receptors status, etc.). Starting from the root of the decision tree, the physician user navigates through the KB while answering questions and thus instantiating patient criteria. Data are collected from the answers given as a conjunction of bindings $< criterion = value >^*$. This conjunction of bindings corresponds to a specific path of the decision tree and represents the best "formal patient" that matches the actual patient.

Guideline-based patient-centered therapeutic recommendations are provided when the navigation is completed, *i.e.* a leaf is reached (see figure 1). When the actual decision is chosen among the system's propositions, it complies by construction with CPGs. When it is different, the physician decision is considered to be non-compliant with CPGs.

Over the last years, regular MSMs have become a standard practice in oncology. They are the place where therapeutic decisions are collectively made by physicians representing the different medical specialties involved in the management of breast cancer patients (surgeons, pathologists, radiologists, oncologists, radiotherapists, geneticists, etc.). At

Figure 1. Formalized patient profile and corresponding guideline-based recommendation with the record of a non-compliant decision.

the Tenon hospital, more than thirty doctors gather once a week for breast cancer MSMs where they discuss together the best treatment for an average number of forty patients per week. First used in a before/after study[8] where it showed it improved the compliance rate of MSM decisions with CancerEst CPGs (from 79 to 93%), OncoDoc2 has then been routinely used as an element of the MSM decision process. For each breast cancer patient, the navigation was performed while the clinical case was orally presented by the physician in charge of the patient. Since OncoDoc2 display was video projected on a large screen, all physicians could check the instantiation of patient criteria and the correct execution of the navigation, as well as read the guideline-based recommendations provided by the system. These recommendations were considered in the discussion between physicians before the decision was actually made, whether it was compliant or not with one of the system's propositions.

MSM decisions datasets. The unit of analysis is made of the path selected by the navigation (*i.e.* the set of patient criteria), the MSM therapeutic decision, as well as its compliance status. However, medical decisions for breast cancer patients differ according to the step of the patient journey. Before surgery, decisions are most of the time surgical, and criteria that need to be documented essentially concern the general condition of the patient (could she undergo general

anesthesia?), as well as the size of the tumor and the existence of axillary lymph nodes. After surgery, decisions are most of the time medical, often made of chemotherapy, radiotherapy, and hormone therapy. Criteria necessary at this time describe physical and biological characteristics of the tumor. It happens that sometimes, the surgery is not carcinological (*e.g.* margins of resection are positive) and a re-excision is needed. Because the KB is structured as a decision tree, the questions asked downstream relies on the answers given upstream. As a consequence, all criteria are not systematically instantiated for a given patient and decisions are not made of the same criteria. Some may look as missing although it is not because values were not available to clinicians, but because they were not necessary/relevant for the decision and the question about their value is not asked. For instance, there is no question about the grade SBR before surgery and no question about contraindication to surgery after surgery. Thus, we have considered 3 groups of patients and 3 kinds of decisions for which the set of criteria are more homogeneous and medically relevant. The 3 studied groups correspond to: (i) *pre-surgery* decisions (mainly initial decisions), (ii) *re-excision* decisions (when prior surgery is not satisfactory), and (iii) *adjuvant* decisions (once the tumor has been surgically removed). Decisions related to "in situ" breast cancers were not considered in the study since they mobilize different decisional criteria.

Method

Pattern mining method. Let $\mathcal{D}$ be a dataset (Table 1), which is an excerpt of the data used for characterizing the non-compliance of medical decisions.

Table 1. An example of transactional dataset

<div align="center">

$\mathcal{D}$

Trans.	Items				
d_1	A	B	C		
d_2	A	B	C		
d_3	A	B	C	D	$\mathcal{D}_{no}$
d_4	A	B	C	D	
d_5	A	B	C	D	
d_6	A				
d_7		B	C		$\mathcal{D}_{yes}$
d_8			C	D	

</div>

Each line of Table 1 is a *transaction*. A transaction represents a patient profile described by features, or *items*, *e.g.* A, B, C, D, denoting the clinical criteria for which the decision has been made. $\mathcal{D}$ is partitioned into two datasets $\mathcal{D}_{no}$ (non-compliant decisions) and $\mathcal{D}_{yes}$ (compliant decisions). A pattern is a set of items (*e.g.* $\{A, B, C\}$ noted by the string ABC). A transaction t contains the pattern X if and only if $X \subseteq t$. The *support* of pattern X, denoted by $Supp(X, \mathcal{D})$, is the proportion of transactions in $\mathcal{D}$ containing X (*e.g.* $Supp(AB, \mathcal{D}) = 5/8$ and $Supp(AB, \mathcal{D}_{no}) = 4/4$).

Pattern mining aims at discovering all the patterns satisfying a given predicate q, named *constraint*, and occurring in the transactional context $\mathcal{D}$. The constraint enables to only select interesting patterns according to the task of finding regularities or contrasts. For instance, the minimal support constraint[11] focuses on patterns having a support exceeding a given minimal threshold $Supp(X, \mathcal{D}) \geq \gamma$ with $\gamma > 0$.

The contrast brought by a pattern between classes is measured by its *growth rate* (GR). The growth rate of a pattern X from $\mathcal{D}_{no}$ to $\mathcal{D}_{yes}$ is defined as:

$$GR_{no}(X, \mathcal{D}) = \begin{cases} 0 & \text{if } Supp(X, \mathcal{D}_{no}) = 0 \text{ and } Supp(X, \mathcal{D}_{yes}) = 0 \\ +\infty & \text{if } Supp(X, \mathcal{D}_{no}) > 0 \text{ and } Supp(X, \mathcal{D}_{yes}) = 0 \\ \frac{Supp(X, \mathcal{D}_{no})}{Supp(X, \mathcal{D}_{yes})} & otherwise \end{cases}$$

Intuitively, an EP is a pattern which frequency is significantly larger in one class[14]. In practice, a pattern is said to be

an EP if its growth rate exceeds a given threshold α. Typically, AB is an EP with $\alpha = 2$ because its growth rate equals to $Supp(AB, \mathcal{D}_{no})/Supp(AB, \mathcal{D}_{yes}) = 1/0.25 = 4 > 2$. In the same way, AC is also an EP with $\alpha = 2$ because its growth rate equals to $Supp(AC, \mathcal{D}_{no})/Supp(AC, \mathcal{D}_{yes}) = 1/0.25 = 4 > 2$. Thanks to their capacity to emphasize differences between classes, EPs have been used for diagnosis [12,13].

Unfortunately pattern mining often leads to an overwhelming number of patterns. For coping with this problem, we propose to filter out expected patterns. We approximate the growth rate of a pattern from its generalizations by jointly considering several generalizations forming a partition. A partition of X is a set of patterns $p = \{X_1, \ldots, X_l\}$ such that the union of X_i equals to X and there is no intersection between any two distinct patterns of p. The set of all partitions of X is denoted by $P(X)$. For instance, $P(ABC) = \{\{A, BC\}, \{AB, C\}, \{AC, B\}, \{A, B, C\}\}$. The *unexpected rate* of X (UR) is defined as:

$$UR(X, \mathcal{D}) = \min_{\{X_1, \ldots, X_n\} \in P(X)} \frac{GR_{no}(X, \mathcal{D})}{\prod \max\{1; GR_{no}(X_i, \mathcal{D})\}}$$

The unexpected rate measures the deviation between the growth rate and the estimated growth rate based on the hypothesis of independence between the sets of any partition. Note that the growth rate of each generalization at the denominator is at least 1 not to favor patterns having low discriminative generalizations. Let us come back to Table 1 and consider the two patterns AB and AC. These patterns have exactly the same support and the same growth rate: $Supp(AB, \mathcal{D}) = Supp(AC, \mathcal{D}) = 5/8$ and $GR_{no}(AB, \mathcal{D}) = GR_{no}(AC, \mathcal{D}) = 4$. Assuming that A, B and C are independent events, we can estimate the growh rate of AB by $\frac{Supp(A, \mathcal{D}_{no}) \times Supp(B, \mathcal{D}_{no})}{Supp(A, \mathcal{D}_{yes}) \times Supp(B, \mathcal{D}_{yes})} = GR_{no}(A, \mathcal{D}) \times GR_{no}(B, \mathcal{D}) = 2 \times 2 = 4$. Similarly, we can estimate the growth rate of AC by $\frac{Supp(A, \mathcal{D}_{no}) \times Supp(C, \mathcal{D}_{no})}{Supp(A, \mathcal{D}_{yes}) \times Supp(C, \mathcal{D}_{yes})} = GR_{no}(A, \mathcal{D}) \times GR_{no}(C, \mathcal{D}) = 2 \times 4/3 = 8/3$. Since $UR(AB, \mathcal{D}) = \frac{4}{4} = 1$ and $UR(AC, \mathcal{D}) = \frac{4}{8/3} = 1.5$, AC is more unexpected than AB, thus more relevant: AC's growth rate deviates more from the estimated growth rate induced by its generalizations (here, A and C). A pattern is said to be an *unexpected* EP if its unexpected rate is greater than a given threshold ρ. In our example, AC is an unexpected EP with $\rho > 1$ whereas AB is not unexpected at the same threshold ρ.

Choice of thresholds. In our application, transactions correspond to patient profiles. Items are the clinical criteria instantiated while using OncoDoc2. Decisions are made by MSMs and may be either compliant or non-compliant with CPGs. The discovery of non-compliance patterns consists in collecting all the patterns that simultaneously satisfy a minimal support γ in $\mathcal{D}$, a minimal growth rate α from $\mathcal{D}_{no}$ and a minimal unexpected rate ρ. This process is repeated for each of the 3 groups of decisions, *pre-surgery* decisions, *re-excision* decisions and *adjuvant* decisions. The thresholds chosen are $\alpha = 4$ and $\rho = 2$ for focusing on the most important and unexpected contrasts. The threshold γ depends on the size of the group. Basically, a pattern X mined with our approach represents a set of clinical criteria, present at least in a given number of decisions in the group, which characterizes non-compliant decisions at least 4 times more than compliant decisions, and this characterization is at least 2 times greater than what was expected.

Selection of relevant emerging patterns. The selection of extracted patterns is made of two steps: aggregation and ranking. The aggregation of patterns consists in grouping patterns that describe the same set of patients, *i.e.* patterns that belong to the same equivalence class [15]. Two equivalent patterns share the same support and the same growth rate. However, two patterns may have the same support and the same growth rate without belonging to the same equivalent class. Each equivalence class of patterns is represented by its most unexpected pattern (the pattern maximizing the UR measure). A score has been assigned to each pattern representative to quantify its overall utility [16] (*i.e.* a trade-off between $Supp$, GR_{no} and UR): $score(X, \mathcal{D}) = Supp(X, \mathcal{D}) \times GR_{no}(X, \mathcal{D}) \times UR(X, \mathcal{D})$. Pattern representatives of the 3 groups of decisions have been gathered. Scores have been sorted in descending order from a (max value) to b (min value). We decided to consider as significant for the analysis only the pattern representatives which score is greater than $(a + b)/2$.

Results

Raw results and distributions by group. Decision data have been collected for 29 months, from February 2007 and September 2009, issued from the routine use of OncoDoc2 during MSMs. A total of 1,886 exploitable decisions was obtained including 1,624 for invasive breast cancers. The global non-compliance rate has been measured at 8.3% with 135 non-compliant decisions. Table 2 reports the distribution of the number of decisions as well as the non-compliance rate for each of the 3 groups.

While the *pre-surgery* and *adjuvant* groups are similar in size, representing respectively 42.6% and 45.2% of the considered sample of invasive cancers decisions, the *re-excision* group is smaller and represents 12.2% of the dataset, but has the worst non-compliance rate. As in hematology (for defining rare erythrocyte phenotypes), we consider that a patient profile is rare as soon as it is less than $4/1000$. We apply this rule to define the minimal support for *pre-surgery* and *adjuvant* decisions as $1,624/250 = 6.496 \approx 7$ patients. However, in the *re-excision* group, since the size of the group is around 3 times smaller, we decided that the minimal support should be smaller, consider that a pattern should cover 3 times fewer patients to be significant and define the minimal support as $6.496/3 = 2.165 \approx 3$. Finally, the threshold γ is $7/691 = 0.010$, $3/198 = 0.015$, and $7/734 = 0.009$ for respectively pre-surgery, re-excision, and adjuvant decisions.

The pattern mining algorithm has been applied to the 3 groups of decisions returning 822,587, 1,131,881, and 3,767,283 unfiltered patterns, reduced to 14, 421, and 78 EPs for pre-surgery, re-excision, and adjuvant groups, respectively, and less EP classes (see table 2). When mixing all 513 EPs, scores vary from 3.5401 to 0.1327. Thus we only considered as relevant patterns, EPs which score is greater than 1.8364.

Table 2. Distribution of decisions in the pre-surgery, re-excision, and adjuvant groups with their non-compliance rate, number of extracted patterns, minimal support, numbers of EPs and EP classes, as well as their score range.

Group	(n)	Non-compliance	Extracted patterns	min. Supp	EPs	EP Classes	Score [max-min]
Pre-surgery	692	5.8%	822,587	0.010	14	11	[3.08-0.13]
Re-excision	198	14.1%	1,131,881	0.015	421	140	[2.82-0.20]
Adjuvant	734	9.1%	3,767,283	0.009	78	34	[3.54-0.17]
Total	1,624	8.3%	5,721,751	–	513	185	[3.54-0.13]

Emerging patterns in the pre-surgery group. The 14 EPs correspond to 11 equivalent classes. Scores range from 3.08 to 0.13. Table 3 reports the 4 relevant classes (scores higher than 1.8364). Each class is made of only one representative, each representative is composed of only one criterion. While P1, P2, and P3 have a high *GR/UR* and a weak support, P4 has a lower *GR/UR* but a larger support. The other classes [P5-P11] have EPs with multiple criteria (max. 3) and lower support, *GR* and *UR*.

Table 3. Relevant EPs characterizing non-compliant decisions in the pre-surgery group.

Class	# EPs	Most representative EPs	Supp	GR	UR	Score
P1	1	Bad_Condition=Yes	0.012	16.30	16.30	3.08
P2	1	Contra_Ind_Surg=Yes	0.012	16.30	16.30	3.08
P3	1	Receptors=HR+_HER2-	0.013	13.00	13.00	2.20
P4	1	Older_80=Yes	0.095	4.79	4.79	2.19

Emerging patterns in the re-excision group. In the re-excision group, the 421 EPs are grouped into 140 equivalent classes of EPs. Scores range from 2.82 to 0.20. Table 4 reports the 6 classes with higher scores (> 1.8364). For each

class, only the most unexpected patterns (the patterns maximizing the *UR* measure) are listed. All these classes have multi-criteria EPs, from 2 (R6) to 4 (R4). It should be noticed that the R5 class contains 20 EPs.

Table 4. Relevant EPs characterizing non-compliant decisions in the re-excision group.

Class	# EPs	Most representative EPs	Supp	GR	UR	Score
R1	4	Insitu_Microinv=Yes Menopause=Yes Sentinel_Node_Dissection=Yes	0.025	24.30	4.58	2.82
		Insitu_Microinv=Yes Menopause=Yes Axillary_Exploration=Yes	0.025	24.30	4.58	2.82
		Microinv=Yes Menopause=Yes Sentinel_Node_Dissection=Yes	0.025	24.30	4.58	2.82
		Microinv=Yes Menopause=Yes Axillary_Exploration=Yes	0.025	24.30	4.58	2.82
R2	2	Inv_Insitu=No Sentinel_Node_Dissection=Yes Inv_Margins=Pos	0.020	18.20	7.50	2.76
R3	3	Age=35-80 Insitu_Margins=Neg Neoadjuvant_Treatment=Chemo	0.025	24.30	4.00	2.46
R4	2	Age=35-80 Insitu_Margins=Neg Inv_Insitu=No Re-Excision_Mast=Yes	0.025	24.30	4.00	2.46
R5	20	Multiple_or_None=No Sentinel_Node=Neg Complete_Mam_Surg=No Insitu_Margins=Neg	0.020	18.20	6.00	2.21
		Multiple=No Sentinel_Node=Neg Complete_Mam_Surg=No Insitu_Margins=Neg	0.020	18.20	6.00	2.21
		Unique_Inv=Yes Sentinel_Node=Neg Complete_Mam_Surg=No Insitu_Margins=Neg	0.020	18.20	6.00	2.21
R6	1	Inv_Size=Greater_2cm Receptors=HR+_HER2-	0.015	12.10	10.50	1.93

Emerging patterns in the adjuvant group. The 78 EPs correspond to 34 equivalent classes. Scores range from 3.54 to 0.17. Table 5 reports the 5 classes with scores higher than 1.8364. It should be noticed that 4 classes contain only 1 EP while A3 contains 6 EPs, each one being made out of 5 criteria ; the other classes are represented by EPs with 1 or 2 criteria. The *GR* of A3 is very high (59.70), but its support is small whereas the *GR* of A1 is small, but its support is high.

Table 5. Relevant EPs characterizing non-compliant decisions in the adjuvant group.

Class	# EPs	Most representative EPs	Supp	GR	UR	Score
A1	1	Receptors=HR+_HER2+	0.068	7.21	7.21	3.54
A2	1	Complete_Mam_Surg=No	0.016	13.90	13.90	3.15
A3	6	Re-Excision_Mast=No Menopause=No Inv_Insitu=Yes Node=N- Contra_Ind_Antracyclines=No	0.010	59.70	5.00	2.85
A4	1	Receptors=HR+_HER2+ Node=1-3N+	0.020	39.80	3.33	2.70
A5	1	Inv_Margins=Pos	0.050	6.79	6.79	2.32

Discussion

We have used pattern mining methods to elicit patient clinical criteria associated with non-compliant physician decisions despite the use of the guideline-based CDSS OncoDoc2. When analyzing EPs, conclusions are different according to the 3 groups of decisions which confirms that it was reasonable to distinguish them. In the pre-surgery group, EPs are mono-criterion patterns and concern elderly patients (Older_80=Yes), in bad general condition (Bad_Condition =Yes), in whom surgery is contraindicated (Contra_Ind_Surg=Yes), with positive hormone receptors and negative HER2 status (Receptors=HR+_HER2-). These patterns described the clinical profiles of old patients with no major risk factor. Physician non-compliance with CPGs may be explained by the fact they do not agree with CPGs contents in these very specific cases. Indeed, they may not believe in the effectiveness in terms of clinical outcomes

of recommendations elaborated for patients aged 35-75 and in average general condition for elderly patients in bad general condition. As a consequence, non-compliant decisions involve either overtreatments to avoid a second surgery (*i.e.* mastectomy and axillary lymph node dissection instead of lumpectomy and sentinel lymph node dissection) or compassionate undertreatments (*i.e.* surgery is withdrawn in favor of a sole hormone therapy).

In the re-excision group, patient profiles described by R1 are related to the surgical management of micro invasive tumors, whether they are associated with in situ tumors (Insitu_Microinv=Yes) or with invasive tumors (Microinv=Yes). CancerEst CPGs implemented in OncoDoc2's KB consider that micro invasive tumors should be managed as invasive tumors whereas national CPGs consider they should be managed as in situ tumors although there is no evidence to support any of the two options. Thus, non-compliant decisions concern specific cases where physicians do not decide the axillary lymph node removal after an initial sentinel lymph node dissection because they may consider that re-excision is not the best choice in terms of patient clinical outcomes. The case of R6 is similar since no adjuvant chemotherapy has been decided despite the negative status of hormonal receptors assessed on micro-invasive *foci* (although adjuvant chemotherapy is recommended for invasive *tumors* with negative hormonal receptors status). In the same group, the other patterns concern patients with incomplete breast surgery either because of positive invasive margins (Inv_Margins=Pos in R2 or Complete_Mam_Surg=No with Insitu_Margins=Neg in R5) or because mastectomy is indicated and not decided (Re-Excision_Mast=Yes in R4 or patients of R3 in whom mastectomy is recommended after neoadjuvant chemotherapy). In the case of positive invasive margins, re-excision although recommended is not decided because physicians considered margins of resection that should be greater than 3 mm according to CPGs were clear at only 2 mm. In R3 and R4, patients refused the mastectomy recommended. Patient preferences represent a small part of non-compliant decisions (11%) and hold only because they are accepted by physicians. In these latter cases, current margins of resection after lumpectomy were negative but since tumor size > 4 cm, mastectomy is recommended. In all cases, by deciding not to go for a second surgery, physicians show they do not believe it would be a benefit for the patient.

In the third group, the most relevant EP concerns the decision of hormone therapy for patient with HR+ and HER2+ receptors status (Receptors=HR+_HER2+ in A1 and A4). Non-compliance reveals again the lack of evidence to chose between "tamoxifen", "tamoxifen with agonists", "aromatase inhibitors", and "aromatase inhibitors with agonists" although CancerEst CPG recommendations are unambiguous on the subject. A2 and A5 patterns concern patient profiles with incomplete surgery (Complete_Mam_Surg=No in A2 and Inv_Margins=Pos in A5) in whom there has been no decision of re-excision. For A3 patterns, re-excision is also recommended but after adjuvant chemotherapy. Non-compliance concerns the inversion of the recommended sequence chemotherapy-surgery and the decision of surgery-chemotherapy for non-severe patients (*i.e.* no lymph node invasion, Node=N-). In this case, the reason of non-compliance might be organisational.

From the KDD point of view, the number of extracted EPs is very low as compared to similar experiments performed in other fields [12,13]. In particular, we notice that there is no pattern having a lot of criteria (> 5) despite we do not use a size limit. This phenomenon is neither due to the dataset size nor to the thresholds, but results from the unexpected rate that filters out redundancies. Only few patterns describe the pre-surgery group because the number of non-compliant decisions is low (only 5.8% of decisions). More contrasts could be obtained by decreasing the minimal support threshold, but in return, the extracted patterns would be less significant. We also notice that the patterns characterizing the non-compliant decisions in the pre-surgery and adjuvant groups are shorter than those in the re-excision group. This is not a bias resulting from the choice of thresholds since the minimal growth rate and the minimal unexpected rate are similar in the three groups of decisions, and the minimal support threshold is even higher. Indeed, the increase of the minimal support threshold often should lead to mine more general patterns which are more frequent and smaller [11]. Thus we can formulate assumptions. Non-compliant decisions in the re-excision group correspond to profiles of patients less homogeneous and harder to characterize. For this reason, there is no criterion that is individually sufficient to explain non-compliance as it is the case, for instance, in the pre-surgery group. Moreover, we observe that it is probably more difficult to characterize the re-excision group as the score is in average lower in this group.

Another observation is that the supports of relevant EPs are often low, corresponding to few decisions, so that even if the *GR* is high and corresponds to actual observations, generalization might be hazardous. The explanation could be

threefold: (i) the low size of our data set as compared to the number of paths of OncoDoc2's KB and the spreading of patient profiles among these paths (more than half of patient profiles are unique), (ii) the fact that not all criteria are assigned for every decision leading to raw low supports for criteria, and (iii) the observed non-compliance rate is basically low (8.3%).

Although effective in producing medically relevant EPs, the method applied has some limitations. The first one is to assess how successful the method is since there is no "gold standard" in the knowledge discovery paradigm. The difficulty is indeed to prune extracted patterns and filter relevant EPs. Three measures have been used: $Supp$, GR, and UR. Other descriptive measures could have been considered as an alternative to GR like Jaccard, Cosine, or Information gain [16]. But we chose GR since the growth rate of a pattern has no relationship with the count of the records that do not contain this pattern and it is constant if there is no counter example to the EP. Besides, if the unexpected rate measure guarantees a minimal robustness of mined patterns, some relevant patterns are inevitably missed due to the thresholds that are chosen with a certain degree of subjectivity. With higher thresholds, we could have obtained less EPs, but in this case, only a small part of non-compliance would have been recognized as associated with EPs. For instance, in the re-excision group, increasing the minimal support would have hidden a lot of EPs. That is the reason why we chose to use low thresholds for these measures to keep as many potentially relevant EPs as possible, while obtaining many EPs that represent nearly one third of the dataset size. Thus we had to introduce a new measure to be used as a unique indicator of interest of EPs, in order to analyse the "most relevant" EPs. However, this appears as a trade-off, since GR is the most significant indicator to measure the strength of the link to non-compliance, while UR measures the "expectedness" of EPs. As a result, giving less importance to UR in the scoring function is an option to be investigated.

Conclusion

Despite the routine use of the guideline-based CDSS OncoDoc2, the compliance rate of MSMs decisions for breast cancer patients is not 100%. Guidelines are indeed suggestions for care, not rules. There is always individual patients who should be managed differently for legitimate reasons. However, most patients do fit guidelines, and this should be reflected in practice. Thus, we have applied a KDD method to elicit patient-profiles-related EPs associated with non-compliance. Although they were using OncoDoc2, MSMs' physicians decided not to comply with local CPGs when they considered recommendations were lacking evidence, they had a doubt on patient outcomes and they chose to optimize the way clinical departments operate (complete surgery even if chemotherapy is recommended first). Non-compliant decisions are then driven by relevant exceptions rather than by "rules". More data should be collected to make the most of KDD techniques.

Acknowledgments

References

1. Cabana MD, Rand CS, Powe NR, *et al.* Why don't physicians follow clinical practice guidelines? a framework for improvement. *JAMA* 1999;282(15):1458–65.

2. Waitman LR, Miller RA. Pragmatics of implementing guidelines on the front lines. *J Am Med Inform Assoc* 2004;11(5):436–8.

3. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes. *JAMA* 1998;280:1339–46.

4. Garg AX, Adhikari NKJ, McDonald H, *et al.* Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293(10):1223–38.

5. Shiffman RN, Liaw Y, Brandt CA, Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. *JAMIA* 1999;6(2):104–14.

6. Séroussi B, Bouaud J. Reminder-based or on-demand guideline-based decision support systems: a preliminary study in primary care with the management of hypertension. In: Kaiser K, Miksch S, Tu S, eds, Computer-based Support for Clinical Guidelines and Protocols.Proc Symposium on Computerized Guidelines and Protocols (CGP 2004), (vol101) of *Studies in Health Technology and Informatics*, Amsterdam. IOS Press, 2004:142–6.

7. Séroussi B, Bouaud J, Antoine ÉC. OncoDoc, a successful experiment of computer-supported guideline development and implementation in the treatment of breast cancer. *Artif Intell Med* 2001;22(1):43–64.

8. Séroussi B, Bouaud J, Gligorov J, Uzan S. Supporting multidisciplinary staff meetings for guideline-based breast cancer management: a study with OncoDoc2. In: Proc AMIA 2007, Chicago, IL. AMIA, November 2007:656–60.

9. Razavi AR, Gill H, Åhlfeldt H, Shahsavar N. A data mining approach to analyze non-compliance with a guideline for the treatment of breast cancer. In: Kuhn KA, Warren JR, Leong TY, eds, MedInfo, (vol129) of *Studies in Health Technology and Informatics*. IOS Press, 2007:591–5.

10. Svátek V, Ríha A, Peleska J, Rauch J. Analysis of guideline compliance–a data mining approach. *Stud Health Technol Inform* 2004;101:157–61.

11. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. In: Bocca JB, Jarke M, Zaniolo C, eds, VLDB. Morgan Kaufmann, 1994:487–99.

12. Li J, Wong L. Identifying good diagnostic gene groups from gene expression profiles using the concept of emerging patterns. *Bioinformatics* 2002;18(5):725–34.

13. Lozano S, Poezevara G, Halm-Lemeille MP, *et al.* Introduction of jumping fragments in combination with QSARs for the assessment of classification in ecotoxicology. *J. Chem. Inf. Model.* 2010;50(8):1330–9.

14. Dong G, Li J. Efficient mining of emerging patterns: Discovering trends and differences. In: KDD, 1999:43–52.

15. Pasquier N, Bastide Y, Taouil R, Lakhal L. Discovering frequent closed itemsets for association rules. In: Beeri C, Buneman P, eds, ICDT, (vol1540) of *Lecture Notes in Computer Science*. Springer, 1999:398–416.

16. Geng L, Hamilton HJ. Interestingness measures for data mining: A survey. *ACM Comput. Surv.* 2006;38(3).