UNIVERSITA' DEGLI STUDI DI PAVIA

FACOLTA' DI INGEGNERIA
DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

DOTTORATO DI RICERCA IN BIOINGEGNERIA E BIOINFORMATICA
XXVIII CICLO - 2015

PERSONALIZATION OF CLINICAL DECISION SUPPORT SYSTEMS IN THE CONTEXT OF SHARED DECISION MAKING

PhD Dissertation by

ENEA PARIMBELLI

Advisor:
Prof. Silvana Quaglini

PhD Program Chair:
Prof. Riccardo Bellazzi

Grandfather
advised me:
Learn a trade

I learned
to sit at desk
and condense

No layoff
from this
condensery

Acknowledgments

I owe much gratitude to my supervisor Professor Silvana Quaglini, her inspiration and guidance have been invaluable during these years as a graduate student. I would also like to thank Professor Riccardo Bellazzi for his precious advices and support to my research work.

I gratefully acknowledge all the people at BMI lab "Mario Stefanelli", who I have enjoyed collaborating and working with in these years. In particular I would like to thank the MobiGuide project partners for the good work carried out together.

A special mention goes to prof. John Holmes for giving me the opportunity of being hosted at UPenn and for his support in my research activity, and to prof. Mor Peleg for the professional and personal insights during the MobiGuide project activities and beyond.

Finally I would like to express my utmost gratitude to my wife Francesca for being the best part of a life that couldn't be better, and to all my friends and family for their invaluable support and encouragement.

Abstract (Italiano)

L'attività di ricerca descritta in questa tesi è stata condotta nell'ambito dell'informatica medica e ha riguardato in particolare i sistemi di supporto alle decisioni cliniche (CDSSs). Il volume, la varietà e la pronta disponibilità di dati relativi ai pazienti costituisce oggi, per la pratica clinica, un'importante opportunità per utilizzare queste informazioni per fornire un'assistenza sanitaria che sia personalizzata al livello del singolo paziente. Tuttavia l'eterogeneità e la dimensione delle informazioni disponibili pongono sfide importanti per quanto riguarda la rappresentazione dei dati e l'integrazione delle diverse sorgenti da cui provengono. Infatti, di fronte ad un problema di scelta medica, un decisore dovrebbe prendere in considerazione numerose fonti di informazione. Una lista non esaustiva delle più importanti potrebbe comprendere pubblicazioni contenenti le più recenti conoscenze scientifiche, database degli ospedali, dati provenienti da sensori collegati agli smartphone dei pazienti, preferenze personali dei pazienti, informazioni sul loro stile di vita e il contesto in cui essi vivono. Tutte queste informazioni possono contribuire alla possibilità di fornire un supporto informatizzato per medici e pazienti che sia più personalizzato possibile e che abbia il potenziale effetto di aumentare l'adozione delle *best practice* contenute nelle linee guida cliniche e di migliorare la *compliance* al trattamento per i pazienti.

Metodologie di intelligenza artificiale e CDSSs sono qui proposti come alcuni degli strumenti più promettenti in grado di agevolare l'attuazione degli obiettivi appena descritti, sia in un contesto medicalizzato (come un ospedale) sia in un contesto di pazienti non ricoverati che vengono quindi assistiti al di fuori di un ambiente medicalmente protetto (ad esempio a casa). In particolare, le metodologie proprie dell'analisi decisionale sono tradizionalmente utilizzate per rappresentare problemi decisionali e fornire strumenti di analisi per identificare meglio le scelte ottimali in base ai dati disponibili. Tuttavia negli ultimi anni è stata proposta una nuova pratica in cui il paziente e il suo medico, quando le condizioni lo consentono, condividono informazioni e responsabilità per giungere assieme a una decisione clinica. Questo processo prende il nome di *shared decision making* e il suo obiettivo principale è quello di considerare sia i dati scientifici e clinici disponibili sia le preferenze del paziente rispetto alle possibili opzioni e alle loro conseguenze. Infatti è bene tenere conto che, anche in un contesto medico strettamente evidence-based, vi sono situazioni in cui il processo decisionale può in ultima analisi dipendere dalle preferenze del paziente. In questi casi, la capacità di trovare una soluzione ottimale (ovvero identificare quella preferibile a tutte le altre)

può dipendere da variabili non strettamente cliniche (quali attitudini personali, abitudini di vita, situazione economica del paziente e le diverse percezioni della qualità della vita associata a diversi stati di salute). Applicare ove possibile lo shared decision making è, in definitiva, un importante valore aggiunto alla pratica clinica moderna e, di conseguenza, ai CDSSs che la supportano.

Per queste ragioni, nel contesto di un progetto di ricerca europeo di 4 anni denominato MobiGuide, è stato progettato e sviluppato un CDSS che offrisse funzionalità di personalizzazione per pazienti e medici. Nel sistema sono state integrate funzionalità avanzate volte ad implementare un framework di shared decision making con l'obiettivo di raggiungere il massimo grado di personalizzazione del supporto fornito dal sistema.

Per far fronte alle problematiche di integrazione dei dati eterogenei è stata sviluppata una soluzione basata sullo standard Virtual Medical Record (vMR) di HL7. In particolare, il modello di dati basato su vMR è stato utilizzato con successo per realizzare un Personal Health Record (PHR) integrato che raccogliesse i dati provenienti dalla cartella clinica informatizzata dell'ospedale, i dati inseriti nel sistema direttamente dai pazienti, i dati provenienti da sensori connessi allo smartphone dei pazienti e infine le raccomandazioni cliniche prodotte dal motore inferenziale di esecuzione della linea guida. Il PHR costituisce uno dei componenti fondamentali del sistema MobiGuide e il suo progetto basato su vMR ha dimostrato di essere una soluzione efficace per il supporto di alcune caratteristiche importanti del sistema finale: flessibilità verso il supporto di diversi domini clinici, possibilità di integrare facilmente fonti di dati diverse, il supporto per ontologie e terminologie mediche standard, la possibilità di estendere il modello dati per rappresentare quelle entità peculiari e uniche del sistema MobiGuide (senza stravolgere il paradigma originale dello standard vMR) e il supporto per un flusso di lavoro e un'architettura distribuita.

Un secondo punto centrale del progetto MobiGuide riguarda lo *shared decision making*. Sono stati sviluppati due alberi decisionali per rappresentare altrettanti punti decisionali che andrebbero sempre condivisi col paziente, riguardanti la terapia anticoagulante orale e l'ablazione, all'interno delle linee guida per la cura della fibrillazione atriale. Tali modelli decisionali sono stati progettati in modo da consentire la personalizzazione delle analisi a livello del singolo paziente. I Quality Adjusted Life Years (QALYs) sono stati scelti come il payoff da considerare nell'analisi decisionale con l'obiettivo di massimizzare la lunghezza della vita attesa ma considerando anche la qualità della vita. E' stato a questo scopo sviluppato un tool specifico per l'elicitazione dei coefficienti di utilità (valori fra 0 e 1 che rappresentano la bontà percepita degli stati di salute). Il lavoro svolto nella tesi propone l'elicitazione delle utilità dei singoli pazienti e il loro utilizzo nella quantificazione degli alberi decisionali come una soluzione efficace per personalizzare modelli decisionali che altrimenti risulterebbero generici. Questa importante modifica rende gli strumenti dell'analisi decisionale (solitamente applicati in studi a livello di popolazione) applicabili anche al singolo paziente, nello

spirito proprio dello shared decision making. La tesi comprende infine ulteriori ricerche sul tema dell'elicitazione delle utilità: si descrive un *recommender system* in grado di predire il valore delle utilità quando i tradizionali metodi di elicitazione diretta falliscono. In particolare viene presentato un prototipo di questo sistema basato sul *collaborative filtering*. Il sistema sfrutta le preferenze di altri pazienti simili al paziente in esame per predire un valore atteso per i coefficienti di utilità ancora sconosciuti. Al fine di dimostrare la validità generale degli approcci metodologici proposti sia lo strumento di elicitazione delle utilità sia il *recommender system* sono stati applicati ad un insieme di pazienti con fibrillazione atriale arruolati nello studio pilota di MobiGuide, mostrando risultati promettenti.

Abstract (English)

The research activity described in this dissertation has been conducted within the field of medical informatics and in particular in the area of clinical decision support systems (CDSSs). The volume, variety and prompt availability of data regarding patients nowadays constitutes an important opportunity for clinical practice to use these information to provide a healthcare that is personalized to a single-patient level. However, heterogeneity and size of the available data pose relevant challenges regarding data representation and integration of heterogeneous sources. In fact, when facing a medical decision, the set of information the decision-maker should consider is very broad. These can include, and are not limited to, latest scientific knowledge, EMRs of the hospitals, data coming from sensors connected to the patients' smartphones, patients' personal preferences, information about their lifestyle and context where they live. All of these information can contribute to the possibility of delivering the most personalized guidance to physicians and patients with the potential effect of increasing adherence to clinical practice guidelines and improving compliance to the treatment for patients.

Artificial intelligence methodologies and in particular CDSSs have been identified as one of the most promising tools able to facilitate the implementation of the abovementioned objectives both in the clinical practice and in the outpatient settings. In particular decision analysis methodologies have been traditionally used to model clinical decision tasks and provide analytical tools to better identify the optimal decision options according to the most updated available evidence. However scientific research has lately been proposing a shift in the clinical decision making paradigm where the patient and his care provider (where it is appropriate and possible) share information and responsibilities and reach a clinical decision together. This process is known as shared decision making and its main goal is to take into account both the available scientific evidence and the patient's perception of the consequences of different decision options. In fact even in in a strictly evidence-based setting, the decision process may ultimately depend on patient preferences. In these cases finding a clear-cut optimal solution may depend on non-strictly-clinical variables such as personal attitudes, lifestyle habits, economic situation of the patient, and different perceptions of the quality of life associated with different health states. Addressing shared decision making and involving patients in medical decisions is ultimately a significant added value to modern clinical practice and to CDSSs supporting it.

For these reasons, in the context of a 4-year European research project named MobiGuide, a personalized CDSSs for patients and their care providers has been designed and developed. Advanced functionalities that implement a theoretical shared decision making framework have been integrated in the system as well as capabilities to make it context-sensitive (and context-responsive) in an overall effort towards personalization.

To face the challenge of data integration a solution based on the HL7 virtual medical record (vMR) standard has been developed. In particular the vMR-based data model has been successfully used to build an integrated personal health record (PHR) which collects data coming from the hospital EMRs, data directly entered into the system by patients, data coming from patient-operated sensors and the outputs of the guideline execution engine in the form of clinical recommendations. The PHR constitutes one of the fundamental building blocks of the entire MobiGuide system and its vMR-based design has proven to be an effective solution to provide support for some important features of the final system: flexibility to support different clinical domains, possibility to easily integrate different data sources, support for standard medical ontologies and terminologies, possibility to extend the standard data model to represent entities that are peculiar only to the MobiGuide system (without impacting the overall standard paradigm of vMR) and support for a distributed architecture and workflow.

A second central aspect of the work carried out within MobiGuide concerns the support to shared decision making. Two decision tree models have been developed to represent as many shared decision points, regarding oral anticoagulant therapy and ablation, in the guideline for the management of Atrial Fibrillation. Such decision models have been designed to allow patient-specific personalization of the decision analysis. Quality Adjusted Life Years (QALYs) were chosen as the payoff to consider when choosing among different treatment options, with the final goal of maximizing length of expected life while also considering quality of life. A specific tool for the computer-assisted elicitation of utility coefficients (0-1 values representing the desirability of health states) was developed as a by-product of the shared decision making implementation. The present work proposes the elicitation of utilities from patients and their use in decision tree models as an effective way of personalizing the otherwise generic decision models and to make them a valuable tool to take better informed decision even at a single-patient level. The dissertation also includes further research on the utility elicitation topic: we propose to build a recommender system to predict the value of utilities when traditional elicitation fails. Moreover a prototype implementation of this system based on collaborative filtering is presented in the dissertation. The system capitalizes on the preferences of other similar patients to predict an expected value for the still unknown utility coefficients of a new patient. In the effort to demonstrate the validity of the proposed approaches both the utility elicitation tool and the recommender system for utilities have been evaluated showing promising results using a set of AF patients enrolled in the MobiGuide pilot trial.

Contents

Chapter 1

Introduction

1.1. Relevance of Clinical Decision Support

The application of information and communication technology (ICT) to health care has a long history starting as early as the 1960s. However the rate of adoption and degree of impact of ICT to clinical practice can still be considered low if compared to the primary role it has assumed in other fields like personal communication or to the radical changes introduced in our daily lives by consumer electronic devices. One of the most appealing applications of computers in health care is their potential to help solve complex problems and make decisions. Health care practitioners have to deal with a wide range of challenges, often competing among themselves, on a daily basis: making difficult diagnoses, avoid errors, ensure highest quality, maximize efficacy and save money all at the same time. These are the basic reasons why clinical decision support is one of the most sought-after functionalities for ICT in healthcare (probably second only to the capability of collecting and storing clinical data in electronic format). However despite the fact that computers can significantly help in achieving most of the abovementioned goals, a widespread adoption of computerized clinical decision support is still to be achieved and their use is far from being accepted as routine in most clinical scenarios. A large number of computer-based aids for decision has been developed over the past 50 years and their usefulness evaluated in several pilot studies both in academic setting and in private organizations. However the need for effective support to decision-making is even more urgent today than in the days of early adoption of these systems. The advances of biomedical discovery including genomics (as well as other "omics" like proteomics or exposomics), the improved understanding of diseases, availability of new technologies for mobile and self-monitoring devices, the exponential increase in the use and penetration of the internet are only some of the factors contributing to the growth of the two main components needed for effective decision making:

information (i.e. data) and knowledge on how to use these information. One of the relevant factors is the increasing adoption of electronic health record (EHR) systems in health organizations. The availability of data in a complete and structured form is a major enabler for the effective implementation of decision support. Moreover, in a setting where ICT is part of the everyday clinical practice, integrating decision aid capabilities into an existing system would be perceived as an added value rather than an extra burden for the daily routine (i.e. extra clicks, use of another, non-integrated system, etc.). Computerized Physician Order Entry (CPOE) systems are a popular example of this fact being one of the earliest field of application of decision support and still one of the most widely adopted. Another important change in the availability and volume of clinically relevant information derives from the fact that the era of "big data" is officially started. This calls for an improved ability to cope with increasing amounts of complex data, being generated and changing at a fast rate, usually unstructured and heterogeneous both in terms of sources and formats. This fact combines with the call for a more and more "personalized" approach to medicine where each patient is unique and also care must be tailored specifically to that individual. Since late 2000s these factors altogether constitute a disruptive trend that could affect health care delivery organizations and practices. Clinical decision support combined with other medical informatics methodologies is one of the most promising candidates to effectively deal with these challenges which will become even more prominent in the near future.

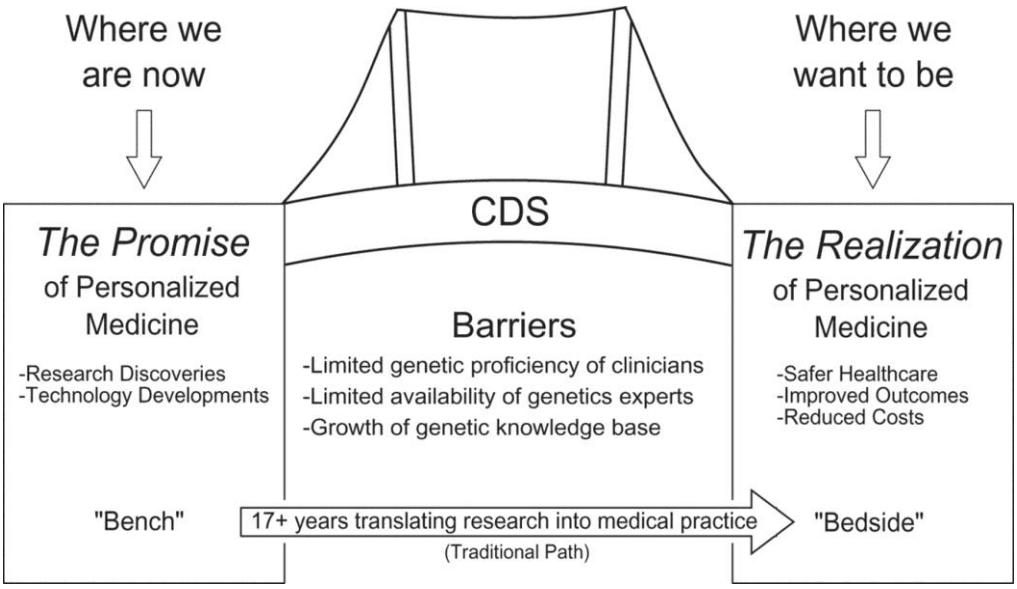


Figure 1 - Clinical decision support can speed up the realization of the promise of personalized medicine [1] while a more traditional path can take up to 17 years to get to the same level of dissemination and adoption [2].

1.2. Dissertation Outline

Starting from the abovementioned scenario this dissertation will explore the topic of clinical decision support systems and their relationship to the current problems of heterogeneity of patients data, need for integration of these systems in the complex ICT infrastructure of health organizations and the concurrent force demanding effective use of all these data sources to provide a personalized support to a single-patient-level.

Chapter 2 gives an introduction to the topic of computerized clinical decision support systems, their definition, scope and a brief history of their evolution. A dedicated section will focus on evidence-based, guideline-driven systems which are particularly relevant to the work described in this dissertation. In the same chapter the concept of personalized medicine is also introduced. The need for attention toward a single-individual-level care process is an essential part of the present work which proposes novel techniques to address the challenge of providing personalized guidance which is both compliant to the most updated medical knowledge available and tailored to the peculiarities of the single patient being treated.

The enabling methodologies used to pursue this goal are described in chapter 3. Firstly the issue of data model standardization is considered. Currently available standards suitable for decision support systems are presented and their strengths and weaknesses analyzed. Secondly we move to the concepts of shared decision-making and the connected framework of methods needed for its implementation. These include decision analysis, decision models like decision trees and Markov processes, and utility theory. Finally chapter 3 also describes some artificial intelligence methodologies that are relevant to the decision support topic when combined with the other presented techniques in an integrated system. These are sentiment analysis, which deals with the problem of analyzing opinions and subjective information contained in natural language text, and recommender systems, a specific class of predictive models able to calculate the expected value of target variable for a specific individual based on data collected on “similar” individuals.

Chapter 4 and 5 apply the methodologies described in chapter 3 to the implementation of an actual clinical decision support system. The EU-funded MobiGuide project lasted 4 years and developed a ubiquitous, distributed and personalized clinical decision support system. Chapter 4 presents the general architecture and functionalities of the project. A particular focus is given to the design and implementation-level solutions adopted for the data model of the personal health record where all the patients’ data is stored and processed. The complex architecture of the components and their distributed nature posed significant challenges to the implementation of such a system where effective data integration between components is a crucial step toward a fully functional system. The

solutions to such challenges are discussed in chapter 4. Chapter 5 on the other hand focuses on one of the most innovative parts of the project. MobiGuide users are not only physicians but, unlike most part of traditional decision support tools, also patients themselves. Indeed guidance about treatment and illness management is provided directly to patients through dedicated user interfaces running in the smartphones of the patients. Having patients as active users however demands for an increased level of personalization of the system where also patient preferences, lifestyle and personal context have an important role in the decision support process. Chapter 5 analyzes the work carried out to implement personalization capabilities in the MobiGuide system along with other tools needed to pursue the same goal. The shared decision implementation framework is presented along with the formal decision models developed, a software tool to perform utility elicitation and a collaborative filtering algorithm for utility function prediction.

Finally chapter 6 gives some final considerations and concludes the dissertation with some remarks about open points and possible future developments of the present work.

Chapter 2

Background

2.1. Clinical Decision Support Systems

2.1.1. A definition

Following Greenes [3] computer-based clinical decision support (CDS) can be defined as:

> *The use of information and communication technologies to bring relevant knowledge to bear on the health care and well-being of a patient.*

It is important to acknowledge the fact that generic clinical decision support can of course be provided without ICT support by means of textbooks, reference material like narrative clinical guidelines and other similar aids. However the scope of this dissertation primarily includes CDS applications implemented by means of general ICT and medical informatics. Such systems are commonly defined as clinical decision support systems (CDSSs). Although decision making process is the primary focus of CDSSs, it is very unlikely that any decision support can be provided without taking into account other aspects such as impact on business process, workflow management and resources management; technological infrastructure; target users of the system; knowledge management and update. A more verbose definition of CDS given by Osheroff [4] explicitly accounts for more of these aspects:

> *Clinical decision support is a process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health and healthcare delivery. Information recipients can include patients,*

> *clinicians and others involved in patient care delivery; information delivered can include general clinical knowledge and guidance, intelligently processed patient data, or a mixture of both; and information delivery formats can be drawn from a rich palette of options that includes data and order entry facilitators, filtered data displays, reference information, alerts and others.*

Some facts are evident from the combination of these two definitions. First of all any CDSS is heavily dependent from two main sources of information: i) patients' clinical data and ii) scientific (medical) knowledge. Secondly, despite the fact that the implementation of a CDSS seems to be a rather simple task, it does imply addressing several challenges regarding which information to use/present, to whom, in which format (e.g. drug order sets, alerts, information snippets from reference material, etc.), through which channels (e.g. integrated in an electronic medical record interface or into a mobile device messaging system) and at what time in the workflow process. In fact the primary objective of CDS is to foster optimal problem solving, decision making and action by a human user. This can be accomplished simply presenting the user with relevant data about the specific situation he's facing (thus easing the decision process) or by actively processing data and providing more elaborate feedback to the user. In both cases the primary task of the CDSS is usually to perform an action that can range from simple filtered representation of a patient's data to a complex feedback in the form of a recommendation of what the optimal decision according to current evidence would be.

2.1.2. Success factors and benefits

Several studies have been carried out to demonstrate the positive effects of the adoption of CDSSs in the medical practice. Among the most pursued effects of CDS it is common to include increased compliance to best clinical practice for physicians or to therapy prescriptions for patients, avoidance of medical errors (e.g. in drug prescriptions, diagnosis etc.) and improved cost management. However the fact that CDS is not widely accepted as an essential part of modern clinical ICT applications tells us that every intervention has to be carefully designed to actually positively impact health care quality measurements. Kawamoto et al [5] identified a set of essential success factors for the implementation of a CDSS in a review of 70 published studies. Results showed that the 4 most important features correlated to CDSSs that improved clinical outcomes where: i) automatic provision of decision support as part of clinician workflow, ii) provision of recommendations rather than just assessments, iii) provision of decision support at the time and location of decision making, and iv) computer based decision support. Of the systems possessing all four features, 94% significantly improved clinical practice. Furthermore, direct

experimental justification was found for providing periodic performance feedback, sharing recommendations with patients, and requesting documentation of reasons for not following recommendations (i.e. reasons for non-compliance).

Positive impact of CDSSs was proven in several published studies [6–8]. For example a recent study [9] on both locally developed CDSSs and commercially available ones proved that they improved health care process measures related to performing preventive services, ordering clinical studies, and prescribing therapies. However few studies still measure potential unintended consequences or adverse effects. Another review [10] showed that practitioner performance was improved in 64% of the 97 studies assessed while, the effects on patient outcomes remain largely understudied and, when studied, inconsistent [10, 6]. However the growing interest in precision and personalized medicine (see section 2.2.2) demands renewed attention to including patients themselves as active users and assessing benefits of CDSSs also on patient outcomes.

2.1.3. Classical applications: expert systems for diagnosis

Several successful historical examples of the application of ICT to clinical decision support have been reported in the literature. In the early days of CDSSs the application of artificial intelligence methodologies to medical problems gave birth to a number of solutions under the name of clinical "expert systems". One of the most studied medical problems that intuitively may benefit from the support of ICT is diagnosis. The promise of automatic diagnosis has long fascinated researchers in the field of medical informatics and artificial intelligence, and a number of methodologies have been applied to build systems which performances were at least on par with the best human medical experts.

One significant example of such a system is Mycin [11]. Developed by a group of researchers at Stanford University medical center, the objective of Mycin was to help physicians decide whether antibiotic therapy was needed or not in case of infections. The system was developed using simple production rules (i.e. if-then statements) and the knowledge base of its final version included approximately 500 rules. The main advantage of Mycin was its ability to use incomplete and uncertain information in the reasoning process. The system underwent a formal evaluation where 10 cases of meningitis had to be labelled as viral, mycotic or bacterial. Seven independent infectivologists were asked to choose a treatment for each case and another 8 medical experts were then asked to evaluate whether the choice of therapy was correct or not. Surprisingly Mycin outperformed all the medical experts suggesting an acceptable therapy in 65% of the cases while experts' performance ranged from 30% to 62,5%.

Another well-known classical example of how ICT could help solve medical problems is PIP, Present Illness Program [12]. Developed by the Clinical Decision Making working group at MIT in collaboration with the

Tufts University School of Medicine and the New England Medical Center Hospital in Boston, the system used frames and slots (a formalisms that can be regarded as a simpler version of classes and attributes of the modern ontologies) to represent knowledge and perform inference. The chosen use case was to determine the causal agent of a set of renal pathologies. PIP was used to acquire the clinical history of patients and output a differential diagnosis list. The most probable diagnosis was then labelled through a special scoring system based on the 70 frames that constituted the knowledge base of the system.

One final example of a classical work where computerized decision support was applied to diagnostic process is CASNET [13]. This system used yet another formalism to model knowledge and perform diagnostic reasoning. Knowledge in the field of physiopathology was used to develop a semantic network connecting observations, pathophysiological states and disease states on three different planes (see Figure 2).

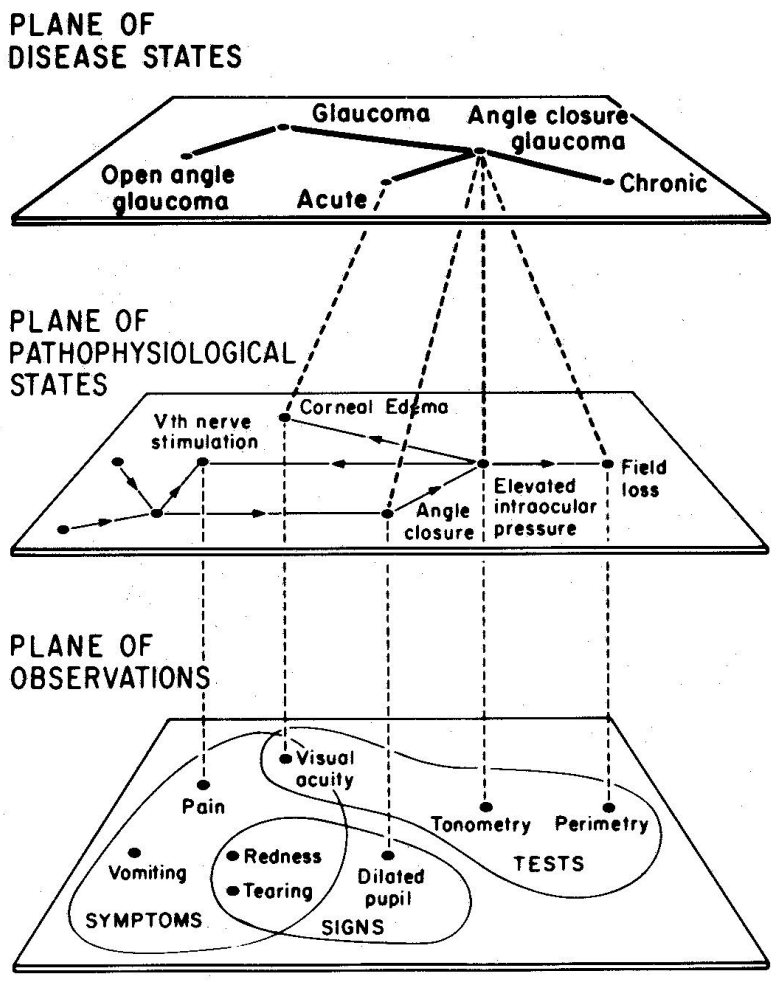


Figure 2 – The example three-layer description of disease abstracted from the CASNET knowledge base modelling the glaucoma use case published in the original article [13].

2.1.4. Current state and methodologies

As highlighted in the previous section, the diagnosis process is probably one of the most representative examples of what clinical decision support can accomplish. However current research trends in the field have significantly broadened both the scope of application and the methodologies used in CDSSs. A dedicated section (2.1.5) will focus on guideline-based DSS, which are of particular interest for this dissertation, while here we present a high-level non-exhaustive overview of the current methodologies used in CDS applications and the their open challenges. Giving an detailed description of these is beyond the scope of this dissertation. Nonetheless a short list of the ones that are relevant to the present work is given below:

- Information retrieval: as previously mentioned one of the key goals of CDSS is to group and present relevant knowledge to the point where it's most needed during clinical practice. Information retrieval methodologies provide an answer to this requirement in the most straightforward way. Various techniques have been applied to the problem of finding information and answering questions, usually posed in the form of natural language text: from text-based approaches for searching scientific literature, to complex statistical methods, to taxonomies and ontologies to organize medical knowledge.
- Logical conditions and rules: these have been widely used to implement intelligent systems able to send alerts, detect possible errors and evaluate logical conditions automatically. Probably the most popular and widespread example of their application is computerized physician order entry (CPOE).
- Probabilistic and data-driven approaches: this family groups together all the machine learning and probabilistic methodologies used to build statistical models from clinical data. Classification (e.g. for diagnosis or any other labelling problem), decision analysis (see chapter 3) and predictive models are some of the most popular applications.
- Heuristic modelling and expert systems: mainly applied to diagnostic and reasoning systems. See previous section 2.1.3 for some examples and further details.
- Workflow management: used to model and execute clinical and organizational process flows (Figure 3) workflow management methodology is particularly relevant for guideline-based systems and guideline modelling languages. These systems will be discussed in further detail in the following section 2.1.5.

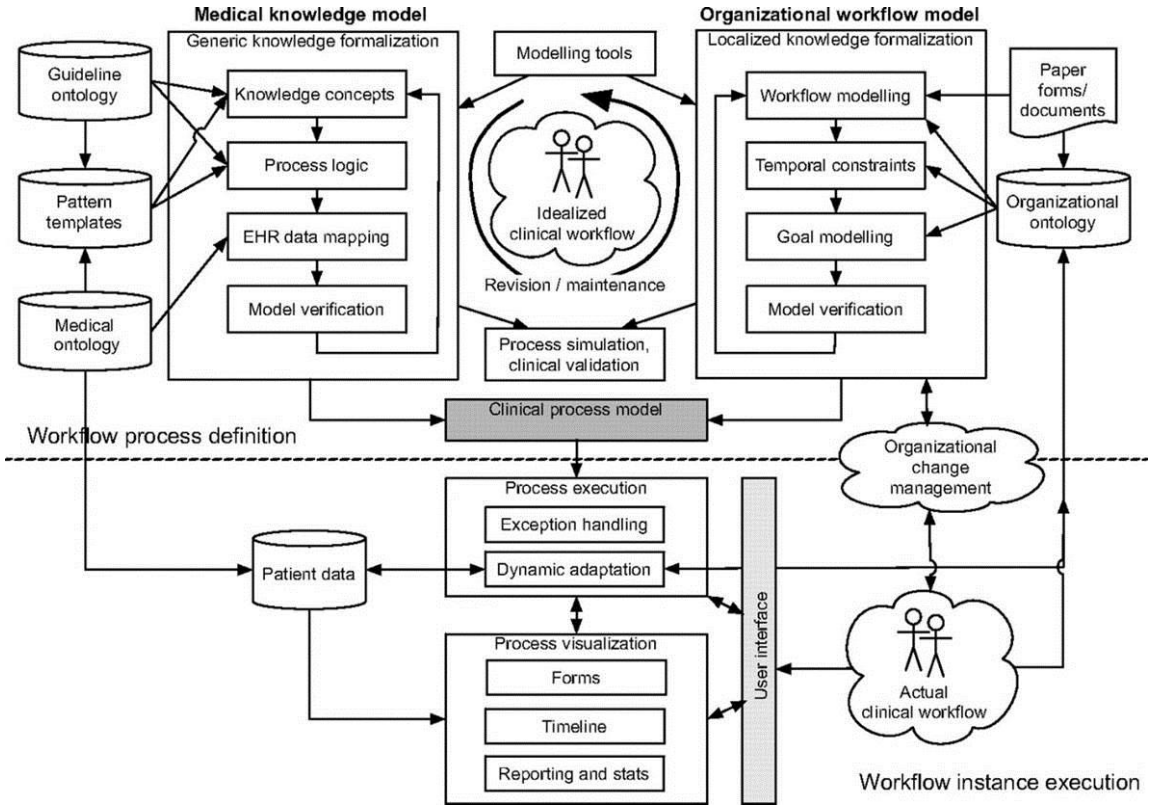


Figure 3 - Conceptual model of modern process-oriented health information systems [14]. The combination of medical knowledge model and workflow model together with patient data enables support for the execution of complex clinical processes.

2.1.5. Guideline-based decision support

Evidence-based medicine (EBM) is a research trend in which every clinical action seeks to be annotated with the sources of scientific evidence that prove its validity. The term has been used since in the early 1990s when a series of articles from the Evidence-Based Medicine Working Group (based at McMaster University) introduced the idea. EBM has the main goal of ensuring high quality and safety of medical practice, reducing costs and speeding up the time needed for the transfer of new research findings into practice.

One of the most relevant products of EBM are clinical practice guidelines (CPGs). A CPG is a collectively produced document (usually authored by scientific committees of experts) that groups the most updated available knowledge for the treatment of a disease or condition with the purpose of disseminating such knowledge and standardize care to ensure highest quality. In particular the American Institute of Medicine (IOM) defines CPGs as statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefit and harms of alternative care options [15]. In order for CPGs to be effective, clinical guidelines need to be integrated

with the care flow of the specific healthcare organization involved, and provide appropriate recommendations when and where needed to healthcare providers. This lead to the development of a specific family of CDSSs based on the formal representation of CPGs as computer-interpretable guidelines (CIGs) [16]. CIGs enable the guideline-based DSSs to perform computer-based reasoning and execution of these formalized models. Such systems match formalized guideline knowledge with patient clinical data to provide patient specific advice at the point of care, increasing the chance of impacting clinician behavior compared to using only the narrative guidelines.

A recent work by Peleg [17] highlighted how CIGs are hot research topic in the area of clinical decision support. A complex framework of research topics including knowledge representation and acquisition, integration with electronic health records and guideline execution engines is involved in the full CIG lifecycle (see Figure 4). Although research in some of these areas (like knowledge acquisition and formal language to represent CIGs) have reached a good level of maturity, there is still important work to be done to reach the same results in other strategic aspects like execution tools and integration with EHRs and workflow systems (Figure 5).

Despite a good amount of research has been carried out on CIGs-based DSSs some areas are still rather unexplored and their relevance is expected to increase in the following years. Some of these topics include advanced functionalities like CIG interactions (i.e. how to combine the execution of concurrent guidelines on patients with multiple comorbidities) or the use of machine learning techniques like process mining to provide feedback back to the guideline definition stage. In this dissertation however we will specifically focus on of these future trends of CIG-based DSS: i) patient-centric systems (refer to the next section on precision medicine for an introduction to the relevance of the topic, and to chapters 3 and 5 for a description of the shared decision methodology and its practical implementation) and ii) ubiquitous, portable, distributed, adaptable and context-sensitive system (see chapters 4 and 5 for more details in the context of the MobiGuide project).

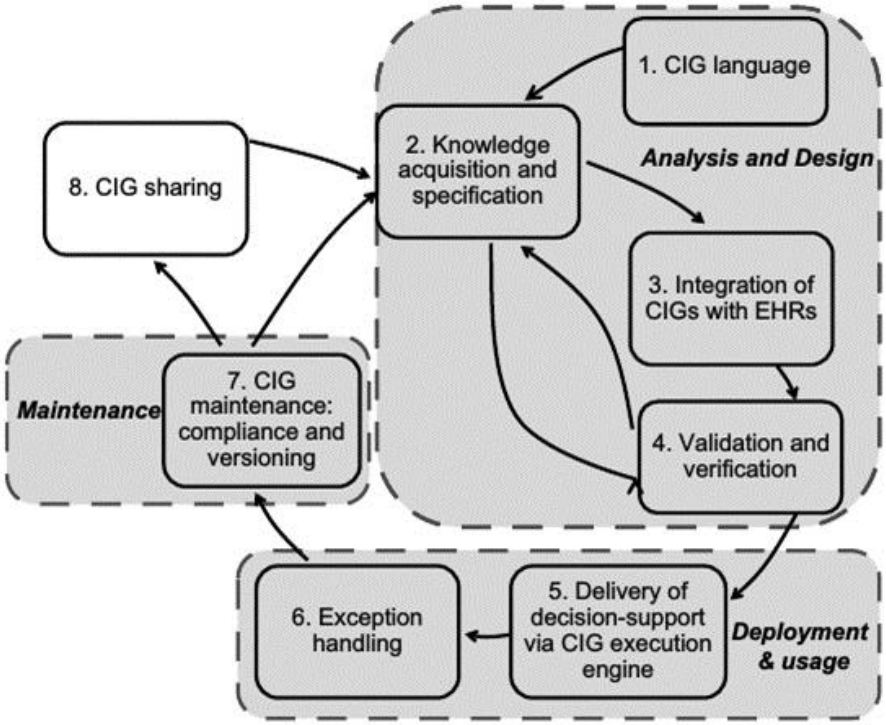


Figure 4 - Topics of CIG related research grouped according to the three main lifecycle stages: Analysis & design, Deployment & usage and Maintenance [17]

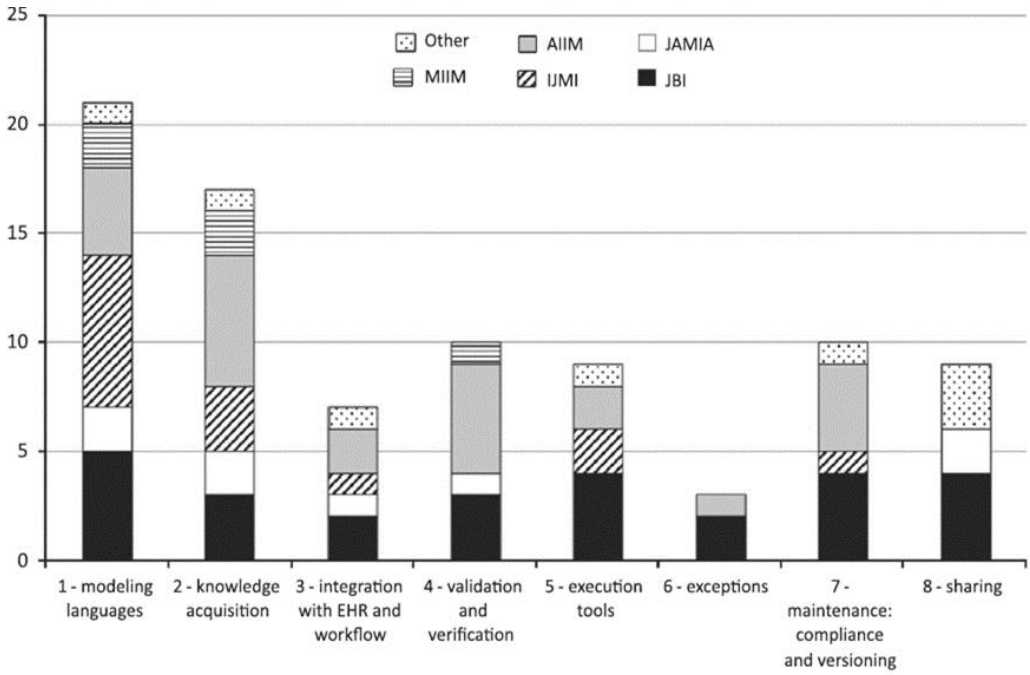


Figure 5 - Number of papers included in the 2013 review by Peleg [17] grouped by topic. Note how integration with EHRs and execution tools are less explored topics than modeling languages and knowledge acquisition.

2.2. Personalized Medicine

Abrahams and Silver in their "The History of Personalized Medicine" [18] point out how Hippocrates himself had the intuition that constitutes the basis of what we call "personalized medicine" today:

> *It's far more important to know*
> *what person the disease has*
> *than what disease the person has.*
> *-- Hippocrates*

The term "personalized medicine" is often described as providing "the right patient with the right drug at the right dose at the right time." More broadly, personalized medicine may be thought of as the tailoring of medical treatment to the individual characteristics, needs and preferences of a patient during all stages of care, including prevention, diagnosis, treatment and follow-up [19].

However the concept of personalized medicine has been applied to a wide and variegated range of research fields and applications. In this section we present some of the most current implications of the broad concept of personalization of the medical treatment, along with its primary implications in the design of CDSSs.

2.2.1. Current trends: Precision Medicine

During his State of the Union address to Congress on 20 January 2015, USA president Barack Obama announced a program called the Precision Medicine Initiative. "I want the country that eliminated polio and mapped the human genome to lead a new era of medicine - one that delivers the right treatment at the right time" [20]. The budget for 2016 consists of 215 million dollars to be managed primarily by NIH, NCI (National Cancer Institute) and FDA. The Precision Medicine Initiative (PMI) will revolve around four key areas of intervention [21] on both near-term and long-term focus:

- More and better treatments for cancer: Oncology is the clear choice for enhancing the near-term impact of precision medicine [22]. Cancer diagnosis has been significantly transformed by the discoveries made by molecular and cellular biology in the latest years and many targeted therapies have been and are being developed. Several of those have been shown to confer significant benefits by targeting tumors with specific molecular signatures.
- Creation of a voluntary national research cohort: On a longer term focus NIH, in collaboration with other agencies and stakeholders, will launch a national, patient-powered research cohort of one million or more Americans who volunteer to participate in

research. Participants will be involved in the design of the Initiative and will have the opportunity to contribute diverse sources of data including medical records; profiles of the patient's genes, metabolites, and microorganisms in and on the body; environmental and lifestyle data; patient-generated information; and personal device and sensor data. This ambitious project will leverage existing research and clinical networks and build on innovative research models that enable patients to be active participants and partners while providing a broadly accessible integrated dataset to qualified researchers to generate new insights [21].

- Commitment to protecting privacy: With the need of such a large scale data sharing initiative proper arrangements must be made to ensure privacy protection and address any legal and technical issues related to the privacy and security of data in the context of Precision Medicine.
- Regulatory modernization: The Initiative will include reviewing the current regulatory landscape to determine whether changes are needed to support the development of this new research and care model (e.g. FDA will develop a new approach for evaluating Next Generation Sequencing technologies).
- Public-private partnerships: A strong collaboration effort with academic medical centers, researchers, foundations, privacy experts, medical ethicists, and medical product innovators would be needed. In addition to that patient participation and empowerment is also a key aspect of the whole initiative.

Despite the PMI was announced at the beginning of 2015 the concept of precision medicine - prevention and treatment strategies that take individual variability into account - is not new [22]. Until recently, most medical treatments have been designed for the "average patient" as a consequence of the strong emphasis placed on large-scale randomized clinical trials and evidence-based medicine. As a result of this "one-size-fits-all" approach, treatments can be very successful for some patients but not for others. Precision Medicine, on the other hand, is an innovative approach that takes into account individual differences in people's genes, environments, and lifestyles [23]. Some early efforts towards the paradigm shift proposed by the PMI were already made as early as 2011 [24] while other initiatives having similar intents were still being labelled with different names like "personalized", and "individualized" medicine [25].

Some examples of how molecular biology and other basic research achievements can positively influence health outcomes and the everyday clinical practice are already well known, especially in the area of cancer treatment. One example consists of non-small cell lung cancer. In this type of cancer two studies [26, 27] demonstrated that 10-15% of patients characterized by the fusion oncogene ELM4-ALK are unresponsive to conventional EGFR inhibitor treatment in cancer therapy. Thanks to this discovery, the drug crizotinib was developed and approved by FDA in

2011. Because of the specificity of this drug, its effect would have never been identified if it were tested in the general population and there would have been the need for a very large cohort to detect a very small effect, present only in the minority of patients bearing the mutation. But thanks to the understanding of the fundamental biology, it was possible to effectively stratify patients and the effects became much more evident in this well-defined population.

It is thus evident that in an era of data deluge, a need to integrate and overlay the information from many different sources into openly accessible and organized datasets is emerging. Indeed without better integration of information both within and between research and medicine, an increasing wealth of information is left unused [24]. Our time has been recognized as the right moment for moving away from the "one-size-fits-all" paradigm mainly because of two enabling factors: i) the fast production pace and increased availability of sheer volumes of data coming from research (e.g. Human genome project [28] and Next Generation Sequencing technologies, molecular biology repositories for sharing knowledge [29], etc.) and patients themselves (e.g. from portable devices for self-monitoring or smartphones); ii) new and improved technologies and tools for handling and analyzing large and heterogeneous datasets [30].

2.2.2. The push toward personalization and patient empowerment

Advances in –omics sciences and their application to tailor a medical treatment to a specific patient or patient group are probably the most straightforward examples of the potential impact of personalized medicine. What is perhaps less apparent but equally important is that initiatives like precision medicine also bring along a revolution in the way doctor-patient relationships and medicine in general work. A certain degree of paternalism and "doctor-knows-best" approach has long dominated medical practice since its very beginning (e.g. see how the AMA's website itself says that "For the more than 160 years, the AMA's Code of Medical Ethics has been the *authoritative* ethics guide for practicing physicians" [31]). Today however this model is changing and shifting to an approach where the patient is the center pivot point. Precision Medicine may be the most fashionable and "hot" topic at the moment (especially since the PMI also meant availability of important funding resources for research in this area) but the implications of a "personalized", "patient-centered" medicine are much broader, and equally relevant.

In the latest years we have assisted to what Topol calls "the Gutenberg moment of medicine" [32]. The author states that much like the printing press took learning out of the hands of a priestly class, the mobile internet, data sharing technologies and enhanced connectivity are doing the same for medicine, giving patients an unprecedented control over their own healthcare. Patients not only have far easier access to their medical data

than before but they more and more often have the ability to create a good share of it themselves. Smartphone connected devices and sensors are increasing in numbers and their cost is rapidly dropping, as it is the one for smartphones [33, 34]. It is realistic to think that soon enough we will be able to run a pretty complete set of lab tests on our own smart device just as many people take their blood pressure or blood glucose measurement on a daily basis. Furthermore the power of the connection between patients (with similar conditions, experiences and needs) provided by social platforms like PatientsLikeMe [35] and the easy access to high quality medical information (e.g. medical literature databases, webMD [36] etc.) has high impact on how patients gather information about their health and treatments. A recent study on physician attitudes showed that not every doctor may be ready to embrace such a paradigm shift: patients who have in-depth knowledge of their condition cause their doctors problems when their expertise is seen as inappropriate in standard healthcare interactions [37]. However some common malpractices like defensive medicine [38] and the consequent willingness to avoid medical errors and improve adherence to clinical practice guideline are contributing to increase the attention to the topic of patient empowerment also among the community of healthcare providers. As of today "the single most unused person in health care is the patient" [39] while "every patient is an expert in their own chosen field, namely themselves and their own life" [40]. Involving patients in decisions about their health management and disease treatments seems like one of the most reasonable things to do in this scenario.

These are the main reasons why personalized medicine also means personal attention to the patients and their individuality. This comprises not only addressing bench science evidence like genetic and molecular analysis (more properly identified as precision medicine or genetically guided personalized medicine [1]) but also considering more subtle differences in patients' lifestyle and personal preferences. These trends obviously also impact CDSSs [41, 42] that can help to effectively harness the sheer amount of data needed to take proper medical decisions. In turn, CDSSs need to be able to include the patient in the decision-making process as well as in the data collection, self-monitoring and analysis of results. A recent Cochrane review, based on 115 randomized clinical trials [43] showed that, compared to usual care, decision aids improve user's knowledge and awareness of personal values for outcomes, create realistic expectations of outcomes, raise active participation in decision making with a positive effect on patients-practitioner communication, and reduce decisional conflict. This type of patient-centered, personalized decision support, that also implies including patients among the active users of CDSS, is one of the main challenges addressed by this dissertation and the effects of introducing patients preferences in a guideline-based clinical decision support system will be further discussed in chapters 3 and 5.

Chapter 3

Methods

3.1. Clinical Data Standards

Recent work in the area of Medical Informatics [44] suggests that the development and deployment of CDSSs in healthcare organizations has the potential to reduce medical errors and increase healthcare quality and efficiency. In particular the use of CDSSs to facilitate evidence-based medicine promises to substantially benefit healthcare services providers as well as clinical outcomes for the treated patients [45]. On the other hand, it is agreed that the implementation of such systems will not be feasible without overcoming traditional barriers for the integration of different patient data scattered throughout several information systems, like Electronic Medical Records (EMR), or more dynamically generated from patient-worn mobile sensors connected to Body Area Networks (BANs). Apart from the traditional terminology standardization issues [46], where different coding specifications can be used to assign an agreed code to a specific clinical concept, more substantial technical and semantic challenges require the use of proper techniques for clinical data representation. Different standards for both the representation and exchange of clinical data between different systems have been developed in the last two decades. Initially, these standards were designed considering the technical and computational issues of clinical data management (e.g., the HL7 v2.x message standard). Despite the fact that this has been a first big help to achieve wide adoption of ICT in healthcare, these standards are still not optimal for data representation and persistence given their low degree of interpretability for non-technical users. Therefore, new standards pursuing a higher abstraction level were developed recently. In the following of this chapter we will present some of the most relevant ones focusing on their strengths and weaknesses when used in combination with a CDSS.

3.1.1. HL7 vMR

HL7 is one of the most well-known Standard Development Organization (SDO) in the healthcare IT market and its standards are commonly used by hospitals for messaging between different existing systems (e.g. HL7 v2.x messages, used in 95% of US healthcare organizations) and recently also in European relevant initiatives (e.g., HL7's Clinical Document Architecture (CDA) is used in the epSOS project [47]). The healthcare data standard models of HL7 are based on the Reference Information Model (RIM).

The RIM is the cornerstone of the HL7 v3.x development process. It is an ANSI (American National Standards Institute) approved standard and it is also adopted by ISO (International Organization of Standardization), concretely ISO/HL7 21731:2014. The RIM is a cross-domain object oriented model that features entities (e.g., living subject, organization, place, material) assuming roles (e.g., patient, employee) participating in acts (e.g., observation about the patient, substance administration, patient encounter). As in all object-oriented models, the classes have attributes and are related to each other via relationships. The HL7 RIM is a high level model which was not designed as a standard for direct implementation but as a reference for all the HL7 v3.x family of standards.

The HL7 Virtual Medical Record (vMR) [48] is an initiative developed by the HL7 Clinical Decision Support Working Group. The standard takes the rich semantic content of the RIM and expresses it in a format that reduces its complexity making it more readily usable for CDSS implementations. VMR is indeed a recent standard specially designed for the purpose of integrating patient data with CDSS. The standard originated from academic research [49] but has been extended based on a multi-institutional analysis of CDS data needs [50] encompassing 20 systems from 4 nations, which included both large scale home grown CDSSs and a number of commercial CDSSs [51].

The standard consists of a small set of classes, simplifying the learning curve for users and the time needed to represent different data items. The model is composed of 22 classes and subclasses and is built upon two main axes represented in Figure 6. The first represents the type of clinical information involved (eight high-level classes including Procedure, Observation, Problem, Substance Administration, AdverseEvent, Goal, Encounter, Supply), and the second the clinical workflow moment (e.g., Proposal, Order, Event), which implicitly represents the source of the information item (e.g. Proposal is produced by a system while Order is produced by a person).

Several papers have highlighted the potential role of a vMR as a solution to the diversity of terminology standards or CIG representations and the issue of institution-specific databases [52, 53]. Health eDecisions initiative [54], a public-private initiative sponsored by the U.S Office of the National Coordinator for Health IT to develop scalable standards for sharing CDS, chose the vMR standard as its foundational data model. Finally the latest version of the vMR is being harmonized with the Quality Data Model [55]

to develop a successor of the vMR known as the HL7 Quality Improvement and Clinical Knowledge model (QUICK) [56].

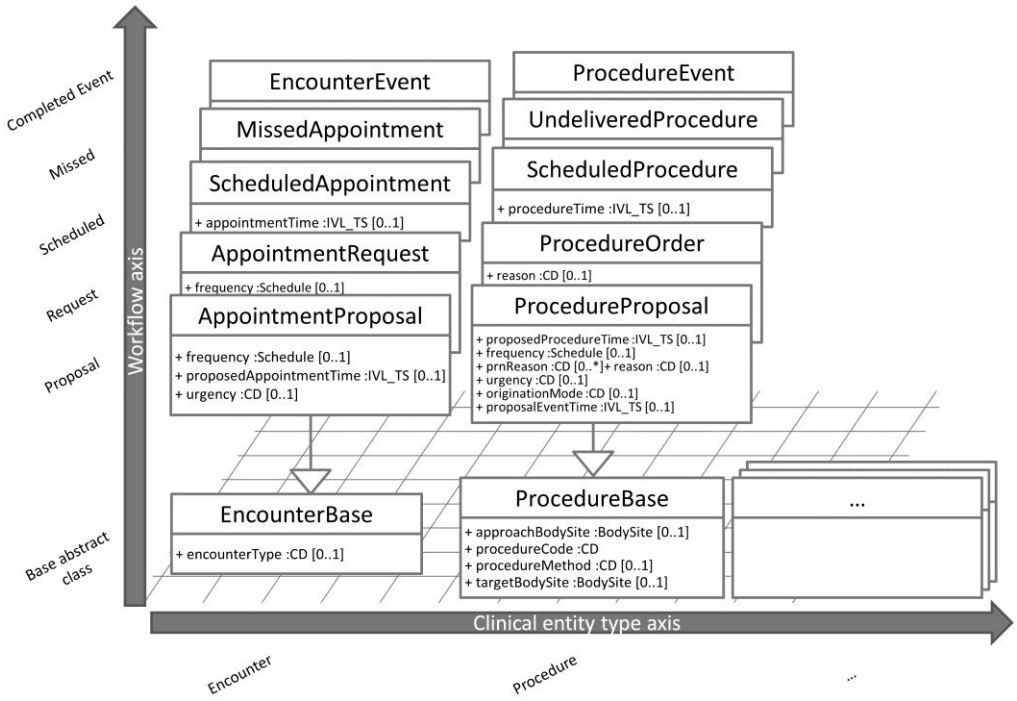


Figure 6 - The two main axes of the vMR model and some example classes.

3.1.2. OpenEHR archetypes

The Archetype model is the cornerstone of the OpenEHR Specification Project [57]. OpenEHR aims at building a set of specifications for a Health Computing Platform consisting of an Application Development platform, a Knowledge Management platform and a Health Integration platform. The project deliverables include requirements, abstract specifications, implementation technology specifications (ITSs), computable expressions and conformance criteria.

The most distinctive feature of the openEHR standard is the archetype model. Via archetypes [58], a separation between clinical concerns and the technical design of the application and data storage is made possible using two-level modeling approach. Many of today's information systems are developed in such a way that the domain concepts which the system is to process are hard-coded directly into its software and database models via an iterative process of writing use cases, finding classes, and building models which will eventually become software. While this single-level approach may allow for relatively quick development, it often results in systems which are expensive to modify and extend, and consequently have a limited lifespan [59]. Indeed some of the main drawbacks of the single level development process are:

- The model encodes only the requirements found during the initial analysis phase. This can be a major issue in the medical domain where domain knowledge is updated on a regular basis and at a fast pace.
- Introduction of new concepts requires software and database changes, and typically rebuilding, testing and redeployment.
- Interoperability is difficult to achieve, since each communicating site must continually either make its models and software compatible with the others, or else continually upgrade software converters or bridges. Heterogeneous computing environments where the software has been created using single-level methodologies typically do not interoperate well, because of the complexity of models underlying each system.
- Standardization is difficult to achieve. With large domain models, it is logistically and technically (and often politically) difficult for different vendors and users to agree on a common model.

The alternative approach proposed is to separate the semantics of information and knowledge into two levels of modelling. While the first level model takes care of the technical concerns and deals with the information structure and data types using an underlying Reference Model (RM), the second level handles the concerns of the clinical domains, which are about how to represent and communicate the semantics of the clinical content. The RM must be small in size, in order to be comprehensible, and contain only non-volatile concepts in order to be maintainable. The second level representing knowledge, requires its own formalism and structure, and is where the numerous, volatile concepts of most domains are expressed. Figure 7 (adapted from [59]) shows the main differences between Health Information Systems developed with the single and two-level methodologies.

The concept of “archetype” is introduced to denote a model defining some domain concept, expressed using constraints on instance structures of an underlying RM [59]. Archetypes indeed act as a bridge to constrain the general technical RM to the specific domain knowledge model at runtime (see Figure 8). Archetypes can be designed from scratch, or adapted from preexisting ones which are publicly accessible from openEHR Clinical Knowledge Manager repository [60]. Furthermore, different archetypes can be aggregated into one by means of archetypes templates, which also support semi-automatic derivation of user interfaces. Another important consequence of using archetypes is that, while technical models are developed by software engineers, knowledge concept definitions are developed directly by domain experts. Specialists are thus empowered to directly produce artifacts which will control how their information systems function.

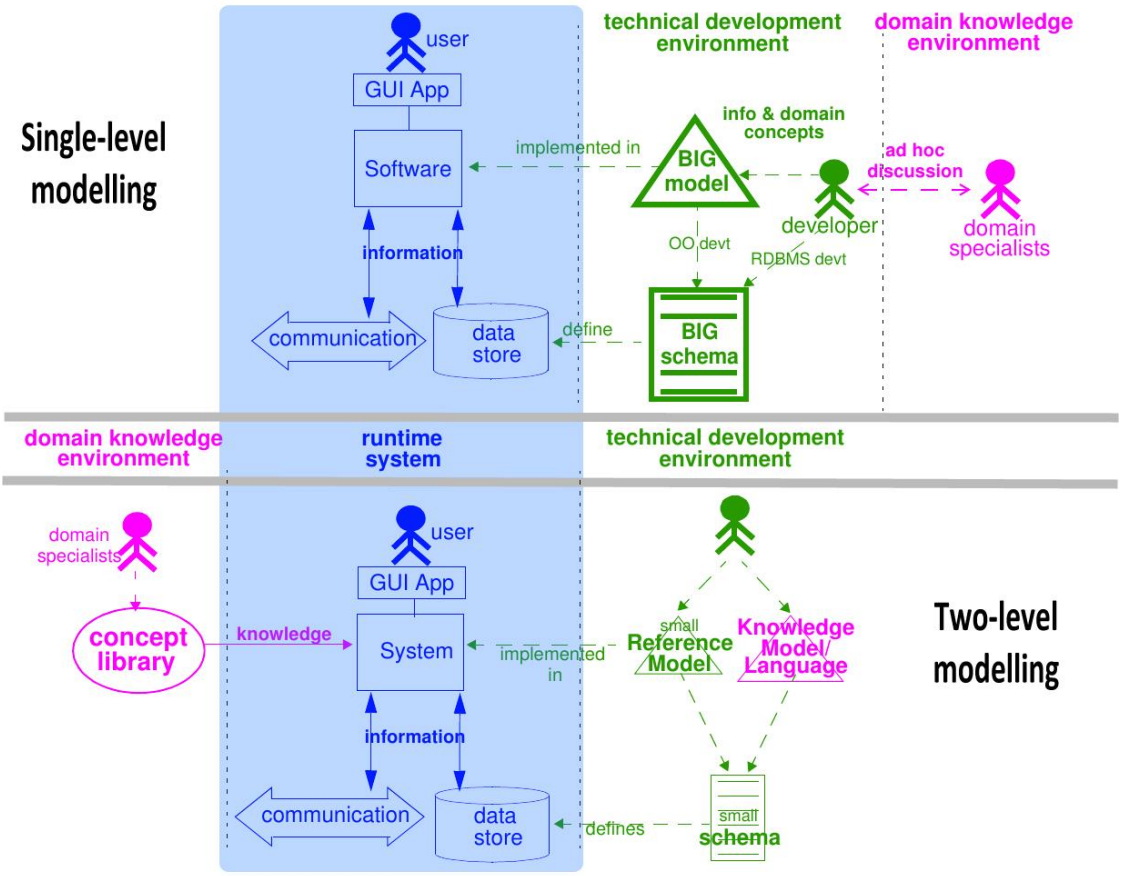


Figure 7 - Single and two-level modelling approaches compared.

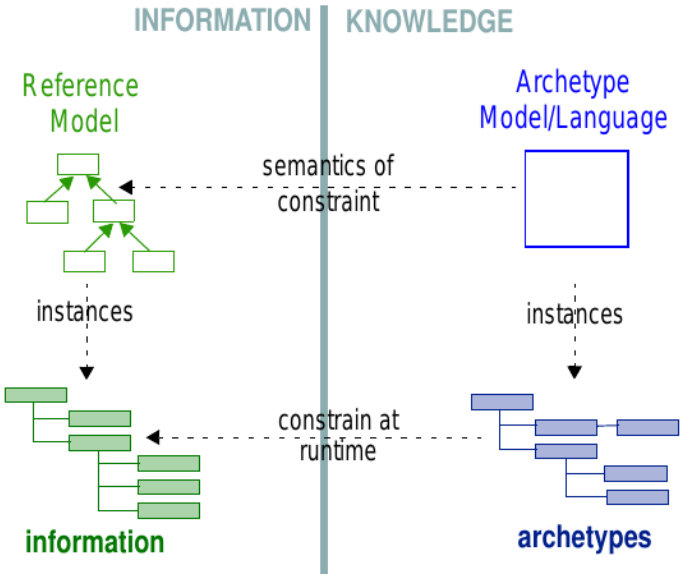


Figure 8 - Information and Knowledge model mutual interactions.

3.1.3. ISO/CEN 13606 norm

ISO/CEN 13606 [61] is a multi-part standard for Healthcare IT systems that includes terminology, security and interface considerations for the standardized exchange of Electronic Health Records and also deals with health information modeling. It proposes to use a dual modeling approach without specifying the format (either the openEHR model or the HL7 CDA combined with templates are expected to be possibilities in this sense [62]). Indeed ISO/CEN 13606 primary objective is to create a standard for the communication of EHR extracts between components, defining a detailed and flexible authorization mechanism, usable in almost any legal situation worldwide. The reason behind this choice is that CEN acknowledges the fact that standards like openEHR or HL7 RIM can provide the semantic level for representing patient data, so their effort has been directed to align the 13606 standard with both initiatives, instead of trying to develop another independent data standard. CEN signed a Memorandum of Understanding (MoU) with HL7 for aligning the CEN information model and the RIM, another MoU with openEHR to adopt the archetype concept [63].However ISO/CEN 13606 norm is still an important standard to comply with especially following its two main considerations: i) the use of a two-level modeling approach and ii) guaranteeing adequate security through authentication and authorization mechanisms which are not natively included in the HL7 or OpenEHR standards (that, being more abstract-level models, delegate it to the implementation and deployment phases).

3.2. Shared Decision Making

Different patients have different behaviors when facing an unequivocal clinical decision that must be taken about their health: some of them opt for a passive approach, and let caregivers fully decide for them. On the very opposite side, there are patients who tend to gather all of the possible information about their disease, and seek for medical advice only after forming a personal opinion about their treatment options. An "intermediate" attitude is that of the patient who methodically inquires the doctor on the details and the rationale behind the available options, in an attempt of obtaining in-depth understanding and possibly participating in the decision [64]. The process during which the patient and his care provider reach a clinical decision together is known as shared decision making [65] and its main goal is to take into account both the available scientific evidence and the patient's perception of the consequences of different options [66]. In fact even in in a evidence-based setting, the decision process may ultimately depend on patient's preferences. In these cases finding a clear-cut optimal solution may depend on non-strictly-clinical variables (such as personal attitudes, lifestyle habits, economic situation of the patient, and different perceptions of the quality of life associated with different health states). Data on long-term survival is arguably the most important outcome information used by clinicians when

selecting a treatment option. In a classical study [67] however, an interesting trade-off between longevity and quality of life was highlighted for patients being treated for laryngeal cancer. In the case of surgery involving removal of the larynx patients had a 60% 3-year survival rate while irradiation had a 30% to 40% 3-year survival rate. Nonetheless 20% of the subjects interviewed in the study would have preferred irradiation, with a decreased 3-year survival rate, but with the benefit of retaining ability to speak. Pros and cons of involving patients in medical decision have been studied by a number of authors [68] highlighting with a good agreement that possible errors (especially in risk evaluation) do not affect shared decision systematically. On the other hand in most cases well informed patients can positively affect the decision process and increase mutual trust between all the involved actors and, finally, increase their satisfaction [32]. Addressing shared decision making and involving patients in medical decisions is ultimately a significant added value to modern clinical practice and to CDSSs supporting it [32, 41, 69]. In the following sections the theoretical foundations (namely decision analysis and utility theory) of medical decision making are presented while chapter 5 will address a case study implementation of the shared decision framework in the context of the MobiGuide project.

3.2.1. Decision analysis and decision models

Decision analysis is a discipline that deals with decision problems involving uncertainty using a systematic, quantitative, rational approach. Decision analysis was first proposed in the field of economics [70] but its application to the medical field is nowadays very popular [71]. Indeed healthcare decision-making can be complex, often requiring decision makers to weigh serious trade-offs, consider patients' values, and incorporate evidence in the face of uncertainty [72]. Medical decisions are made implicitly by clinicians and other decision-makers on a daily basis but the common practice is to base them largely on personal experience. This obviously introduces some biases and limitations like the actual experience of the involved decision-makers and familiarity with the latest scientific evidence available. Decision analysis and cost-effectiveness analysis on the other hand are systematic approaches used to support decision-making under conditions of uncertainty that involve important trade-offs. Some of the main characteristics of decision analysis are: i) it forces the decision-maker to explicitly reason upon and represent the structure of the decision problem in a model; ii) it is quantitative since it involves uncertainties in the form of probabilities, quantifies outcomes with numerical values and thus enables mathematical methods to be used to find the best solution; iii) it is prescriptive and not merely descriptive since the output of decision analysis is ultimately an actual choice of one decision option that is "optimal" (i.e. the one which will maximize the expected outcome) given the available evidence and knowledge.

The basic workflow in clinical decision analysis has been described by several authors [71, 72] and can be described as a list of the following actions:

- define the decision problem (including specifying the decision-maker and the ultimate goal or objective of the decision);
- identify all the decision alternatives;
- list all the possible outcomes of each decision alternative;
- define the relevant time horizon;
- map out the sequence of events leading from the initial decision to the relevant outcomes including chance events and secondary decisions;
- quantify uncertainty: determine the probability of each chance outcome;
- quantify values: assign a value to each outcome;
- calculate the expected value of each decision alternative.

Decision analysis thus requires the use of formal models to represent the decision task and calculate the optimal strategy. Some of the most popular models for such purpose are Decision Trees (DTs) [73]. A graphical representation of a simple DT is given in Figure 9. One of the basic components of a DT is the decision node. Decision nodes (represented as squares in Figure 9) are nodes where different alternative options can be chosen by the decision maker. There can be multiple decision nodes in the same DT and each branch outgoing a decision node represents a decision option. Chance nodes (represented by circles in Figure 9) on the other hand are points where chance, not the decision maker, determines which event will follow. Each branch after a chance node represents a possible alternative and is assigned a probability of occurrence. The sum of all the probabilities of the branches emanating a chance node must equal 1, as one of the modelled events must occur. Finally payoff nodes (represented by rectangles in Figure 9) represent the value of the final outcome of each possible path in the tree. Each leaf of the tree must be represented by a payoff node and each of these nodes can have multiple, often competing (e.g. survival and cost), payoffs.

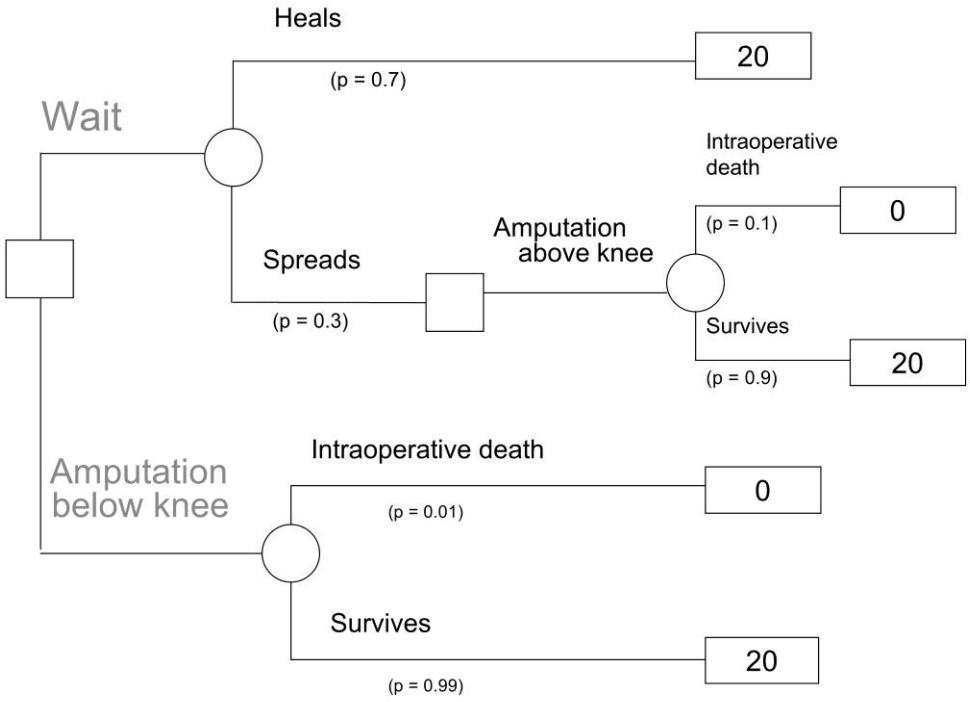


Figure 9 - Example DT modelling the decision about whether or not to immediately amputate an infected foot.

Figure 4 models a DT from an simplified example clinical scenario. Patient x is 58 years old and has been suffering of a peripheral vascular disease for some years. After a severe traumatic injury to the left foot he develops an infection in the foot with significant risk of gangrene. One option would be to immediately amputate part of the leg (below the knee) to prevent the infection from spreading. A more conservative strategy could be to wait and see if the foot would be able to heal without requiring surgery but, in case this does not happen, a more serious surgical intervention (amputation above the knee) would be needed. Finally both the types of interventions, albeit relatively safe, carry a risk of intraoperative death which must be taken into account when deciding for a strategy.

The DT in Figure 4 models the decision problem and has been quantified with probabilities for each of the chance node branches. The final payoff nodes have been assigned with the values of the life years that the patient is expected to live (0 in case of death during surgery, another 20 years if the patient heals from the infection) in each of the possible sequence of events the DT represents. The value of each alternative is calculated starting from the payoff nodes and traversing the tree from the leaves to the root in a procedure known as “rollback” or “folding back” of the DT. The value of each payoff is multiplied by its probability and these partial results are then added at the previous chance node. The process is repeated iteratively until the root decision node is reached. At the end of the rollback the expected

value of each strategy is compared to the other alternatives and the optimal set of decisions is identified accordingly. For our example in Figure 4:

$$"wait"\ strategy: 20 * 0.7 + (0.1 * 0 + 20 * 0.9) * 0.3 = 19.4\ years$$

$$"amputation"\ strategy:\ 20 * 0.99 + 0 * 0.01\ =\ 19.8\ years$$

Thus the decision to immediately treat the patient with surgery is the preferable one since it gives (on average) an higher expected life. However the previous example is based on life expectancy as a single payoff. The DT indeed does not consider at all other important factors such as the various levels of disability that different surgeries would cause to the patient, or costs related to each of the events in the model. The following section 0 will focus on methodologies to incorporate the concept of Quality of Life (QoL) into the decision analysis framework and DTs while in the last part of the current section a brief overview on how to incorporate costs in the analysis will be presented.

Decision analysis can be used to assess the expected costs of decision alternatives. Often public healthcare allocation decisions have the goal of attaining maximal health benefit for a given budget, and this requires information about program effects on health and associated costs. In fact in a regime of scarce resources, there is frequently a trade-off between health outcomes and costs. Cost-effectiveness analysis addresses these kind of problems using the standard techniques of decision analysis but introducing costs as a competing payoff for the different decision alternatives. The objective is to maximize the balance between options where improved health outcomes often come at higher cost. It is important to say that, despite the fact that a good share of cost-effectiveness studies are conducted by public health policy makers, other perspectives can be used to conduct the same type of analysis. These perspective include healthcare organizations, insurance companies or even the single patient perspective. The costs to be considered in the model heavily depend on the perspective chosen. The societal perspective for example requires the broadest approach to characterizing costs and health outcomes of treatment, regardless of who experiences them. On the other hand for a single patient the best choice is to include only his/her out of pocket costs in the model and use values for the outcomes that are closer to his specific situation (e.g. values calculated on a sample population matching his age, general health status etc.) while other costs like the ones covered by health insurance or public health services might not affect the decision at all, being “free of charge” for the patient perspective.

The primary outcome measure of cost-effectiveness analysis is the incremental cost-effectiveness ratio (ICER). The numerator of the ICER is the costs, in monetary units (e.g. euros) while the denominator is the health improvement related to the strategy, typically measured in expected life years, survival at a certain timeframe or other similar measure. In case a decision option is more effective and less expensive it is labelled as

"dominant" and it is clearly chosen as optimal. In the more common case of a treatment is more effective but also more expensive, or conversely somewhat less effective but cheaper, the ICER quantifies the trade-off between the added cost and the health improvement gained. The value of each intervention is then compared with the value of commonly accepted treatment alternatives (or to thresholds defined by organizations like the World Health Organization) to finally assess if they are cost-effective or not. An even more advanced version of cost-effectiveness studies is cost-utility analysis where, in addition to costs, also measures of QoL are taken into account in the selection of the optimal decision. The following section will give more detail on the theoretical background of this methodology and its relevance to decision analysis.

3.2.2. Utility theory and utility elicitation

Cost-utility analysis is a modified version of cost-effectiveness analysis where the effectiveness of a decision is measured in Quality-Adjusted-Life-Years (QALYs). Although they have been criticized by a number of authors and some alternative proposals have arisen in the literature [74–76], QALYs are still one of the most widely adopted measures in studies that compare the outcomes of different healthcare programs. They incorporate mortality and morbidity in a single score, by combining a patient's expected life years with the quality of those life years. More precisely, the number of years spent in a specific health state is "weighted" by a utility coefficient (UC), which values the quality of life in that health state. UCs range from 0 (death) to 1 (perfect health), and they are, in principle, very subjective values, since they reflect a patient's feeling about a health state. QALYs are often used in population studies and in particular in cost-utility analyses, which result is summarized into the incremental cost/utility ratio (ICUR). Assuming that the program under evaluation is more costly and more effective than the actual practice, the ICUR is the cost for each additional QALY achieved by adopting that program. Similarly to ICER, the ICUR is then used to compare alternative interventions or directly compared to standard thresholds (e.g. the World Health Organization, for most of EU countries, suggests a threshold of 30,439$ for QALY gained) for economic appraisal.

Using UCs to quantify QoL in a health state has its theoretical basis in utility theory first proposed in the field of economics by the physicist John von Neumann and his colleague Oskar Morgenstern [70]. The basic idea was that a rational decision-maker, facing a decision problem involving uncertain outcome, would choose the option which maximizes its expected utility. Utility theory is based on 4 basic axioms:

- Completeness: For any couple of states S1 and S1

$$S1 > S2 \; or \; S1 < S2 \; or \; S1 = S2$$

- Transitivity: Given three states S1, S2, S3

$$if\ S1 > S2\ and\ S2 > S3, then\ S1 > S3$$

- Independence (or Reduction of compound lotteries): Given three states S1, S2, S3 where S1>S2 and $\alpha \in [0,1]$ then

$$\alpha S1 + (1 - \alpha)S3 > \alpha S2 + (1 - \alpha)S3$$

- Continuity (or Archimedean property): Given three states S1, S2, S3 where S1>S2>S3 and $\alpha \in [0,1]$ then there exists some probability $\alpha \in [0,1]$ such that

$$\alpha S1 + (1 - \alpha)S3 = S2$$

Preferences of each individual are defined as UCs and assigned to the corresponding possible outcomes by utility functions. In the healthcare setting UCs are defined for each health state and their values are used to quantify decision models like DTs in cost-utility analysis. Analysts have three choices in determining the utility values for use in a study: they can estimate the values using judgment, they can look for suitable published values in the literature, or they can directly measure UCs [77]. The judgmental approach can be a quick and inexpensive way to obtain utility values (e.g. estimates can be provided by the analyst himself or by a few expert physicians) but it is obviously affected by strong limitations and its use should be limited to decision problems where sensitivity analysis shows that the problem is subject to very little change even when performed on a wide range of possible utility values. In most of the cases however UCs should be obtained by more reliable sources like previously published literature. A growing number of scientific journals (e.g. Medical Decision Making, Pharmacoeconomics, Health policy, International journal of health planning management) publish cost-utility studies on a regular basis and UCs are available for a good number of health conditions. However differences in the characteristics of the population where UCs were measured and the sample of the current study might be present. Disregarding the fact that these differences might exist can lead to incorrect analysis results, especially after recent studies have acknowledged that environmental data play an important role in QoL assessment [72]. Generally the most accurate way to obtain utility values for a study is to measure them directly. This involves precise identification of health states for which utilities are required, preparation of health state descriptions, selection of subjects, and the use of a utility measurement instrument [77].

The process of measuring UC values is always referred to as utility coefficients elicitation. Elicitation techniques can be divided into direct and indirect methods. Indirect methods usually involve the administration of a questionnaire to assess QoL experienced by the patient. Some popular

questionnaires like EuroQoL [78, 79] and SF36 [80] are domain independent while others are specifically built to be disease specific. Generic tools like questionnaires however rely on the use of a mathematical formula to convert the questionnaire score to a proper UC. These are usually obtained by regression models where the target variable is a UC elicited with one of the other methods and the predictors are the answers of the questionnaire [81]. Direct methods on the other hand directly elicit a UC from a patient during an encounter and some specific questions. Many elicitation methods have been described in the literature. Here we will focus on three of the most well-known and representative ones, namely Rating Scale (RS), Standard Gamble (SG) and Time Trade-Off (TTO).

In RS an analogue scale is presented to the patient, ranging from 0, associated to the worst imaginable health state (usually death), to 100, corresponding to perfect health. The patient is asked to place a marker on the scale according to the degree of desirability of the health state being evaluated. RS is usually quickly understood by patients and it is often used to rank health states from the less to the most desirable, as multiple states can be placed along the scale in one go. However, the value produced by RS is not a true UC (which must always be based on formal utility theory and derived from a choice between alternatives [82]) but rather what it is called a "value", calculated as x/100 where x is the marker position on the scale.

In the SG method [83] the patient is asked to choose, within a hypothetical scenario, between living the rest of his/her life in the health state that is being evaluated or accepting a gamble whose outcomes are complete healing or sudden, painless death with probability p (see Figure 10a). The more a patient is experiencing poor QoL, the higher risk he would accept to have a chance of healing. The probability p is varied until the patient is indifferent between the two choices. UC is then calculated as $(1 - p)$.

Finally in TTO [84], the patient is asked to choose between living his entire remaining life (t1) in the specific health state being evaluated or to live shorter ($t2 < t1$) but in a perfect health state (see Figure 10b). If the patient is experiencing poor QoL, he will be willing to trade some of his remaining expected life (i.e. to live shorter) for a better QoL. Similarly to p in SG, the amount of time a patient is proposed to give up to heal completely is varied until the patient is indifferent between the two choices. The UC is then calculated as t2/t1.

3.2.3. Markov Models

A Markov Model (MM) is a particular type of time and state-discrete stochastic process that is of particular use in medical decision making [85] to follow a cohort of individuals over time and to determine how the intervention affects costs and benefits. MMs are often used to model processes, like the evolution of a chronic disease, where the patient goes

through a series of states (e.g. with different degrees of severity) over time in a non-deterministic fashion. The model represents a series of states and the probability of all the possible transitions between them in the form of a transition probability matrix. The states must be mutually exclusive and the transitions from one state to another are referred to a particular time range, called the Markov cycle. The basic property of a Markov process is the markovian assumption: i.e. the probability of transitioning at a certain state at time i+1 is only dependent of the current state at time i. MMs are often used in combination with DTs to drive decision analysis or conduct micro-simulations. In the graphical representation of a standard MM, each state is denoted by a circle. Transitions are represented by directed edges. An edge that returns to the same state from which it originates indicates that it is possible to stay in the same state for more than one cycle. The numbers above each edge identify the probability to move from one state to another; and the probabilities outgoing from each node must sum up to 1. A simple example of the graphical representation of a MM is given in Figure 11 along with the corresponding transition matrix.

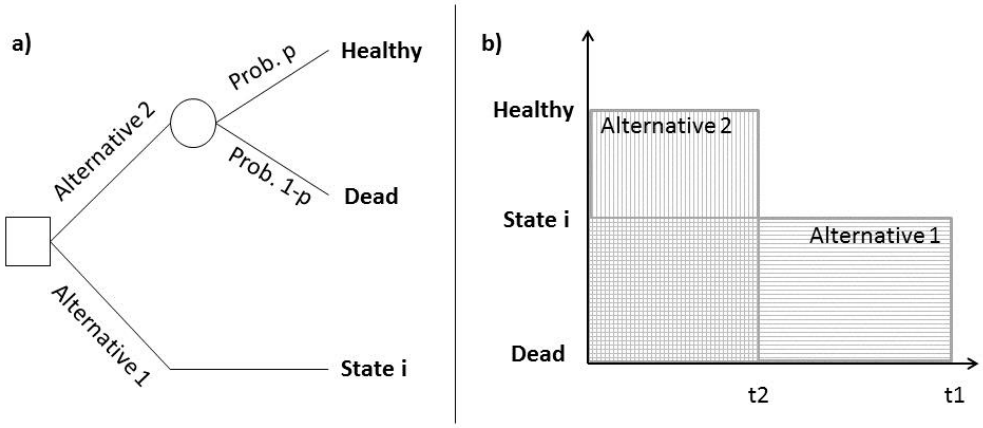
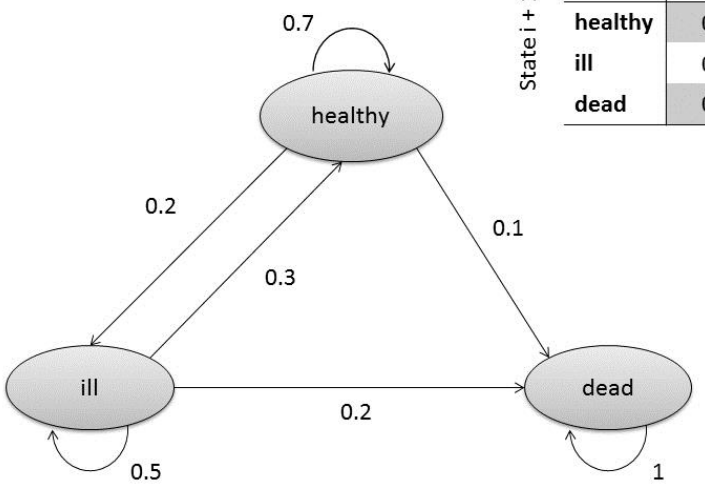


Figure 10 - graphical representation of (a) Standard Gamble and (b) Time Trade-Off utility elicitation methods.



State i

State i + 1	healthy	ill	dead
healthy	0.7	0.3	0
ill	0.2	0.5	0
dead	0.1	0.2	1

Figure 11 - Graphical representation of a Markov Model and the corresponding transition matrix.

3.3. Sentiment analysis

3.3.1. Relevance to medical decision making

When facing a decision problem one of the most natural behaviors that people pursue is to find out what "other people think". Apart from objective data coming from scientific research and other traditional sources also subjective opinions play a very important role in decision processes. The medical decision making context we presented in the previous sections is also subject to this kind of effect, especially if we consider the intention of shared decision to systematically consider patient preferences and opinions in the decision process. The increasing amount and popularity of opinion rich resources like online review sites, communities, social networks and personal blogs enable more and more people to use information technologies to seek out information and understand the experiences and opinions of others before taking their own decisions [86, 87]. This is true also in the medical domain where the success of communities like PatientsLikeMe [35, 88], Treato [89], WebMD [36] and countless health related discussion boards (e.g. DailyStrength [90]) are common examples. The relevance of this phenomenon recently attracted the interest of an increasing number of research groups trying to exploit the potential of such information sources [91–94]. The main challenge consists in the fact that most of this information is in unstructured form (i.e. natural language text) and includes mostly subjective information which traditional techniques are unable to effectively process. These are the reasons that

gave rise to a specific research area named "sentiment analysis" which is particularly relevant to some of the work presented in section 5.2.2 of this dissertation.

3.3.2. Definition and common applications

Sentiment analysis is a subarea of text mining that deals with the computational treatment of opinion, sentiment, and subjectivity in text [95]. The term has been used somehow interchangeably with other definitions like opinion mining or subjectivity analysis especially in the early days of this new research area when a consensus on terminology was still to be achieved. Sentiment analysis is a rather young field that, apart from very few pioneer projects, dates back to the early 2000s when a number of works in the area of finance and marketing fields started to be published [96–98]. Since then marketing, market surveillance and customer relationship management have steadily been some of the fields where sentiment analysis has been successfully applied to use customers opinions (e.g. in product reviews) as a feedback to improve product development and marketing strategies. In addition to these, applications to areas like politics where people's opinions are an important asset have acquired increasing interest over the years [99, 100]. Lately several applications to the medical domain have also started to appear. Among those some of the most significant include epidemiological studies [101, 102], post-marketing drug surveillance [103] and analysis of risky health-related behavior (e.g. smoking or drugs consumption) in the general population [104, 105].

3.3.3. Mining subjective information in texts

Common tasks of sentiment analysis

Sentiment analysis deals with a number of different problems that share a common requirement: the need to automatically process ad interpret subjective information contained in text. Many sentiment analysis applications can indeed be reduced to a combination of a relatively small set of basic problems. Some of the most relevant are:

- Subjectivity detection and extraction: The objective of this kind of tasks is to determine what portions of a text contain subjective information and to distinguish them from purely fact-based information;
- Polarity classification: Deals with the assignment of a polarity class (positive, negative or neutral) to a document or specific portion of text;
- Degree of positivity/negativity assessment: Similar to polarity classification but with the additional aim to assess an "intensity" value for positive and negative sentiments;

- Opinion summarization: Automatically generate a summary of the opinion-oriented portions of a text. This can be done in either purely textual form or with a numerical scoring system (e.g. 0-5 stars scales of many popular review websites);
- Detection of fake or deceptive opinions: Aims at detecting anomalies in the sentiment of a corpus of texts in the effort of identifying fake opinions or fraudulent behaviors. Examples include filtering of opinion spamming [106] (e.g. automatic posting by bots) and detection of counterfeit content to discredit competitors (e.g. fake news or product reviews).

Supervised and unsupervised approaches

The aforementioned tasks share a significant amount of characteristics with traditional text mining problems. In fact many of the techniques used in sentiment analysis are a combination of traditional Natural Language Processing (NLP), information retrieval and other text mining methodologies. However sentiment information is more subject than facts to a number of effects like negations, irony, reported speech and spelling variations. These, together with other factors, make sentiment analysis a rather challenging task which needs dedicated methodologies [107]. Two main approaches have been proposed to face the challenges of sentiment analysis: i) supervised learning and ii) rule-based approaches. In the following we will rely on the polarity classification problem to highlight the main features of the two methodologies and their strengths and weaknesses.

Assigning a positive or negative polarity to a document is basically a binary classification problem. Many traditional text classification problems consist in assigning a document to a class corresponding to its topic and supervised machine learning is a popular approach to these classification problems. In such classifications, topic-related words are usually the key features on which statistical models (e.g. decision trees/forests, naïve Bayes, support vector machines) are trained. Intuitively in sentiment classification, sentiment or opinion words that indicate positive or negative opinions are more important. In a classical paper by Pang, Lee and Vaithyanathan [108] a bag-of-words approach using unigrams was applied to classify movie reviews as positive or negative. Much alike other supervised machine learning applications, the key for effective sentiment classification is the engineering of a set of well-designed features. Term frequency and their variations (e.g. frequency of n-grams, TF-IDF), part of speech tags (e.g. focusing on adjectives or adverbs that imply mood and sentiment) and syntactic dependencies (e.g. parsing or dependency trees) have all been used in numerous works regarding polarity classification. One of the most evident barriers to the adoption of supervised approaches to sentiment analysis is the availability of good quality annotated data to train the classification models. Although this kind of data can be readily available for some applications (e.g. in movie reviews the text of the

review is usually available together with a star-based rating score. In this case feature vectors are calculated from the text while the star score acts as the target variable. 0-2 stars are assumed to be negative while 3-5 are labelled as positive) this is not necessarily true for all domains. Interactions between specific topics and sentiment play an important role in assessing the polarity of text documents. Some expressions that have negative polarity in one domain might have a negative connotation in a different one. "Unpredictable" might be a positive thing to say about a movie plot but something highly undesirable for a surgical intervention outcome. This is one reason why available annotated dataset poorly translate from one domain to the other.

Unsupervised approaches on the other hand do not rely on available training data to classify document polarity. Instead the core idea is to develop a sentiment lexicon, add rule-based knowledge to handle sentiment modifiers (e.g. negations, intensifiers, etc.) and finally define a scoring function whose output classifies the sentiment as positive or negative. The generation of the lexicon, a collection of words and expressions together with their polarity, is obviously a strategic step. A good number of complete resources developed in previous research work are publicly available [109–112]. However, especially in the medical domain, the performance of lexicon-based approaches are often dependent from the specificity to the particular task (e.g. drug adverse effects monitoring vs. disease outbreaks surveillance) of the dictionary resources used. For this reason a good share of the research work in this area focuses on novel approaches to build such sentiment lexicons. Several approaches have been proposed, some of which rely on machine learning algorithms and domain-specific corpora [113], while others use semantic information from structured knowledge sources to build the lexicon [103]. Lexicon based approaches usually outperform supervised learning in some specific settings. For example, rule-based approaches tend to perform better on shorter texts (e.g. single sentences) where the presence of just a few words make the construction of a rich feature vector problematic. However the two approaches are not mutually exclusive. In fact, in recently published works, combined methodologies using the output of rule-based methods as features for machine learning models, have been shown to achieve better performances than each of the two approaches used separately [114, 115].

3.4. Recommender systems

In this section we introduce a family of techniques known under the name of recommender systems (RSs). The relevance of this topic for the present dissertation follows from what has been anticipated in the previous section 3.3.1. In particular, in the shared decision setting, patient preferences play an essential role in decision processes. Despite the fact that utility theory provides a sound framework for preferences elicitation, there might be some cases where data on patient preferences might be unavailable or very

difficult to obtain. We aim to explore RSs as an alternative solution to obtain patient preferences in this setting.

Recommender systems are a popular and widespread approach used to predict user preferences in several industry settings today. The following sections provide an overview of the different approaches for the design and implementation of RSs in general. Then, the following chapter 5, section 5.2.2, will propose an application of this methodology to the medical domain to facilitate personalization of the decision process.

3.4.1. Industry applications

In their most classical definition recommender systems (RSs) are software tools and techniques providing suggestions for items to be of use to a user [116]. Usually the suggestions relate to various decision-making processes, such as what items to buy, what music to listen to, or what online news to read. The growth in volume and variety of the information available on the web and the introduction of new e-business services (e.g. e-commerce or product comparisons websites) have led to the availability of countless alternative choices for potential customers. This variety however, instead of producing a benefit, is often overwhelming for the user who then seeks for guidance to select only those items that are best fitted to him. Many prominent players of the e-business sector (e.g. Netflix or Amazon) are currently investing a lot of resources in the design and implementation of RSs with a double objective: i) guide users in a personalized way to interesting or useful objects in a large space of possible options, improving user experience and ii) maximize their profits, conversion rate[1] and cross-selling by offering a catalogue of products that is personalized to each user. RSs account for a significant part of e-businesses revenues. Recent data suggest that 2/3 of the movies watched on Netflix are those being recommended, 35% of the sales on Amazon.com are from recommended products and, also in the news sector, recommendations generate 38% more clickthrough on Google News [117].

3.4.2. Problem formulation

The problem of recommending items can ultimately be reduced to the task of inferring user preferences from available data. In fact, personalized recommendations are often offered as ranked lists of items and, to perform such ranking, RSs try to predict what the most suitable products or services are, based on the user's preferences and constraints. In order to complete such task, RSs collect from users their preferences, which are either explicitly expressed (e.g. as ratings for products), or are inferred by interpreting user actions (e.g. visiting a particular product page can be

[1] conversion rate = number of sale transactions / number of visits

interpreted as an implicit sign of preference for the items shown on that page).

In the following of this section we will rely on the popular example of movie ratings to provide a formal representation of the RS problem. One of the essential building blocks of every RS is the utility matrix.

Given:

$n_u := number\ of\ users$

$n_m := number\ of\ movies$

$r(i,j) := 1\ if\ user\ j\ has\ rated\ movie\ i, 0\ otherwise$

$y(i,j) := rating\ given\ by\ user\ j\ to\ movie\ i\ (given\ r(i,j) = 1)$

We define a utility matrix of $n_m \times n_u$ elements representing the ratings (0-5 stars) given to the different movies by all the users. Table 1 gives an example representation of this matrix.

Table 1 - example utility matrix for the movie ratings use case

Movie	Alice	Bob	Carol	Dave
Love plus love	5	5	0	0
Romance forever	5	?	?	0
The bride	?	4	0	?
Car crash	0	0	5	4
Guns vs karate	0	0	5	?

The utility matrix can be thought as an extensive "database" of user ratings/preferences and is the basis of all RSs methodologies. The basic task of a RS is to fill in the gaps of the utility matrix, predicting the missing values using the available information about users, preferences and characteristics of the items considered. Several different approaches have been proposed for the implementation of RSs. In the following we will present the details of two of the most popular families of systems, namely content-based recommendations and collaborative filtering.

3.4.3. Content-based approach

Content-based RSs try to recommend items to the users based on the properties of the different items. In this kind of approach the system learns to recommend items that have similar properties to the items a user has liked in the past. In our example the content of each movie *i* is represented through a vector $x^{(i)}$ of features $x_1 ... x_n$ and used together with the utility matrix (Table 2) as an input to the RS.

Table 2 - in addition to the utility matrix content-based approach rely on a vector of features $x_1 ... x_n$ that describe the content of each item. In our

example two features representing the amount of romance and the amount of action are defined for each movie.

Movie	Alice	Bob	Carol	Dave	x_1 (romance)	x_2 (action)
Love plus love	5	5	0	0	0.9	0
Romance forever	5	?	?	0	1	0.01
The bride	?	4	0	?	0.99	0
Car crash	0	0	5	4	0.1	1
Guns vs karate	0	0	5	?	0	0.9

Every user is also characterized by different preferences that are formalized in the form of a user profile. Profile of the user *j* is usually represented by a vector $\theta^{(j)}$. Profiles are then used together with content of each movie to recommend items that have the highest chance of being fit for the user profile. An high-level architecture for a general content-based RS is represented in Figure 12.

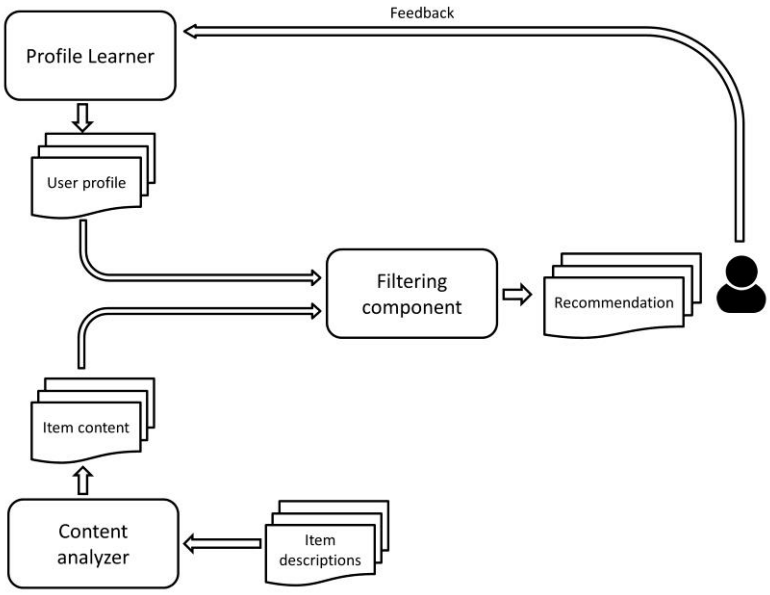


Figure 12- High-level architecture of content-based RSs. The three main components are: the content analyzer, the profile learner and the filtering component.

Different approaches are available to build patient profiles. New users can explicitly provide their preferences associated to each feature x_i (e.g. I love action movies and totally hate romantic ones). As an alternative user profiles $\theta^{(j)}$ can be learned directly from the previous ratings the same user gave to other movies.

RSs can be divided in two main categories: memory-based (where all the raw data in the utility matrix is used each time a new recommendation is requested) and model-based (where a proper model is built from the available data during a training phase and then used to recommend new items). Both these approaches can be used to build a content-based RS. A simple memory based system would just look at item features to determine what items the user hasn't rated yet are similar to those he liked in the past. Those items, identified by using a similarity measure based on item feature vectors $x^{(i)}$ (e.g. Pearson correlation or cosine similarity are popular choices), would be the ones recommended to the user. On the other hand also a model based approach can be chosen. For example the movie rating problem in our example can be thought as regression problem where:

$x^{(i)} := feature\ vector\ of\ the\ movie\ i$
$\vartheta^{(j)} := vector\ representing\ patient\ profile\ (i.e. preferences)$
$r(i,j) := 1\ if\ user\ j\ has\ rated\ movie\ i$
$y(i,j) := rating\ for\ the\ user\ j\ of\ the\ movie\ i$

And the predicted rating for the user j of the movie i given $r(i,j) \neq 1$ can be calculated using a linear regression model:

$y(i,j) = (\vartheta^{(j)})^T x^{(i)}$

In this setting the parameter vectors of the regression are represented by the user profiles $\vartheta^{(j)}$. Those parameters can be learned from the available data minimizing a typical mean squared error cost function:

$$\min_{\vartheta^{(j)}} \frac{1}{2} \sum_{i:r(i,j)=1} \left((\vartheta^{(j)})^T x^{(i)} - y(i,j) \right)^2 + \frac{\lambda}{2} \sum_{k=1}^{n} \left(\vartheta_k^{(j)} \right)^2$$

The main advantages of content-based RS are their relatively simple implementation and the possibility to easily justify the rationale behind each recommendation. However these approaches also have some limitations. First of all they assume it is possible to know, for each new item in the catalogue, its content (i.e. feature vectors $x^{(i)}$ have to be known for each item). Secondly, since they rely on the preferences of only one user at a time, each user has to rate at least some items before any recommendation can be generated for him. This is also known as the "cold start" problem for RSs. Finally information about the similarity of preferences between users with similar profiles are not exploited at all by content-based recommender algorithms. On the other hand user-user interaction is the core idea behind the second family of RSs that is presented in the next section.

3.4.4. Collaborative filtering

The term "collaborative filtering" (CF) was probably introduced for the first time by the creators of one of the first RSs [118], Tapestry. Since then

CF has been used to denote a family of systems where all the different users collectively collaborate to enable the system to provide effective recommendations for everyone. Indeed, instead of using features of items to determine their similarity, the CF approach focuses on the similarity between different users. The recommendation process then consists in identifying users that are similar to the current active user u_a and recommending items they liked, assuming that there is a good chance they will also be liked by u_a.

CF systems use the user rating data (i.e. $y(i,j)$ contained in the utility matrix) to calculate the similarity between users and do not rely on item content analysis at all. This allows to overcome some of the limitations of content-based approaches allowing, for instance, items for which the content is not available or difficult to obtain to still be recommended through the feedback of other users. Furthermore, collaborative recommendations are based on the quality of items as evaluated by peers, instead of relying on content that may be a bad indicator of quality [116] (e.g. length of a movie may not be a good predictor of its quality).

As for content-based systems also CF can be implemented following either a memory-based or a model-based approach. Memory-based CF systems are better known as neighborhood-based. One of the most common techniques they use is indeed K-nearest-neighbors (or another clustering algorithm) to find the subset of most similar users according to their rating patterns on the common item catalogue. On the other hand several models like Bayesian networks, neural networks or their variants like Boltzmann machines, and dimensionality reduction techniques like singular value decomposition have been successfully applied to build model-based CF systems [119]. Despite the fact that state of the art model-based systems have proven to be very accurate in predicting ratings[81], neighborhood-based still enjoy popularity due to their efficiency (i.e. no costly model training and periodic update has to be performed), simplicity and stability [120]. Moreover neighborhood-based systems tend to perform particularly well on specific task like recommending items that are relevant to the user and that he might not have discovered otherwise (this is also known as *serendipity*). Model-based approaches on the other hand characterize preferences of a user using a less extensive set of latent factors, which makes it harder for them to perform well on serendipity. Findings from the popular Netflix prize competition on RSs [121] also confirm that neighborhood-based and model-based methods can explore very different levels of data patterns. Thus a combination of the two approaches is often used to achieve optimal performance in the most advanced implementations.

3.4.5. Hybrid approaches

We reported in the previous section how the combination of model-based and memory-based CF can be a good compromise to take advantage of both approaches when complexity of the final implementation is not an issue. This is also true for the duality of content-based and collaborative filtering systems. In fact the two approaches are not mutually exclusive and a combined hybrid system can be implemented exploiting both item content analysis and similarity among users.

Furtherly elaborating on the previous example of movie ratings, a hybrid collaborative filtering algorithm can be implemented with a slight modification of the linear regression model presented in the previous sections 3.4.2 and 3.4.3. In particular in the content-based approach we learned the parameter vectors $\vartheta^{(j)}$ representing the user profiles minimizing the following mean squared error cost function:

$$\min_{\vartheta^{(j)}} \frac{1}{2} \sum_{i:r(i,j)=1} \left((\vartheta^{(j)})^T x^{(i)} - y(i,j) \right)^2 + \frac{\lambda}{2} \sum_{k=1}^{n} \left(\vartheta_k^{(j)} \right)^2$$

Adding a collaborative filtering step to this algorithm allows us to do a similar learning process on the ratings data to improve the feature vectors $x^{(i)}$ as well. In fact, following the basic idea of CF, using the ratings given by other users we are able to improve the values of $x^{(i)}$ which represents the content of movie *i*. In turn these values of $x^{(i)}$ are used to recommend the item to potential new users, that thus benefit from the feedbacks of their peers, enacting the "collaborative" nature of the RS. In our example implementation formulating the problem as a regression enables us to learn $x^{(i)}$, given the utility matrix of $y(i,j)$ and user profiles $\vartheta^{(j)}$. This can be accomplished minimizing the same cost function with respect to $x^{(i)}$:

$$\min_{\mathrm{x}^{(i)}} \frac{1}{2} \sum_{i:r(i,j)=1} \left((\vartheta^{(j)})^T x^{(i)} - y(i,j) \right)^2 + \frac{\lambda}{2} \sum_{k=1}^{n} \left(\mathrm{x}_k^{(i)} \right)^2$$

Having defined these two concurrent optimization objectives an iterative algorithm can then be implemented to alternatively derive $\vartheta^{(j)}$ from initial values of $x^{(i)}$ and vice versa. This would constitute a basic example of an hybrid algorithm exploiting both item contents and collaboration between the different users.

$$initialize\ x^{(i)} \rightarrow learn\ \vartheta^{(j)} \rightarrow update\ \vartheta^{(j)} \rightarrow learn\ improved\ \mathrm{x}^{(i)} \rightarrow \dots$$

To conclude this overview of RSs it is worth mentioning that some interesting alternatives to the well-established content-based and CF approaches are being explored in recent research works. These systems are also traditionally included in the family of hybrid RS since they usually combine CF with other techniques [117]. Some examples are context-aware RSs where recommendations are selected considering user context like

geographical location or demographics, knowledge-based systems where a-priori knowledge about items and preferences is added to the recommendation logic, and trust-based RSs where items to recommend are selected among those that are popular in the social proximity of the user (e.g. items that the friends of a user liked). Albeit these kind of RSs have reached a good level of popularity in some settings (e.g. trust-based recommendations are a natural fit for a social-networking environment) a detailed description of these kind of systems is beyond the scope of the present dissertation.

Chapter 4

Integrating data and components of a distributed CDSS

4.1. The MobiGuide Personal Guidance System

The topics presented in chapter 2 and the methodologies of chapter 3 have been applied, over the past 4 years, to the development of a personal guidance system named Mobiguide. Chapter 4 will describe the Mobiguide project in general, its main goals and the solutions adopted to achieve technical and functional integration between the system components through a shared data model. Chapter 5 will be more focused on the results of Mobiguide, and other satellite projects, regarding personalization of the DSS and patient-centric personalized guidance.

4.1.1. The project

MobiGuide is a Collaborative Large-scale integrated project, supported under the European Commission Seventh Framework Programme (FP7) for years 2007-2013. The project is carried out by an international consortium of 13 partners including academia, industry, research hospitals and patients' organizations (see Figure 13 for the geographical distribution of the institutions).

Academia:

- University of Haifa (HU), Project Coordinator
- Ben-Gurion University of the Negev (BGU)
- Universita Degli Studi di Pavia (UNIPV)
- Universidad Politecnica de Madrid (UPM)
- Universiteit Twente (UT)

- Technische Universitaet Wien (TUV)

Industry:

- Beacon Tech Ltd (BTL), Administrative Management
- MobiHealth BV (MHBV), Technical Management
- ATOS Spain SA (ATOS)
- ZorgGemak BV (ZORG)

Clinical partners:

- Fondazione Salvatore Maugeri (FSM), Deployment Management
- Corporacio Sanitaria Parc Tauli de Sabadell (CSPT)

Patients' organizations:

- Associacio de Diabetics de Catalunya (ADC)

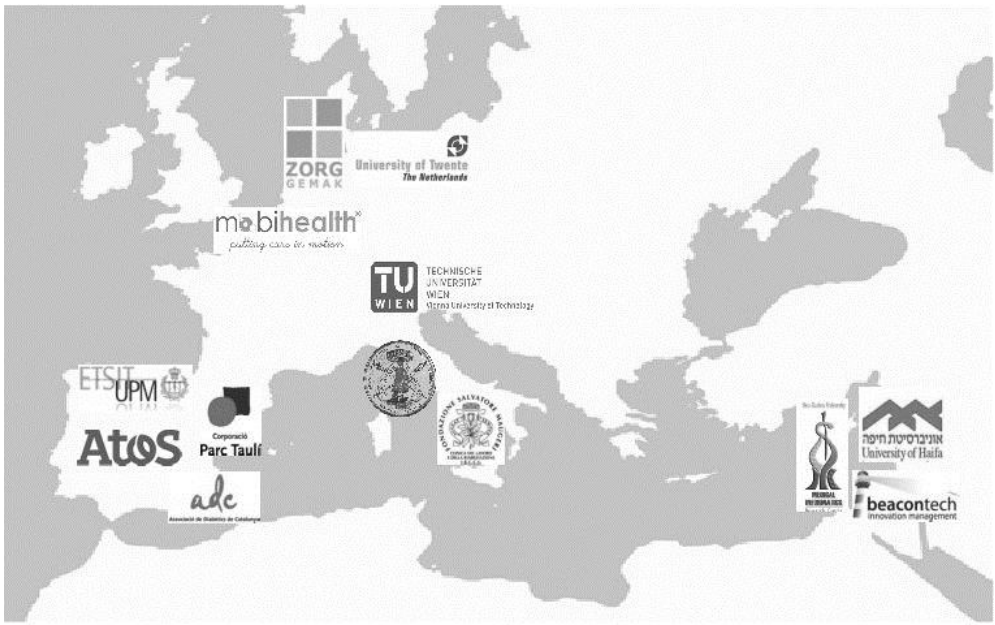


Figure 13 - Geographical distribution of the MobiGuide consortium partners.

The main objective of MobiGuide research is to create a solution for designing, deploying, and maintaining a mobile Patient Guidance System (PGS) that could be scalable, secure, ubiquitously accessible, and user-friendly. Basically, the system supports physicians by delivering them evidence-based recommendations about a patient's treatment according to clinical practice guidelines and personal health record data. Once the physician has defined a treatment plan for the patient, this is translated into a series of instructions for the patients themselves (hence the PGS). By

delivering personalized evidence-based clinical recommendations, MobiGuide aims at increasing patients' satisfaction and compliance to evidence-based clinical guidelines, while also reducing risk to patients and healthcare costs. In particular the CDSS is intended for patients with chronic illnesses. The system accompanies the patients wherever they go and helps them and their care providers in managing their illness, whether they are at home, at work, out and about or travelling abroad on holiday or for business [122]. It is important to stress the fact that the users of MobiGuide, unlike many other CDSSs, consist of both physicians, who can benefit from it thanks to an increased compliance to evidence-based clinical practice guidelines, and patients themselves who can improve their compliance to therapy, have the possibility of being closely monitored from remote through a set of sensors for increased safety and generally benefit from better support for an easier management of their chronic condition. In fact by involving patients in their healthcare they become more motivated and, in turn, motivated patients try to comply with healthcare recommendations more fully, resulting in better health outcomes. What is more, MobiGuide provides personalized decision-support by exploiting patients' personal preferences and their personal context (e.g., being on holiday, family members are temporarily unavailable to help with daily care). Thus, the MobiGuide system delivers recommendations that are more appropriate to each patient, and customizes the treatment to him/her and further facilitates adherence to it [123]. A more focused description of the adaptive and personalized nature of the MobiGuide system will be given in chapter 5. In the following the general architecture of the CDSS and its domains of application will be presented.

4.1.2. Clinical domains of application

MobiGuide has been designed to be general enough to support multiple clinical domains and their relative clinical practice guidelines. However in the scope of the 4-year project the developed PGS has been validated in three important clinical domains: atrial fibrillation and two possible complications of pregnancy (which may co-occur): gestational diabetes, and hypertensive disorders of pregnancy. These domains represent two different monitoring scenarios: chronic patients with possible re-acutizations, that require monitoring all life-long, represented by the cardiac (atrial fibrillation) domain, and monitoring of biosignals and patient's life-style of patients for a limited amount of time, represented by the gestational diabetes domain. Gestational diabetes was especially selected to account for the complex set of patient's context during pregnancy, including the mother and the fetus, with long-term impacts on the born child [124]. Moreover, during pregnancy, other complications, such as hypertension may occur. Hence, we could use these domains to demonstrate the generalization potential of MobiGuide to multiple types of patients.

The three clinical domains chosen are important due to their prevalence and burden on healthcare systems. Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia, occurring in 1–2% of the general population and is estimated to double in the next 50 years as the population ages [125]. Likewise, 2–5% of pregnancies involve women with diabetes, and 87.5% of those pregnancies are estimated to be due to gestational diabetes (GDM). Moreover, it has been well established that the fetuses submitted to hyperglycemic media have an increased risk of developing diabetes in the long term. Thus gestational diabetes is a relevant topic associated with an enormous burden. Hypertensive disorders during pregnancy carry risks for the woman and baby. One-third of severe maternal morbidity is due to hypertensive conditions. The long-term consequences of hypertension during pregnancy include chronic hypertension and an increase in lifetime cardiovascular risk. Hypertensive disorders also carry a risk for the baby as, for example approximately 5% of stillbirths in infants without congenital abnormality occurred in women with hypertension during pregnancy.

4.1.3. Architecture and components

MobiGuide PGS intends to securely interact with the patient and his/her care provides, deliver health-related recommendations and enable access to the patient's health data, whenever a need arises, from any place, and in a user-friendly way using web and Smartphone interfaces. The recommendations delivered will be based on evidence-based clinical practice guidelines (CPGs) that are formally represented in a knowledge base and are executed according to the patient's clinical data. The clinical data will be semantically integrated into a Personal Health Record (PHR) that will persistently store:

- the patient's life-long data from hospital and care centers electronic medical records (EMRs);
- physiological data that will be acquired using body-wearable and portable monitoring devices at non-clinically controlled environments;
- additional personal information that will be acquired through a dedicated Smartphone;
- events (i.e. symptoms or life-style related information)
- temporal abstractions of patient data calculated by the PGS.

The high level architecture of the system just described is represented in Figure 14.

The functionalities of MobiGuide PGS are not carried out by a monolithic system but by a complex ecosystem of interconnected components arranged in a service oriented, distributed architecture (SOA) represented in detail in Figure 15. One of the basic blocks of the system is

the knowledge base. Narrative CPGs are formalized by knowledge engineers together with clinical domain experts [126], represented as Computer Interpretable Guidelines (CIGs) in hybrid-Asbru language [127] and finally stored in the DEGEL digital library by means of its knowledge acquisition tool named Gesher [128, 129]. The concepts and abstractions defined in the knowledge store need to be mapped to the patients' actual clinical data of in order to execute the guideline and provide relevant recommendations. For this purpose two dedicated components complement the knowledge module of the system. These are the Mediator [130], responsible for the constant real-time monitoring of patients' clinical data and computing all the temporal abstraction, and KDOM [131] a Knowledge to Data Mapper based on ontologies. All the clinical data needed for the MobiGuide system execution is stored in a centralized PHR which accessed by a dedicated Data Integrator (DI) component. The DI constitutes the centralized point of access for the PHR and enables all the read and write operations on the patients' data through appropriate web services to be used by other MobiGuide components. Also data coming from the EMRs of the hospitals is merged into the PHR by the DI through appropriate periodic and event-driven import procedures. Given the importance of the DI component in the MobiGuide ecosystem and its central role in the implementation of the data integration strategy of the whole project the following section 4.2.3 will discuss these aspects in more detail.

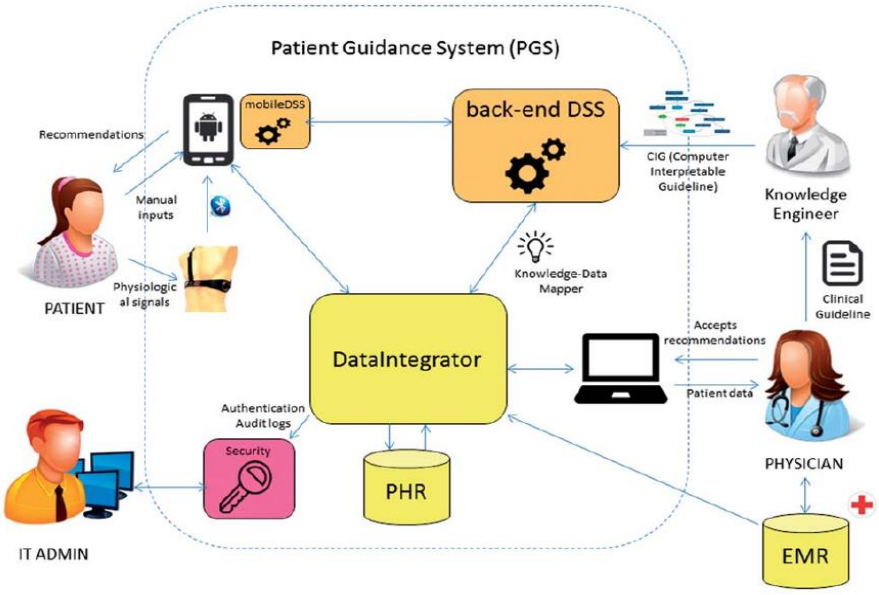


Figure 14 - High-level architecture of the MobiGuide PGS [132].

Formalized domain knowledge and patients' clinical data together constitute the inputs needed to apply the guideline and provide case-specific recommendations. The distributed DSS component consists of two main blocks: i) the back-end DSS (BE-DSS), a guideline execution engine

based on Picard framework [45] and ii) the mobile DSS (mDSS), which runs directly on each patient's smartphone and incorporates the quality of data broker (QoD broker) [133] for constant monitoring of the quality of evidence provided by patients' devices. A more detailed discussion of the distributed nature of the MobiGuide DSS will be presented in section 0. Part of the MobiGuide system runs directly on the patients' mobile phones. Mobile components are orchestrated by a BAN frontend system which dispatches messages among them or interacts with the back-end parts of the system when needed. Other mobile components include the mDSS and QoD broker already mentioned, the set of sensors that are specific to the domain of application (e.g. Activity Intensity Detector and glucometer for GDM or ECG monitor for AF) and the patients' user interfacing app. Other interfaces are offered to the various users of the system: i) a web-based caregiver interface and an integrated advanced data visualization module for physicians, ii) a user interface for system administrators and iii) a knowledge acquisition tool (the aforementioned Gesher) and a GUI for the configuration of the KDOM mapping component for the knowledge engineers. Finally the last important set of components for the MobiGuide architecture is the one comprising security and privacy. All the accesses to the MobiGuide PGS are managed by a single sign-on system implemented in the Identity Provider (IdP) using the openAM framework [134]. A specific ad-hoc component encrypts and secures data in the local storage of the mobile device and grants authentication and authorization to access the data to the mobile components. Patient data confidentiality and anonymization is granted by the Personal Identification Information (PII) hub that securely stores all the internal identification codes for the patient (e.g. EMR-ids, BAN-ids) and maps them to a unique, anonymous MobiGuide ID which is the only patient identifying information used across the system.

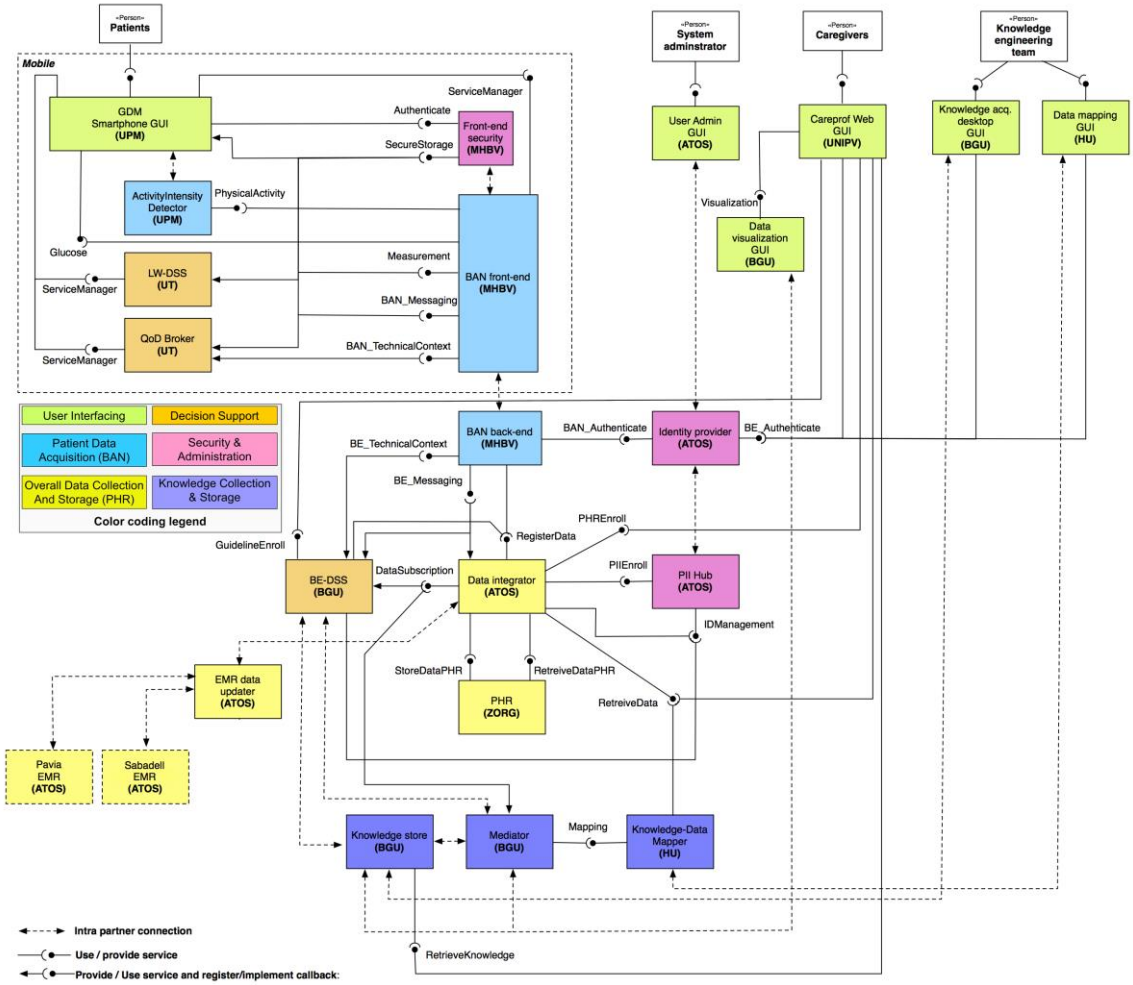


Figure 15 - Detailed architecture of the MobiGuide system highlighting service-based interactions between components. The responsible partner for each component is shown in brackets.

4.2. Data model implementation results

Recent work in the area of medical informatics and clinical data standards point out that integrated PHR models have true transformative potential to strengthen consumers' ability to manage their own health care [135]. The authors state that integrated PHRs improve the quality, completeness, depth, and accessibility of health information provided by patients; enable easier communication between patients and providers; provide access to health knowledge for patients; ensure portability of medical records and other personal health information; and incorporate auto-population of content.

In the context of the MobiGuide project two components, acting as a single entity, play a central role for data storage and components semantic integration across the entire PGS: the DI and the PHR. The following

section describes them along with the implemented solution for the MobiGuide data model based on HL7 vMR model.

4.2.1. Data Modelling with HL7 vMR

Being a PGS where both patients and physicians are active users, the sources of data that the MobiGuide system has to deal with are highly heterogeneous. the amount and variability of the data that must be integrated for such purposes, already daunting even for a single CIG and implementing organization, becomes immense when thesystem is scaled to several organizations and several CPGs. The PHR has to include BAN data, clinical data coming from hospital EMRs, recommendations delivered to the patients by the DSS, abstractions or patterns found in the data by the temporal reasoning components and responses provided to DSS recommendations by users (patients and physicians). By integrating patient data into a PHR, MobiGuide aims to have access to more dynamic information than the hospitals' EMRs usually include, thus being closer to provide patient-centric decision support. This patient data can be related to several aspects of the evaluation of the patient condition, considering both inputs and outputs of the MobiGuide system: the clinical history of the patient, his/her socio-demographic aspects (e.g., environment, habits or family support), the information coming from different medical sensors in order to monitor and evaluate the actual patient condition (e.g., blood glucose, physical activity monitoring), specific knowledge abstractions derived from inference processes made by the reasoning components, or guideline-based recommendations and instructions provided as output by the system [136]. What is more the continuous need to assess the patient's current condition from BAN signal data, supplemented by data that is proactively reported by patients (e.g., patient with GDM reporting eating extra carbohydrates) radically changes the traditional workflow where interaction with the patients is usually limited to periodic face to face encounters with the doctor. The system may also require users to take actions and provide feedback during the automatic guidance process and reacts accordingly. This is true both for patients (e.g. confirmation for drugs intake) and for physicians when confirmation from a human expert is required before taking an action (e.g. suggestions to change diet or therapy dosage should always be confirmed by the doctor). Furthermore given the distributed architecture of MobiGuide presented in the previous section the potential of having patients' information scattered throughout several information systems or devices calls for the use of a PHR which not only stores clinical data but also provides an effective mechanism for semantic integration between the several components of the system and for their communication. The type of PHR developed is known as integrated or interconnected PHR [135], since the data imported may be generated in different hospitals, medical devices, etc., and where the patient, the physicians and possibly other roles likes nurses or patient relatives

supporting the care process are allowed to enter information into selected areas of the record. The data stored in the PHR should later be viewed, searched, and analyzed (e.g., for compliance check, for finding temporal patterns for data analysis) by clinical staff and researchers. Therefore, the data should be provided in a way that is easily understandable for all these stakeholders. In such scenario, not only the representation of data is relevant but also the interfaces provided for external systems to access and exchange data.

Given these requirements the data model for the PHR was designed following two basic needs [136]:

- integrate and represent patient information from different sources.
- facilitate the integration of patient data with a guideline-based CDSS considering the different stakeholders involved in the process of designing and setting up the system.

Following the recommendations of recent work published in the literature [137] these high level integration goals can be translated to implementation-level requirements for both the back-end solution for data storage and the front-end interfaces managing the communication between the different component. Both the HL7 vMR and OpenEHR archetypes standards presented in chapter 2 were considered as possible solutions in a preliminary study [136] that assessed the possibility to express representative subset of data needed/produced by the CDSS including: quantitative EMR data. (average heart rate result of 62bmp, measured on 10/07/1015), qualitative EMR data (e.g. family history of myocardial infarction), BAN data. (e.g. heart rate results recorded every second for 5 minutes starting at 8 a.m. on 10/07/2015), abstractions. (e.g. tachycardia, i.e. heart rate > 115 bpm, during the interval of 8:00-8:30 on 10/07/2015) and decision-support outputs (e.g. a recommendation to increase the frequency of blood pressure measurements from twice a week to once a day).

The results of the preliminary assessment suggested that both the vMR and OpenEHR standards had a good semantic coverage for the test set of data types considered. However the vMR standard is the one that provides the best support. This derives from the fact that it has been natively designed for clinical decision support, and its conceptual model is very similar to what physicians are used to (e.g. observations, problems, procedures, or clinical assessments and recommendations for care plans). Another important factor to consider is that knowledge engineers and database administrators can understand the vMR model quite straightforwardly, since it encompasses a small set of classes with attributes clearly defined in HL7 documentation (see chapter 3 for a description of the classes included in the standard), and where all types of patient data are instances of these classes; this user-friendliness should also ease the process of connecting the hospital information systems with

MobiGuide. Note that this differs from using archetypes created from scratch for a specific purpose, for which there is no predefined structure, hence each data item could be defined differently in each specific implementation project or, even in the same project, in different implementation sites. Another advantage of vMR over OpenEHR, regarding the integration of MobiGuide with the pre-existing EMR systems of the hospitals, is that the HL7 working group in charge of the standard is developing implementation guidelines for transforming HL7 v2.x messages to and from the vMR model. This is very relevant, since it is a known fact that most hospitals are already able to export messages in the HL7 v.2 format. This would simplify tremendously the process of exporting patient data from new hospitals that want to use MobiGuide to the vMR service model.

Despite HL7 vMR has been selected for the front-end implementation of the data model, some requirements about the back-end solution for data modeling and storage still had to be properly addressed. The best solution found consists in using openEHR archetypes in the backend for representing the vMR classes of the front-end model. On one side archetypes provide a high flexibility for possible adaptations of the vMR that might be needed (e.g., complex data like ECG biosignals are difficult to represent with the vMR standard alone) and, on the other side, the conceptual linking between the guideline concepts and a vMR-based PHR is possible and more comprehensible for all the stakeholders than using either HL7 CDA or concept-specific archetypes as the ones found traditionally in the openEHR Clinical Knowledge Manager. Furthermore, using an archetype-based representation of the vMR standard also helps to keep compliance with the ISO/CEN 13606 norm, regarding the two-level modeling approach that is one of the foundations of the OpenEHR archetype-based approach. Adopting a two level modeling approach also allows to choose the storage implementation-level solution to be selected independently of the high-level archetypes representation. One of the industrial project partners (Zorg) provides an openEHR-based middleware that also has the advantage, when combined with the MobiGuide security components like the IPD and PII hub, to be fully compliant to the security guidelines suggested by the ISO/CEN 13606 norm. Those include full access log tracing, possibility to provide an awareness service to system users that want to check how their data was used and by who, complete role based access control and authorization mechanisms for use of specific portions of the PHR and functionalities of the system. Regarding scalability of the data storage solution adequate performance is guaranteed by the use of a NoSQL database as low-level data infrastructure (more details about the DI component and the technical solution adopted for its implementation are given in the following section 4.2.3).

Based on the previous discussion, augmented by further feasibility assessments [132, 136, 137] carried out during the formalization of the two CIGs considered in the MobiGuide pilot implementation (GDM and AF guidelines) the final solution consists in two main decisions. First, the HL7

vMR structure is ideal to address the different front-end implementation needs. Second, using openEHR archetypes designed following the structure of the HL7 vMR is ideal for addressing the back-end needs and the back-end/front-end communication while also being portable to different domains. Figure 16 represents an overview of the overall implemented solution. Looking at the front-end side, the conceptual mapping between the CIGs Knowledge Base used by the DSS and the PHRIs made possible by developing openEHR archetypes, designed to comply with the structure of the HL7 vMR classes, specially designed for the goal of supporting CDSS. On the back-end side archetypes can be easily connected to any medical vocabulary needed. Furthermore, an openEHR infrastructure can provide a powerful dedicated query language (AQL), where data values can be retrieved to feed and support DSS in a more efficient way than using XML-based query languages, making it easier to build more complex queries.

Finally it is worth mentioning that the alignment of different norms like HL7, openEHR, and EN13606 is an interesting objective of data integration research in general. In this scenario the proposal of using openEHR on the back-end, (using an ISO/CEN 13606 compliant component), and the HL7 vMR on the side of the EMRs' interfaces and for inter-component communications, could be a first step towards demonstrating how all three initiatives could be integrated to pursue a common direction for interoperability for CDSSs

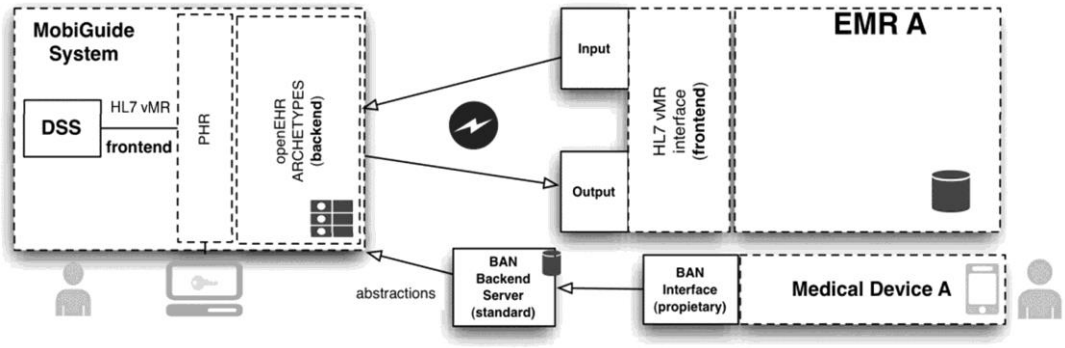


Figure 16 - High level view of the data integration solution.

4.2.2. Supporting CDSS distributed workflow

The reasons behind the choice of HL7 vMR as the standard data model for MobiGuide PHR have been analyzed in the previous section 4.2.1. This section will present a more detailed example (adapted from [138]) of how the vMR classes were used in the context of the GDM guideline and how, with proper adaptations and extensions, the model can support complex asynchronous communication among the different components of the distributed guidance system. The example use case shown in Figure 17

consists in a subsection of the GDM guideline where an interaction with multiple stakeholders is needed to correct a wrong behavior of the patient regarding diet. The corresponding interactions and communications between components needed to support this workflow are represented in Figure 18 in the form of a UML sequence diagram. For the sake of clarity the full complexity of the system has been reduced by merging some of the basic components in higher-level macro components of the sequence diagram. Thus the Patient-GUI also incorporates the mDss and other mobile components like QoD Broker; the Extended DSS includes the guideline execution engine, the temporal abstractions mediator, KDOM data-mapper and the knowledge base; while the PHR stands for both the PHR itself and the data integrator.

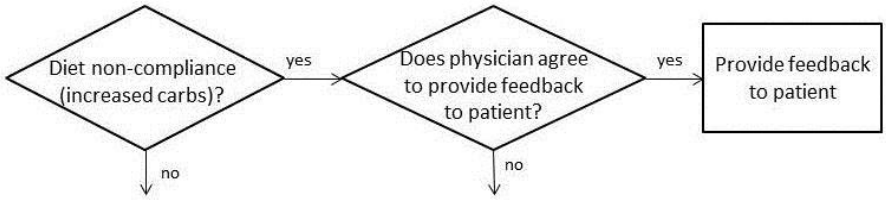


Figure 17 - Example workflow taken from a subsection of the GDM guideline regarding diet compliance.

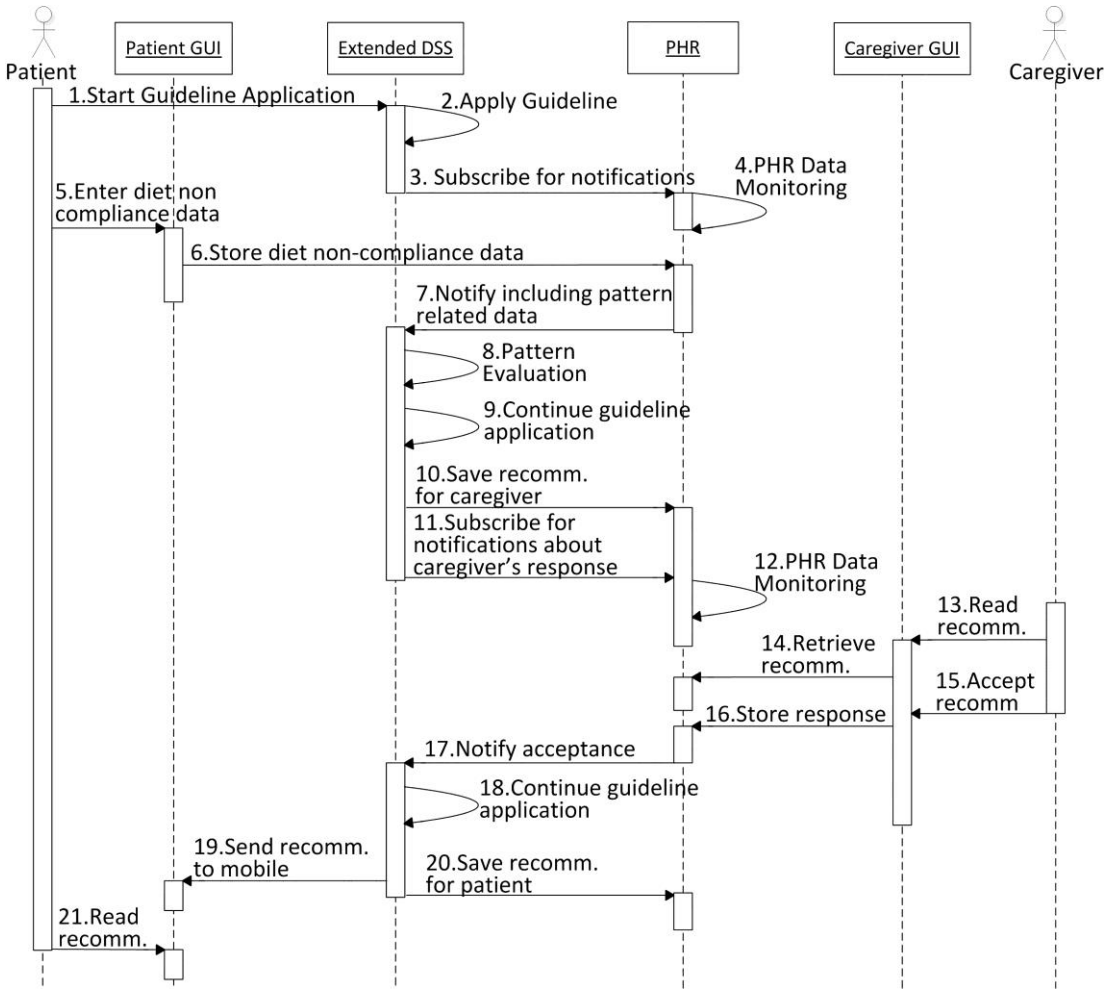


Figure 18 - Sequence diagram of the interactions between components [138].

At the time of the enrollment in the system the Extended DSS starts the guideline application and subscribes through the DI to the set of relevant patient's data items that will need to be monitored from that moment on (Figure 18, steps 1 to 4). The condition that our example considers is a pattern of two diet non compliances in a period of one week. At each insertion of a new diet non-compliance event (step 5, 6) by the patient the DI notifies the extended DSS which, in turn, calculates whether the available data meet the conditions to trigger a pattern of "diet non-compliance twice in the last week" (steps 7-9). The instructions formalized in the CIG are then executed and a new recommendation for the caregiver asking "two non-compliance to the prescribed diet have been detected in one week. Do you want to send the patient a reinforcement message about the importance of following diet prescriptions?" is stored to the PHR and the DSS subscribes to wait for and answer (steps 10, 11). When the doctor logs-in to the caregiver interface he reads the received recommendation and has the possibility of accepting/denying the proposed action (steps 13-15). Upon acceptance a new data item is stored to the PHR and the DSS will

perform the actual generation of the final recommendation “remember it’s important to follow your diet prescriptions” and simultaneously send it to the patient GUI and store it to the PHR (steps 17-21).

The interactions between the different components just presented are made possible by the vMR model via the Proposal-Order-Event pattern and using the PHR as a centralized asynchronous communication enabler. Different classes like ProcedureProposal, ProcedureOrder and ProcedureEvent are generated and saved along the execution of the workflow to keep track of its status of execution. Table 3 summarizes the vMR objects used in the example use-case and relates them to the specific moment of the sequence in Figure 18 in which they are generated. The same Proposal-Order-Event pattern can be applied to support workflows regarding substance administrations, encounters, or observations. Other simpler patters like Proposal-Result can be used in case no confirmation from a physician is required: for example, when the DSS recommends that the patient should measure her fasting blood glucose level at a certain time the DSS issues an observationProposal waits for an observationResult from the patient containing the actual measurement without requiring an "order" step [138].

Table 3 - Use of vMR classes in the GDM example scenario

Seq #	Use Case	Use of vMR in the Scenario
3	The extended DSS monitors for patient non-compliance to diet	**ObservationResult** focus: Increased carbohydrates (123995008) value:+/++ GL_ID (extension):201
6	Non-compliance to diet stored in PHR	
10	The extended DSS stores recommendation to the physician to provide feedback message to his patient	**ProcedureProposal** procedureCode: Notification (C0422202) target: Physician (C0031831) originalText:“The patient didn’t follow compliancy recently. Consider sending him/her the following recommendation message: Remember that it is very important that you comply to diet recommendations and blood glucose measurements schedule” DSS_ID (extension): 111 GL_ID (extension):201

11	The extended DSS monitors for physician's acceptance	**ProcedureOrder** procedureCode: Notification (C0422202) target: Patient (C0030705) originalText: "Remember that it is very important that you comply to diet recommendations and blood glucose measurements schedule" DSS_ID (extension):111 GL_ID (extension):201
16	Physician stores in the PHR his agreement to provide feedback to the patient	
20	Extended DSS stores the message that has just been delivered to the patient on his Smartphone	**ProcedureEvent** procedureCode: Notification (C0422202) target: Patient (C0030705) originalText:"Remember that it is very important that you comply to diet recommendations and blood glucose measurements schedule" DSS_ID (extension):112 GL_ID (extension):201

However, despite the fact that vMR provides a good out-of-the-box fit for most of the data items needed by most CDSSs, some adaptations were needed to be able to represent the whole set of information considered by the MobiGuide PGS. One first example coming from the GDM example workflow discussed above is the importance of maintaining a link between the original recommendation generated by the DSS, its acceptance and the following generated data items. One of the main advantages of this is that it allows the data notification system of the DI to be triggered as soon as the caregiver's acceptance enters the PHR. This avoids confusion with other similar previous recommendation records that could exist in the PHR, improving performance too. This linkage is not part of the vMR model but the standard (since its release 2) includes a native way to be extended with attributes of any possible HL7 data type. Hence, the linkage is stored by tagging the recommendation with proprietary IDs that the extended DSS internally uses to identify recommendations and the guidelines from which they originated (see the DSS_IDs and GL_ID in Table 3). The DSS_ID is then used by the interacting components when saving their reaction in the PHR. For example, in the use case shown above, the caregiver GUI extracts the DSS_ID from the ProcedureProposal and uses it in the subsequent ProcedureOrder instance that is saved upon acceptance of the doctor (step 16 in Table 3 and Figure 18). The extension mechanism of vMR was also

used to cope with a set of other requirements which are summarized in Table 4 [132].

Table 4 - Extensions to the vMR model.

vMR Class Affected	Description	Reason
None	Added codedName ValuePair to vMR datatypes	Added to make the XML more compact, thus providing extension mechanism to ClinicalStatement and EntityBase classes, and to ensure future compatibility with the latest HL7 vMR release, which already supports an "attribute" extension mechanism.
EntityBase	Used extension mechanism with references to previously defined CodedName ValuePair	Enables inclusion of MobiGuide specific attributes like enrollmentDate, and qualityOfData attributes.
ClinicalStatement	Added the TransactionTime attribute	Converts our vMR-based storage solution into a temporal database, thereby improving simplicity and performance
ClinicalStatement	Used extension mechanism with references to previously defined CodedName ValuePair	Enables inclusion of PGS additional attributes like GuidelineID and GuidelideStepID.
ObservationOrder	Added the observationSig attribute	Enables representation of patient preferences, also stored in the PHR. Added directly to the schema instead of using RelatedClinicalStatement or RelatedEntity extensions for simplicity.

In some cases were adaptations of the standards were needed, a less strict application of the definitions of vMR classes and their attributes was performed. For example ProcedureEvent was initially thought to be the class used only for clinical procedures (e.g. "cardiac surgery"), while in MobiGuide it is also used to represent new scenarios like the event of sending a recommendation message to the patient mobile, which is

clinically meaningful in the ubiquitous environment of MobiGuide but was not considered in other CDSSs and thus not included in the original vMR standard with a dedicated class. In some other cases data items generated in the context of MobiGuide needed more expressivity than provided by the standard attributes of the vMR classes used (e.g., blood glucose measured after lunch and entered automatically by a specific glucometer that the patient used). In these cases existing vMR attributes were used for capturing parts of the added semantics. For example, in the observationResult class the method of data input (manual vs. automatic) was captured by the "method" attribute while the glucometer type was captured by the "datasource_type" attribute. Finally the "focus" attribute represented the detailed type of blood glucose (BG after lunch) using the post-coordination of SNOMED codes [139]. In other cases, we used an alternative extension mechanisms provided by the relatedClinicalStatement and relatedEntity classes of the standard. For example, to record the fact that a reminder to measure blood glucose was issued by the DSS and delivered directly to the patient, we used an instance of ObservationProposal whose dataSourceType value was "DSS" while to express the target of the recommendation (the patient) we included a relatedEntity representing the patient.

After all the requirements for data modeling were addressed with the necessary adaptations of the vMR schema the entire set of data items was mapped to the vMR standard and the complete PHR data model was built accordingly. Table 5 summarizes the number of instances of each vMR class that was needed to support the two pilot GDM and AF guideline implementations.

Table 5 - Instances of vMR classes in the two pilot guideline implementations.

vMR Class	Instances in GDM	Instances in AF
ObservationResult	34	61
ObservationProposal	13	19
ObservationOrder	-	4
ProcedureProposal	10	5
ProcedureOrder	9	7
ProcedureEvent	16	37
UndeliveredProcedure	5	5
Problem	5	3
EncounterBase	3	1
SubstanceAdministrationOrder	-	1
AdverseEvent	-	1
ScheduledAppointment	-	1
Total number	95	145

4.2.3. The Data Integrator

An important component that acts as a wrapper around the PHR is the Data Integrator (DI). The DI can be used as gateway between data sources and PGS components, easing their interoperability. The DI encapsulates the data storage, hiding its complexity from the rest of the components, while at the same time providing an application programming interface (API) suitable for the implementation needs of the components that add new data and data consumers. Specifically, the DI publishes service-oriented interfaces through Web Services. Centralizing all communication through the DI has two main advantages [132]: i) the API provided by the DI can be reused later to integrate new incoming components and ii) improved privacy and security measurements can be easily implemented so patients can know who is accessing their data and why. To guarantee patients' privacy and security, demographic data is stored separately from clinical data at all levels (physical storage, web service calls, etc.) and the data is merged back at the final endpoint of the interaction just before being presented to the user. The vMR schema enables strict separation of demographic and clinical data in separate branches in its XML implementation, which is the one adopted by the DI. The DI also manages the imports from the EMRs of the hospitals to the PHR of MobiGuide while also providing adapters and converters from the proprietary data format of the EMRs to the vMR data model. For the current release of the DI, data from the hospital's system is read-only, mainly due to security policies of the participating healthcare institutions. However, in principle, a two-way connection (reading and insertion) to feed back into the hospitals EMRs the data generated from the MobiGuide system sensors would be feasible in a future release. This would solve two issues that currently exist about i) consistency of the two repositories (PHR and EMR) and ii) consequent double data entry.

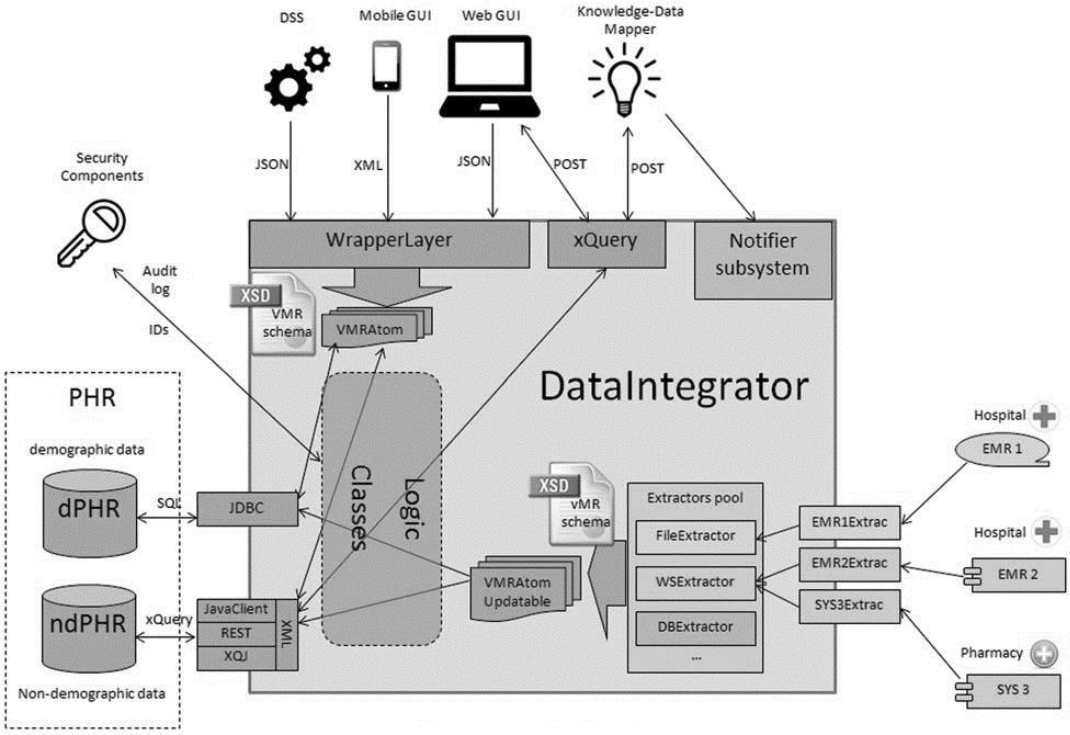


Figure 19 - Detailed architecture of the Mobiguide Data Integrator.

4.3. Distribution of Decision Support

One of the main features of the MobiGuide system is its distributed nature. Section 4.1.3 gave an overview of the several different components of the system and how they interact in a SOA to provide the monitoring and guidance functionalities to the users. This section gives a deeper insight into the distribution of the proper decision support components. In fact the MobiGuide DSS is distributed through a back-end decision-support system (BE-DSS) and the mobile decision-support system (mDSS). The BE-DSS provides full decision-support based on all patient data available at the PHR and on the full CIG representation, formalized and available in the knowledge base. On the other hand the mDSS has access to a limited set of relevant CIG knowledge and to a limited set of data: biosignals arriving from the BAN sensors, the patient's preferences and daily schedule (e.g., meal times, reminder times), and data entered by the patient him/herself.
The distribution of chunks of knowledge from the full-fledged BE-DSS to the lightweight mDSS happens through the proprietary methodology of *knowledge projections*. Each portion of the guideline which can be identified as a self-contained executable knowledge package to be potentially projected and applied in the mDSS, is called a projection unit. Each of these is tagged as a projection point at guideline acquisition time by the knowledge engineer (using a special flag of the extended hybrid-asbru language which is used to formalize MobiGuide CIGs). Only parts of

the guideline that are applicable to the current state of the patient are projected. Hence, the projected knowledge includes implicitly also knowledge about the current state of the patient. The main challenge in designing the projection process during the guideline specification time is to choose at which level (mDSS or BE-DSS) the plans and decisions should be executed, and which patterns should trigger callbacks to the central server. In fact when analyzing requirements for the projection-points in the guideline, consideration should be given not only to technical analysis methodologies, but also to clinical properties, such as whether the decision could be taken by the patient alone, supported by the mobile device or whether it should be taken by the physician using the patient's full historical record and the complete guideline knowledge [140].

A list of principles has been defined to guide the decision about whether or not to project a specific subsection of the guideline to the mDSS:

- actor of the decision (patient or physician)
- temporal horizon and impact of future decisions
- data and knowledge resources needed for the decision
- need for PHR access
- need for a potential personalization of the guideline knowledge

As an example of the application of these principles consider a process of monitoring each patient's high-frequency ECG sensor signals by means of patient-worn biosensors. In the MobiGuide system such a sensor is linked via bluetooth to the mobile device, and the detection of a pattern of Atrial Fibrillation (AF) from the biosignal is delegated to the mDSS. Afterwards this abstraction pattern of "AF episode" is sent to the central PHR to support further guideline-based reasoning and generation of recommendation to the patient and/or physician by the BE-DSS. The advantages of such a distribution of labor, in which the AF detection for each patient is done by the mDSS is rather evident, and prevents an overburdening of the central BE-DSS server. Furthermore, the local mDSS is also essential for continuity of care when for some reason there is no internet connection to the central DSS. In this case MobiGuide would still be able to provide the patient with alerts relevant to latest projections received by the mDSS. However, not all decisions can be taken locally by the mobile components alone. Apart from computational power and data storage limits (which are less and less relevant due to the latest advancements of Smartphones technologies) some decisions still require access to the full historical (longitudinal) patient data, which are only stored in the centralized PHR and should not reside on the mobile device [140]. In some other cases the decision to be made is a part of a long-term plan in the complete clinical guideline knowledge. Such knowledge resides only on the central knowledge-base server. In these cases, the mDSS sends a callback message to the BE-DSS, asking for further instructions, resulting in a BE-DSS recommendation, and in most cases in new projections of updated guideline knowledge.

It is also important to stress the fact that the willingness to distribute subsections of the guideline to the mobile components also affects the way the guideline elicitation process is performed [141]. MobiGuide PGS requirements helped to improve existing guideline formalization methodologies considering also some additional steps like: i) identifying guideline recommendations that will require patients to take actions (e.g., take measurement, take drug), thus impacting patients' daily-life behavior, ii) eliciting from the medical experts the corresponding set of personalized operationalized advices that are not explicitly written in the guide-line (patient-tailored workflow patterns) and iii) delivering this advices to patients. The MobiGuide improved knowledge elicitation process is described in Figure 20.

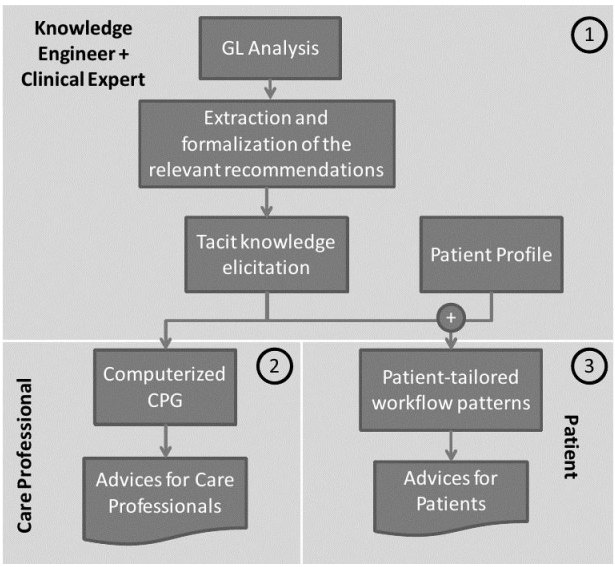


Figure 20 - Methodology followed for guideline elicitation and patient-tailored workflow patterns identification.

The first phases (box 1 in the figure) are the traditional CPG to CIG transformation steps which require the knowledge engineer (KE) to work with clinical experts on the analysis of the guideline and extraction of relevant recommendations. The following steps (boxes 2 and 3) require an even closer interaction with clinical experts to elicit tacit knowledge and to create a patient profile template. The elicitation process splits into two parallel workflows: the first (2) is a more traditional workflow directed at the care professional while the second is a parallel part of the process that focuses on the patients' behavior and their interaction with the MobiGuide system. For example, a traditional guideline regarding GDM may define a plan for monitoring the patient's compliance to diet as a set of instructions to a nurse to check whether the patient reported in her diary of eating too many carbohydrates more than twice during the past two weeks (non-compliance). In the MobiGuide parallel workflow methodology this

recommendation is translated to an automatic evaluation of the patient's non-compliance condition every two weeks, by retrieving data from the patient's digital log book, and delivering an alert to the patient through the Smartphone in case of non-compliance. This part of the parallel workflow whose customer (i.e. the user receiving recommendations) is the patient himself is managed by the mDSS, hence the relevant CIG plans are indicated in the CIG as projection points and allow at guideline application time the passing of control to the mDSS, which receives the relevant knowledge in the projection format.

In the case of the CPGs considered in MobiGuide the AF guideline elicitation provided the most significant example of extraction of parallel workflow patterns involving the delivery of reminders and recommendations to the patient. In particular 4 different types of workflows to cope with different situations: i) therapy-related advisors, to help the patient comply with his/her pharmacological treatment; ii) measurements advisors, to remind the patient to take measurements such as heart rate, weight or blood pressure; iii) suggestions for dealing with personal situations that may necessitate modulating the patient's therapy; iv) personalized packages for specific close monitoring and follow-up of patients. In the first type of parallel workflow the DSS extracts a "drug therapy calendar" from the therapy prescription data structure in the PHR for each patient. Every time the physician prescribes a new therapy, the DSS updates the patient's calendar using information about patient's profile (stored in the PHR) and drugs (stored in the knowledge base) and projects the needed knowledge and data to the mDSS that is then able to execute this set of instructions independently from the back-end. The therapy-related instructions include also a non-compliance monitoring feature to control the day-to-day adherence of the patient to therapy. To this end, the patient can indicate whether he/she has taken the medication or not, and will be asked to provide a motivation in the case he does not comply with the prescription. In case the drug code reported in the therapy calendar identifies a medication that does not tolerate non-compliance the BE-DSS will generate a special alert if a patient declares he has not taken the pill, e.g. "The anticoagulation therapy must not be interrupted. If you have specific motivations please call your doctor". Non-compliance cases are then classified according to motivation (drug side effects, drug unavailable, etc.) for rating the severity of the non-compliance itself [141]. The second type of patient workflow regards the periodical evaluation of some measurements. As the MobiGuide project provides the patient with a set of devices able to take those measurements, the management of this procedure is a responsibility of the patient himself. As in the case of drugs, a measurements calendar is be filled in by the BE-DSS and when the time comes to take a specific measurement, the patient will receive an alert generated by the mDSS on his smartphone. Furthermore alerts are also sent back to users (to the patient to solicit the measurement in case of a single non-compliance, to the nurse/physician when a number of consecutive non-compliances occur) when a parameter exceeds some personalized

thresholds reported in the PHR and defined by the physician during a visit. Another important set of recommendations that require the activation of a patient-tailored workflow pattern is the one related to the events that might necessitate modulating the patient's therapy, with particular attention to oral anticoagulant therapy (OAT). One of the most severe consequences of AF is the risk of stroke and, for this reason, the majority of AF patients undergo OAT. On the other hand these patients are also exposed to higher risk of bleeding as an adverse effect of the therapy and some procedures thus require particular caution (e.g. temporary suspension of the OAT). mobiGuide deals with these situations through a personal calendar where the patient to records his events (e.g. an appointment to the dentist) which the DSS is able to recognize among a set of codified events carrying risk of bleeding. In the presence of one of these events, the mDSS will be responsible for sending out a reminder to the patient to talk with his doctor to define the OAT administration regime during the days around that particular event. The last type of patient-tailored workflows that were identified are personalized packages. Care providers can activate these kind of monitoring plans whenever they need to monitor some patient-specific parameters more closely for a limited period (e.g., activating 48h holter ECG anytime the patient undergoes a therapy change). Examples of such situations are new therapy activations or therapy suspension, the onset of new symptoms or the need to follow-up on specific patient's variables. Personalized packages, after being projected, can be executed by the mDSS and help physicians receive more relevant alerts on specific parameters that need close monitoring and immediate attention.

Chapter 5

Applications of CDSS personalization

5.1. Personalization of decision support

In the previous chapter 4 the MobiGuide project and its primary output, namely the MobiGuide patient guidance system, have been introduced. In the current chapter we will focus more on its features regarding personalization of the decision support functionality and the solutions adopted for its realization. As stated before, among the challenging objectives of the project, MobiGuide seeks to provide personalized decision-support to patients with the purpose of delivering recommendations that are more appropriate to each of them, and customize the treatment to further facilitate adherence to it. For these reasons two main action points will be described in the following sections: i) special arrangements aimed at taking into account personal preferences and contexts in the guideline-driven scenario and ii) design and implementation of a shared-decision framework integrated in the MobiGuide DSS.

5.1.1. Context aware guideline support

Guideline-based DSS generate diagnostic and therapeutic recommendations on the basis of a computerized representation of evidence-based established clinical practice guidelines and of the patient's medical record. CPGs are written on the basis of scientific evidence and by definition cannot include personal context and preferences for which evidence does not exist. However a major goal of the MobiGuide project is to empower and engage patients by supporting the personalization of their treatment. Personal context variables that may affect care recommendations include among others the patient's ability to comply with treatment, the ability to maintain routine diet (which may change during travel), daily activities, time required to reach the medical center, support level (from family members

or live-in help), and exercise level. A special 4-steps process for knowledge acquisition and guideline execution has been designed to cope with these requirements regarding personalization. The 4 phases are described in the following.

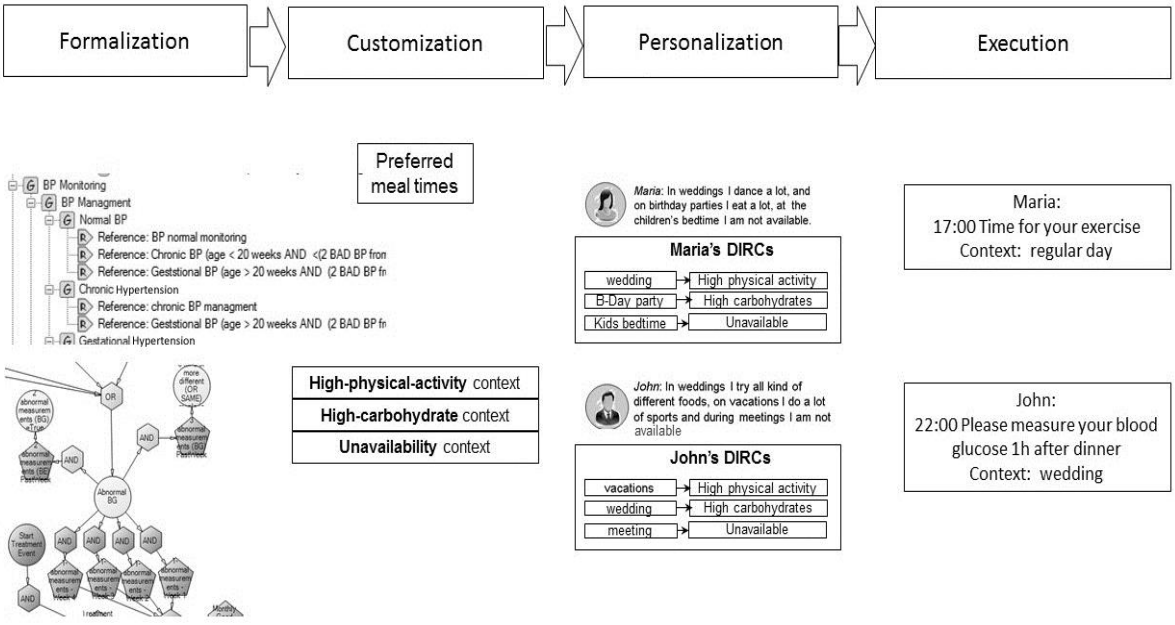


Figure 21 - Customization and personalization of CIGs through the 4-steps process.

- Formalization: This is a process involving knowledge engineers ad clinical experts in which the free-text source clinical practice guideline is represented as a formal CIG, using the Asbru language and the GESHER knowledge-specification module, without any additions. The output of this phase is a formal, executable CIG that should generate exactly the same recommendations of the original textual guideline without further customization.
- Customization: This is a process performed per CIG, again by a knowledge engineer together with a clinical expert. The customization process expands the CIG to include all the different contexts that could affect the CIG that were not taken into consideration in the source version. A first example of guideline customization can be derived from the specific hospital where the guideline is to be implemented. For example a guideline might suggest an MRI scan as the best option for a diagnostic phase. However if the specific hospital implementing the guideline does not own a MRI machine it might use a CT scanner instead. Thus technological or economical resources of the hospital might ask for customization of the, otherwise generic, guideline. In the MobiGuide project however customization is more focused on the effects on the

guideline of the contexts of the patient. These include how the CIG should change when the patient lives alone, or when the patient is in a high-exercise-level or a technological context, such as having no Internet access or experiencing a low mobile-device battery. We call these contexts CIG-customized Contexts. Each CIG-customized context (e.g., 'patient-alone') defines how the CIG changes for any patient that enters into this context. At this point, the CIG is customized for different universally occurring contexts, but is not personalized to any particular patient. The output of this phase is a context-sensitive, customized, but generic (universal) CIG.

- Personalization: This is a process that usually takes place during one of the first encounters of a patient with his or her care provider (no knowledge engineer is involved at this stage). Together, the patient and the care provider define which events or concepts (specific for the patient) might induce any pre-defined CIG-customized context and the patients' preferences regarding their treatment. According to the patient's habits, the care-provider and patient specify events (e.g., patient actions) or concepts (derived from measurable patient data, such as 'high blood pressure'), which are specific to the patient, and which lead to one or more of the predefined CIG-customized contexts. The mapping between events or concepts and their induced contexts, is called a 'Dynamic Induction Relation of a Context' (DIRC). DIRCs are part of the dynamic temporal interpretation contexts theory [142] and allow to define temporal relationships and constraints between the inducer event and the induced context. Figure 22 shows a graphical example of a DIRC where a wedding event (that might be a specific case of a more general event "eating a lot") induces a high carbohydrate intake context for a specific patient.

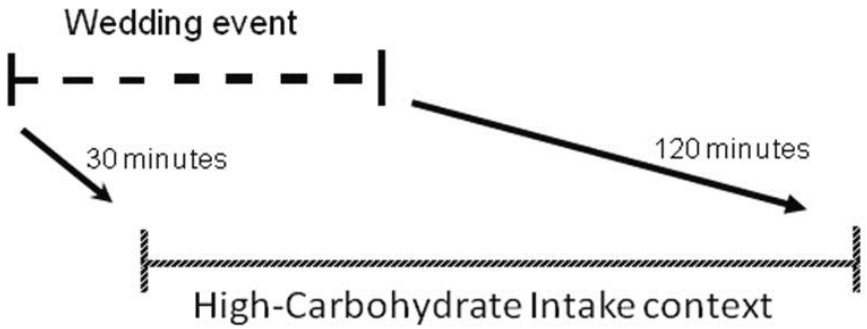


Figure 22 - Example of a dynamic induction relation of a context between a wedding event an a CIG-customized high carbohydrate intake context.

A second part of the personalization step consists in acquiring personal preferences of the patient to personalize the system behavior at specific action points. For example in the GDM scenario a

gestational diabetes patient might usually have breakfast around 7:00 a.m. on a regular week day while at 9:00 am on the weekend. The alert to measure blood glucose before eating should occur 30 minutes before that time, which is context dependent. Meal times preferred by patients should then be acquired and stored to the PHR (through the caregivers' interface) for each of the different contexts available for later use in the guideline application phase.

- Application: This is the process through which, during the CIG-application time, the MobiGuide system loads a series of data that include the customized CIG (which is generic for all patients) from the MobiGuide knowledge base, and the patient-specific DIRCs (used to induce the patient's CIG-customised context) from the personal health record. In addition, the specific patient preferences are also loaded. The system applies a personalized treatment for each specific patient, within any of the predefined CIG-customized contexts, while considering the patient's personal preferences.

5.1.2. Shared decision implementation

One of the challenges faced in MobiGuide is to empower patients by involving them in reasoning about the therapy most appropriate for them, together with their caregivers, within a rational framework like shared decision making. The chosen approach consists in combining a probabilistic approach based on decision analysis with the more comprehensive guidance of the guideline-based system. The intent was to make decision analysis accessible, if not to all, at least to that portion of patients who want to make a more informed choice, that considers their personal preferences [143]. As we pointed out in chap 3 where the shared decision general methodology is introduced, the rationale behind this choice is that when scientific evidence is not strong enough to recommend one option versus another one, the patient's involvement might is a preferable choice to an arbitrary decision by the care provider.

The MobiGuide DSS must be flexible, providing evidence-based recommendations when possible, and offering patients several options whenever more than one is reasonable, eliciting their preferences, and proceeding to provide detailed support to the option indicated by the patient's preferences. This is done explicitly marking, at knowledge acquisition time, the CIG model for the specific points where a shared decision is desirable. Then the task-network model execution is halted by the Picard engine and the control passes to the shared-decision module, to resume after the chosen option has been identified. To provide the needed support the shared decision module should contain 2 main components (Figure 23): i) decision models to represent the decision task and ii) a preferences elicitation tool to gather the information needed to tailor the decision problem to the specific patient. These two components and their actual implementation in the context of the atrial fibrillation guideline will

be described in the following sections of this chapter (5.1.2.2 on decision models and 5.2.1 on the utility coefficient elicitation tool).

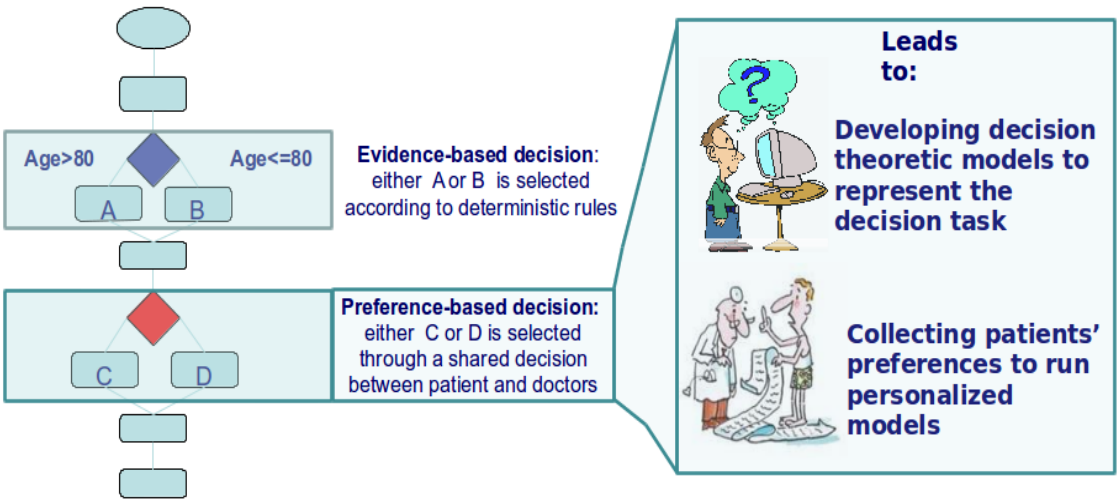


Figure 23 - Even in an ideal evidence-based and guideline-driven decisional environment, there might exist situations in which it is desirable to involve the patient and his/her preferences in the decision.

5.1.2.1. Framework ontology

The general framework for shared decision implemented in MobiGuide takes into account several aspects, starting from the implementation of decision theoretic models relying on CPG recommendations and providing different facilities for eliciting patients' preferences and automatically including them in the models. The ontology depicted in Figure 24 represents how all the involved concepts and methods are combined to enact the SDM scenario [4].

In general, a guideline presents one or more decision points, which are the clinical problems in which a shared decision is suggested or needed (class DecisionProblem). Every shared decision problem has multiple possible solutions (class DecisionOption), which are the options from which the patient and the care professional will select the final decision. The selected decision (class FinalSharedDecision) is an Option and it is chosen from all the available decision options. A decision problem can be represented by a decision model (class DecisionModel). A decision model considers several decision options and includes a set of variables. These are probabilistic variables (class ProbabilisticVariables), which can be either health states (class HealthStateVariable), the results of diagnostic tests (class TestResults) or some relevant patient behaviors (class PatientBehavior). Since, in a decision model, events are generally represented following their temporal sequence, we have included the arc *precedes* to take this into account. An important concept in decision

analysis is quantification. It represents the step in the modeling that allows the variables that are part of the model to be assigned values. Health states can be associated with specific values (class HealthStateValue), which represent the patient's perception of a health condition. These values can be either utility coefficients (class UtilityCoefficient) or some ranking values (class RankingValue). The class ProbabilityValue is used to represent the quantification of the probabilistic values of the model variables. Demographic data, such as age and gender, can be useful in selecting the correct probability value and have been included in the ontology using the class Demographic. Another quantification component, although not exploited in all the shared decision problems, is the cost (class CostValue). In general, costs can be divided into those with an impact on the national healthcare service (class NHSCost), costs that impact on society (class SocietyCosts) and costs that directly impact on the patient (class OutOfPocketCost). This latter class is the most interesting one for our framework, as it can be quantified through the patient's direct participation. The decision model output consists of a set of results (class Result), obtained on the basis of the quantification settings and related to the decision options. Results can be of a varied nature: they can be the expected values for some quantities (class ExpectedValues), they can be indices (class Indices) or they can be the results of additional analyses, such as Monte Carlo simulations (class MonteCarloSimulationResults) or sensitivity analysis (class SensitivityAnalysisResults). The expected values that are calculated by our decision models are: life years (class LifeYears), Quality Adjusted Life Years (QALYs – class QALYs) and costs (class CostValue, which, in this case, can be either a quantification step or a result when it represents the expected value of a decision option). Among the indices, we have selected ICUR (Incremental Cost/Utility Ratio) and ICER (Incremental Cost/Effectiveness Ratio). These two indices are computed starting from the analysis results, as the change in costs divided by the change in benefits. In particular, ICUR is calculated using QALYs and costs and ICER is calculated using costs and life years (see chapter 2 on decision analysis for further details). The individuals taking part in the shared decision process are represented by the class Agent in the ontology. The patient (class Patient) and the care professional (class CareProfessional) work together to take the shared decision. In this representation the care professionals have been divided into physicians (class Physician) and psychologists (class Psychologist), to allow for the presence of a specialist supporting the physician in the process. These two professional figures are considered the most suitable to present the proposed framework to the patients. Although other professionals might be added, the psychologist is the most skilled for supporting the utility elicitation task.

As a matter of fact, it is important to point out that this ontology includes mainly the features that have been identified during the definition of our system. It thus represents only a first step toward the definition of a comprehensive ontology of shared decision making.

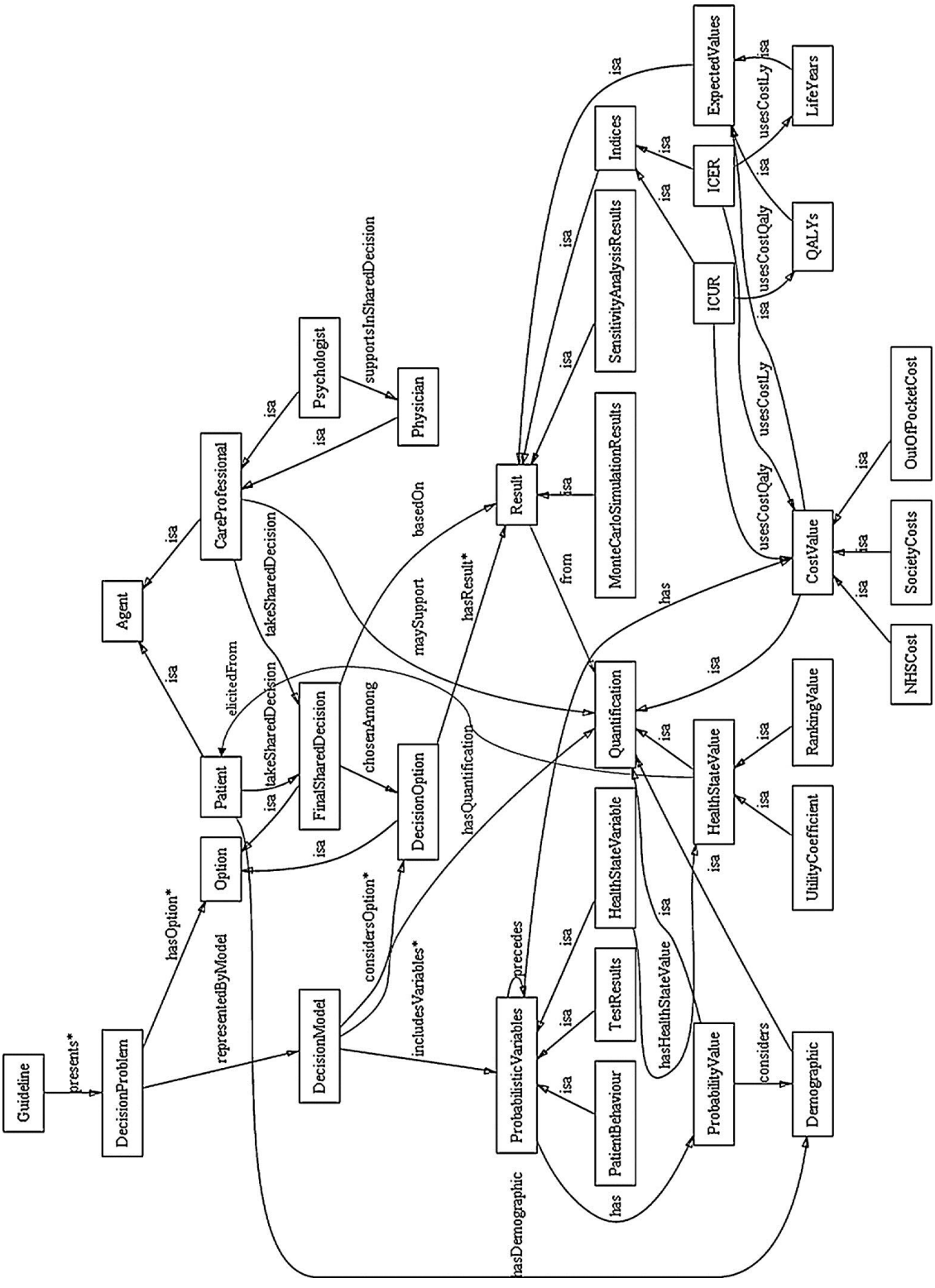


Figure 24 - The ontology of the shared decision framework. Rectangles represent classes of concepts and arrows represent the relationships connecting them.

5.1.2.2. Decision models for the atrial fibrillation guideline

One of the core features of the MobiGuide approach to shared decisions is to use decision trees (DTs) with embedded Markov models as a suitable probabilistic, graphical decision-theoretic formalism for representing and communicating the critical parameters for decisions. In order to support caregivers and patients in shared decision-making as part of a DSS, the knowledge base should contain not only the CIG, formalized as in the hybrid-Asbru language in the case of MobiGuide, but also include a set of decision-theoretic model that support reasoning with patient preferences. Such customized decision models have been developed as decision trees (DTs) starting from the recommendations of the CPG and the related medical literature relevant to the disease states resulting from alternative treatment options. A significant effort was spent in the literature review for the thorough design of the models, since we must consider that each DT embeds variables and states regarding both the current condition of the patient and all the co-morbidities and complications that he could experience in the future.

For the implementation of the models the TreeAge [144] software suite was chosen. TreeAge is a commercial tool that embeds a module completely dedicated to health care (TreeAge Pro Healthcare). This module provides some very useful analysis tools, such as Markov processing, comparative effectiveness and cost-effectiveness analysis. Moreover the availability of the TreeAge Pro Interactive module allows to create web interfaces to make the models available to remote users.

Among the two pilot implementation domains of MobiGuide two decision points where shared decision was desirable have been identified for AF (while none were found for GDM). In particular, 2 different sections of recommendations included in the latest AF management clinical guidelines [145, 146] gave origin to 2 different decision models that will be presented in the following.

Oral anticoagulant therapy decision model

The first model is related to the selection of treatment for preventing thromboembolism in low-risk AF patients. The 2011 version of the AF management guideline [145], on which the project started being developed, stated that:

> *For primary prevention of thromboembolism in patients with non valvular AF who have just 1 of the following validated risk factors, antithrombotic therapy with either aspirin or a vitamin K antagonist is reasonable, based upon an assessment of the risk of bleeding complications, ability to safely sustain adjusted chronic anticoagulation, and patient preferences: age greater than or equal to 75 y (especially in female patients), hypertension, HF, impaired LV function, or diabetes mellitus. (Level of Evidence: A).*

Thus the individual preferences of patients, at least for those at low risk of thromboembolic events, should be considered in the choice between antithrombotic therapy (with either aspirin or a vitamin K antagonist) and no treatment. This lead to the development of a first version of a DT to be used in a shared decision scenario.

However significant updates have been published on the topic of thromboembolism prevention in recent releases of the guideline documents [146, 147]. These caused some revisions of the DT model which final version is shown in Figure 25. The first major improvement is that the model has been modified according to new evidence related to the use of novel oral anticoagulant (NOAC) drugs recently introduced on the market. In particular, the model has been enhanced to add a new decision option, according to the following recommendation [148]:

> *Dabigatran is useful as an alternative to warfarin for the prevention of stroke and systemic thromboembolism in patients with paroxysmal to permanent AF and risk factors for stroke or systemic embolization who do not have a prosthetic heart valve or hemodynamically significant valve disease, severe renal failure (creatinine clearance 15 mL/ min) or advanced liver disease (impaired baseline clotting function) (Class I, Level of Evidence B1).*

The second relevant update of the 2014 release of the AF guidelines aims at identifying low-risk patients with more precision. This is done with a new stroke risk stratification index (the CHA_2DS_2-VASc) that labels very low-risk patients with a score of 1. These patients are subject to a new recommendation that gives more importance to the work carried out so far:

> *In patients with AF, antithrombotic therapy should be individualized based on shared decision making after discussion of the absolute and RRs of stroke and bleeding, and the patient's values and preferences. (Level of Evidence: C)*

The developed DT (Figure 25) models 4 different decision options to compare the clinical pathways of an AF patient who may undergo different treatment strategies for stroke prevention, namely warfarin, dabigatran, aspirin or no antithrombotic therapy at all. A Markov process, also represented in Figure 25, starts at the end of each branch and represents the possible health states the patient can go through:

- AF only (i.e. relatively healthy apart from AF);
- ischemic stroke (IS), that can be temporary, mild, moderate-severe, fatal;
- intracranial hemorrage (ICH), that also can be temporary, mild, moderate-severe, fatal);
- myocardial infarction;

- extracranial bleedings (minor and major);
- death.

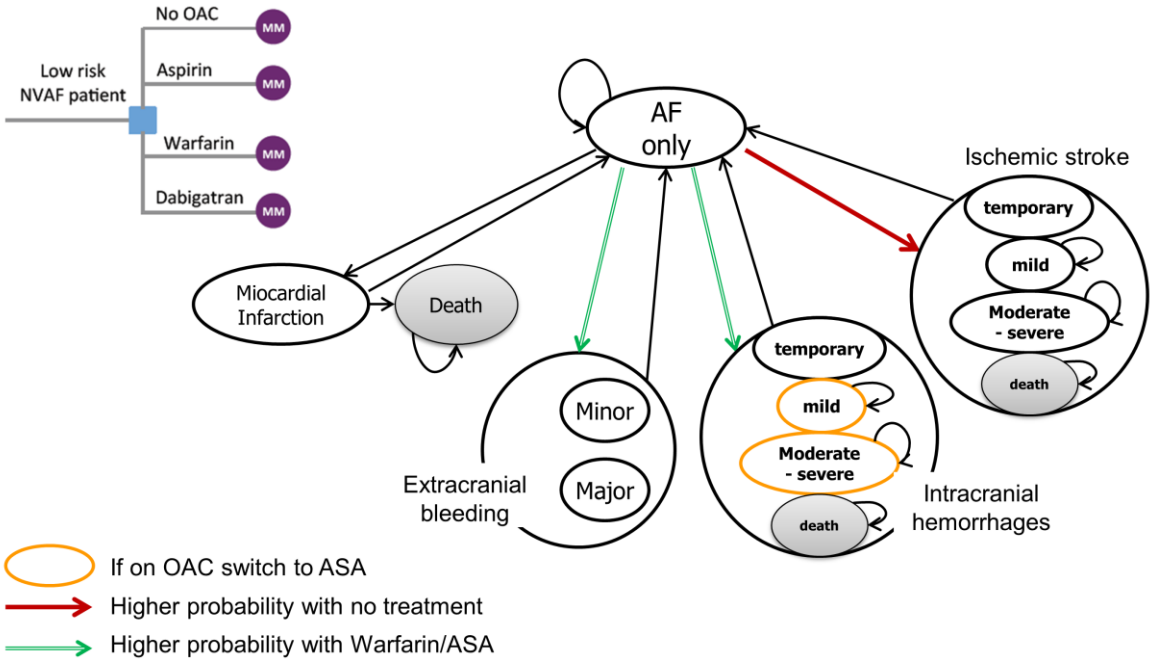


Figure 25 - Decision tree and embedded Markov model for the selection of anticoagulant therapy for low risk AF patients.

During the Markov process, individuals move among health states according to transition probabilities that are defined for each state, and that can vary over time. Running the model over a number of cycles, the expected values of outcomes (i.e. QALYs, costs, etc.) of the process associated with the different strategies can be estimated. In the model, a 3-months Markov cycle length and a lifetime horizon were chosen. An option has also been introduced to discount (e.g. at 3.5% annually) QALYs if required.

Warfarin, dabigatran and aspirin therapies decrease the probability of IS occurrence but increase the probability of ICH and extracranial bleedings. On the other hand, no therapy increases the probability of experiencing an IS but decreases the occurrence of ICH and extracranial bleedings. Temporary IS or ICH are events that cause only a temporary disability and let the patient recover and go back to the AF-only state. A patient experiencing a more severe event, as mild/moderate-severe IS or mild/moderate-severe ICH, acquires a certain level of permanent impairment that results in quality of life worsening. If a patient in the AF-only state is on anticoagulant therapy, and he experiences a mild/moderate-severe ICH or a major extracranial bleeding, it is supposed that he/she interrupts OAT/NOAC therapy that is replaced by aspirin, to decrease the probability of bleedings. An AF patient that has been affected by IS or ICH has 2.6 fold her risk of experiencing new events in comparison with

"healthy" AF patients. In case of MI, if death occurs, it is supposed happening within 3 months from the event, otherwise the patient recovers back to AF-only state. Events like temporary IS, temporary ICH and extracranial bleedings lead to temporary decrements in quality of life (disutilities), that have been estimated from the literature, while the utility coefficients of the states related to the permanent disabilities are elicited from the specific patient (see section 5.2.1 for further detail on the elicitation tool).

Ablation decision model

The second DT model relies on a model presented in the literature [149] and which was further refined on the basis of domain knowledge, with the help of the FSM cardiologists. This DT attempts to model the specific section of the AF management guideline where patients eligible for an ablation procedure need to choose either to undergo such a procedure or stick with pharmacological treatment only. Ablation of the AV node is a surgical procedure able to keep the patient in normal sinus rhythm that, despite good success rates and relatively low risk, often needs to be reiterated after some AF episodes recurrence. The latest AF guidelines [148] report that "when the rate cannot be controlled with pharmacological agents or tachycardia-mediated cardiomyopathy is suspected, catheter-directed ablation of the AV node *may be considered* in patients with AF to control the heart rate." (Level of Evidence: C). A further update dated 2011 states that catheter ablation is useful in maintaining sinus rhythm in selected patients with significantly symptomatic, paroxysmal AF *when performed in experienced centers*. However a footnote further comments on the "experienced center" stating that "evidence-based technical guidelines including operator training and experience necessary to maximize rates of successful catheter ablation are not available" [148]. Ablation is thus a fairly controversial topic for both patients and physicians which, ultimately, may benefit from the application of our shared decision framework.

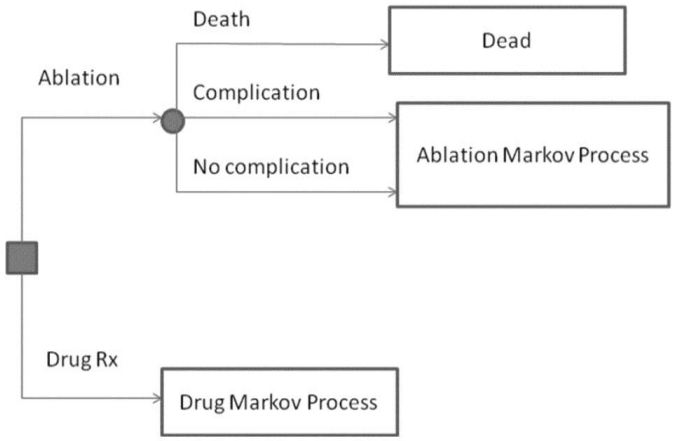


Figure 26 – Simplified view of the decision tree model for the choice between ablation and pharmacological therapy.

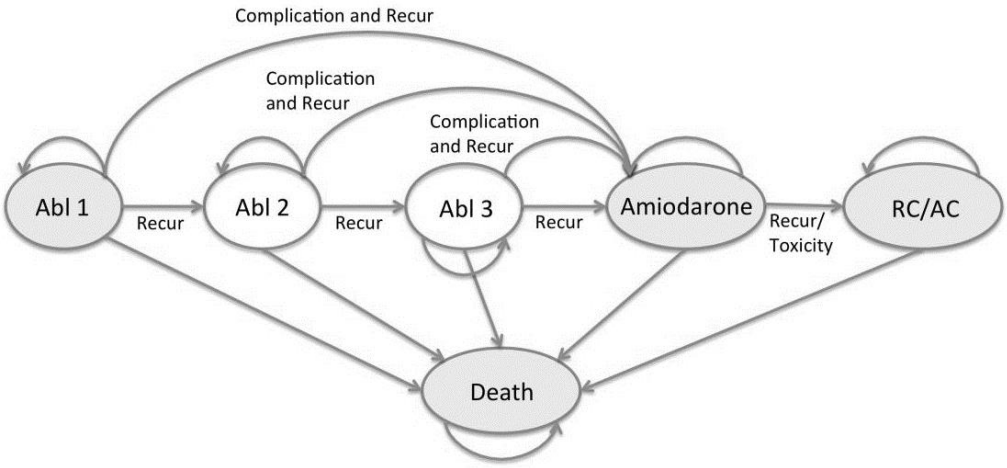


Figure 27 – The Markov model of the process following the ablation procedure.

For the design and implementation of the model a DT, represented in Figure 26, was combined with two Markov models, one for patients completing an ablation, and another for patients undergoing an Antiarrhythmic Drug (AAD) therapy. The DT starts with a decision node, which distinguishes the strategies in comparison (i.e., Ablation and AAD). After the initial decision, continuing on the ablation branch, patients may die as a result of the procedure, have non-fatal complications, or have a normal course without complications. All patients who survive ablation enter the Markov process for the ablation shown in Figure 27.

Such a process assumes that patients will progress stepwise from one therapy to the next, based on whether or not they experience symptomatic AF recurrences on their current treatment. Since there is consensus that recurrences of AF after ablation procedures can be best controlled with re-ablation procedures, those patients showing a recurrence after a first ablation may repeat ablation up to two times. Following guideline recommendations (catheter ablation of the AV node should not be attempted without a prior trial of medication to control the ventricular rate), the process includes the treatment with the previously ineffective guideline-recommended first-line AAD drug, either sotalol or flecainide, for the first two months after ablation. Patients with recurrent AF despite the third ablation proceed to treatment by the Amiodarone AAD drug. Reynolds et al. [149] assume that patients undergoing ablation will not be subject to drug treatment with Amiodarone. After consultation with our clinical partners, and taking into account what is indicated in the guidelines, we decided instead to consider Amiodarone at this stage. Moreover, we considered the possibility of encountering death due to the ablation procedure and its complications also after both the second and the third ablation. It was assumed that only patients

who have undergone ablation without any complications could repeat the procedure. On the other hand, patients who have experienced non-fatal procedural complications incur costs and disutilities in the short term, and in the case of AF recurrence, proceed to treatment with Amiodarone.

Patients failing second-line drug treatment cease further efforts at rhythm control and are treated with pharmacologic rate control. The AAD Markov process is shown in Figure 28.

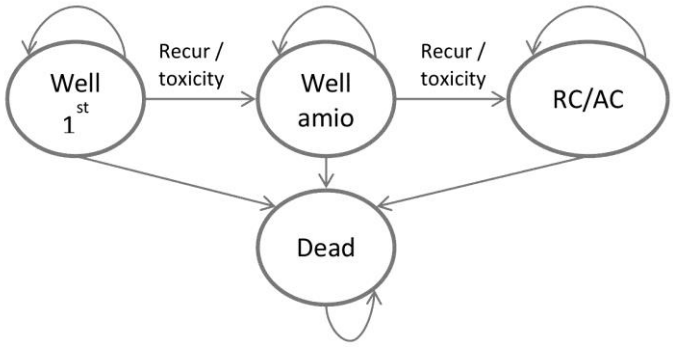


Figure 28 - Anti-arrhythmic drug therapy Markov model.

Patients initially receive a first-line drug (sotalol or flecainide), entering the "Well 1stdrug" state. In the event of toxicity or therapeutic failure, they start treatment with amiodarone ("well amio" state), and in the event of amiodarone failure, are treated with rate control ("RC/AC"). Amiodarone was chosen as the second-line agent for all patients in the drug "arm" based on its superiority over other drugs at maintaining sinus rhythm, however it is associated with more severe side effects. For all patients should be taken into account the mortality rate related to age and sex, so each state can lead to death. Except for the very small risk of death associated to ablation and the fatal toxicity of the drug, the Reynolds model assumes that the risk of death is the same for all health states, except for stroke after ablation. We, however, do not consider stroke different from other complications as in the study by Reynolds et al., assuming that the incidence of stroke is the same for both therapies and we do not consider explicitly the stroke as an outcome. As reported in previous studies, after the success of ablation, we consider a three months therapy with anticoagulants and antiarrhythmic drugs. The risk of toxicity related to antiarrhythmic drugs are obtained from the literature and are applied both to patients who undergo ablation and to those who follow pharmacological therapy only.

5.1.2.3. Collecting patients' preferences: Utilities and costs

The decision trees presented above are models developed from the original decision task directly taken from the guideline specification and available published literature. For this reason however they represent general models that still need to be personalized to each patient. This step is accomplished

through quantification of some of the model variables with patient specific values. In our framework these values are of 2 different types: health-related utilities and costs.

Quality-adjusted life years (QALYs) were chosen as the primary payoff for our models in the effort to combine in a single value the life expectancy and the subjective perception of the health states considering physical, mental and social aspects. Different patients may indeed have a very different perception of the quality of life related to health states [150]. To define QALYs, we thus need to characterize each health state included in the DT model by a utility coefficient (UC), ranging from 0 (usually associated to death) to 1 (perfect health). Despite scientific literature and the web provide UCs for several health states a dedicated utility elicitation tool has been developed in the effort to achieve the maximum degree of personalization through the use of UCs directly elicited from the single patient. The detailed description of the utility elicitation interface UceWeb developed for this purpose is delegated to the following section 5.2.1.

Another important aspect of the model quantification is the evaluation of costs associated to the different decision options. The patient is asked to provide the information needed for quantifying the monetary costs related to the clinical paths that are generated as a consequence of the different decision options. A dedicated questionnaire for cost assessment has been designed for the patients (a portion of which is represented in Figure 29). Of particular interest for the shared decision use case are the "out-of-pocket" costs, which are the costs directly burdening the patient and causing an economic impact on his/her activities. A general cost model, considering four categories of costs has been designed: i) costs related to the appointments the patient has to undergo during his/her treatment; ii) costs related to domiciliary care the patient may be in need of; iii) home adaptation costs and iv) costs related to the drugs the patient has to purchase. For the first and last category it is important to consider the context of the national healthcare service. In some countries, such as Italy, some patients might have these costs entirely covered, while, in some others, costs might wholly impact upon the patient's resources. The inclusion of costs related to domiciliary care (ii) takes into account the possibility of domiciliary assistance required after severe events (e.g. a stroke in our AF scenario). In cases where the assistant is a professional employed by the patient, the cost is quantified by the salary given to the assistant. If, on the other hand, the assistant is a member of the patient's family, this cost is quantified in terms of productivity loss. Finally it is worth stressing that since several cost components are related to the specific patient's context, the quantification holds as long as the context remains the same. When the physician perceives the context of the patient might have changed, a reassessment of the costs should be considered.

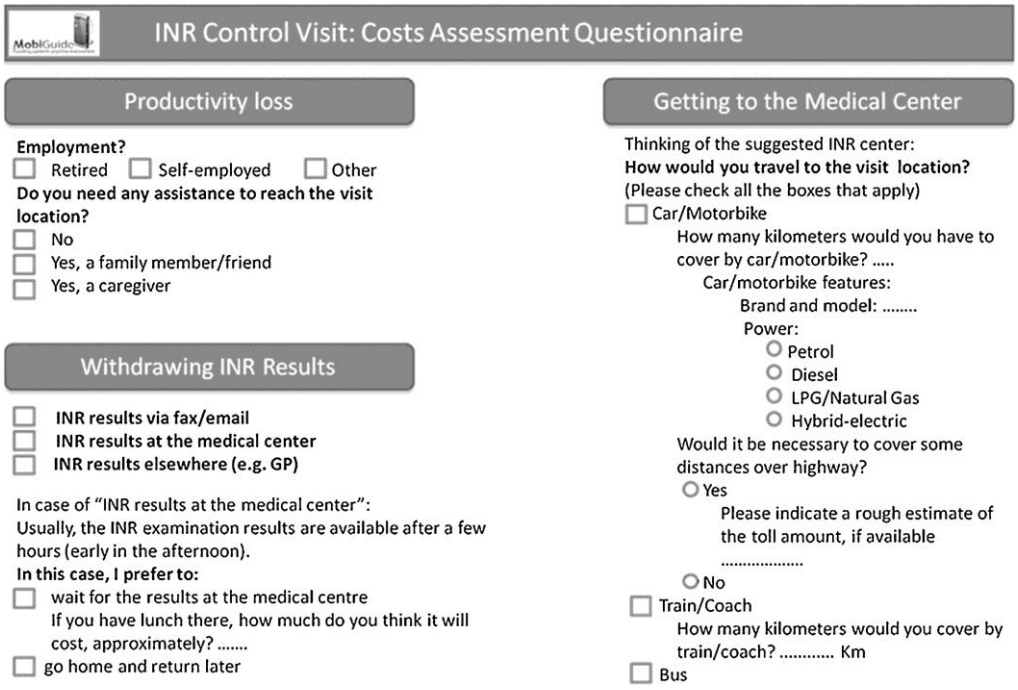


Figure 29 - A portion of the questionnaire administered to AF patients to quantify the costs included in the decision model. NOTE: INR is specific laboratory test required to monitor patients undergoing anticoagulant therapy with vitamin K antagonists (like warfarin).

5.1.2.4. Communicating DTs results

Apart from their obvious usefulness in decision analysis DTs are also an effective communication tool able to clearly and simply highlight the key outcomes of a decision problem for the patients [64]. To effectively participate in shared decision making patients should be made aware of the possible options, of the main scientific results already obtained about them, the risk levels of the major complications, and additional non-medical consequences of possible interest (e.g., costs). At the same time, they should not be burdened with too much information that may result overwhelming and confusing. To this extent DTs and Markov models, especially in their simplified graphical representation, are very effective tools that improve patient-physician communication, ease information sharing and ultimately benefit shared decision scenarios. The same requirements for effective and easy-to-understand communication also apply to the presentation of results to patients. Decision analysis results are traditionally reported as expected values of the outcomes of interest, usually tabular, purely numerical format. However, patients (and providers as well) may also be interested in knowing, for each decision option, the associated distribution of uncertain disease courses and outcomes over time [151]. On the other hand presenting such comprehensive information may be problematic, due its complexity.

For these reasons we developed a new functionality named DT-Health-related Life Paths (DT-HeLP) [152] dedicated to the graphical representation of decision analysis results. A health path is a representation of the expected temporal sequence of different health events that could arise from each alternative decision, over a lifelong time horizon. Most of the available decision analysis tools offer graphical facilities to show results of sensitivity analysis and probability distribution of outcomes. However, they do not allow the users to capture the temporal distribution of the health states and events that lead to those probabilities and expected values. Classic Markov models use the Markovian assumption, which makes state transitions memoryless. Our interest in tracking patient history, forces memory (i.e. what has occurred before entering each state and the amount of time a patient remains in the state) to be considered in the model. In our setting Monte Carlo microsimulations (MCMS) are a valuable tool to capture the inherent variability of real world contexts as well as to efficiently track prior history. Unlike traditional cohort analysis, MCMS retains memory of previous events from one cycle to the next one, recording information about individuals' history through the Markov model as values of tracker variables. In MCMS, individuals traverse the model one by one; at each transition, random number sequences, generated according to given probability distributions, are used to select a single path through the model. To capture the experienced health events, their duration, and the Markov cycle in which they occur, we used a combination of trackers and global matrices, another advanced TreeAge Pro feature that allows the dynamic storage of information. After the simulation is complete, the entire content of the matrix can be saved for visualization.

Tracking patients' history

As a first step, we defined the set of events to be recorded. In particular, for the purpose of our work, we included all the Markov health states but grouping IS and ICH in a single event that we called "stroke". The rationale for this grouping is that the quality of life of the affected patients does not depend on the type of events (IS/ICH) but on its consequences, that in principle are the same. As a matter of fact, even if with different probabilities, both stroke types can cause mild/moderate/severe physical and mental disability. Eventually, we identified the following health states: NVAF-only, temporary stroke, mild stroke, moderate-severe stroke, minor ECB, major ECB and MI. We defined a tracker variable associated to each of these states.

Tracking a patient's history within a Markov process requires to know the health events he experiences, their duration as well as the Markov cycle in which they occur. Hence, in addition to each tracker variable, we used a global matrix to store all the information. Global matrices are an advanced TreeAge Pro function that allows saving values globally and then using them for calculation or reporting purpose. The expression *GlobalN (n; row; col; value)* sets the cell *(row, col)* in the matrix *n* equal to the specific

value. All the information we need can be simply retrieved by defining the tracker variable as shown in Figure 30.

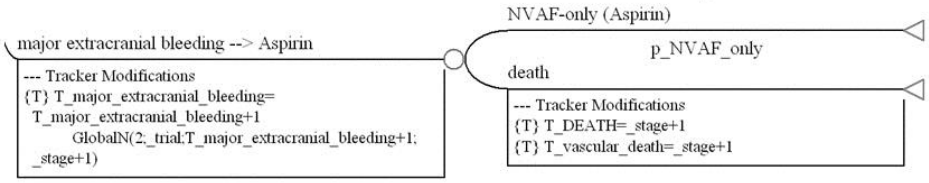


Figure 30 - Tracker variable definition at the "major extracranial bleeding" event node of a decision tree implemented using TreeAge Pro software.

Please note that the variables *_trial* and *_stage* are some built-in TreeAge Pro variables which count the MS iterations and the Markov process cycle, respectively. When running MS, if a simulation trial encounters a node with the tracker modification above, the current value of the tracker for that patient's trial is incremented by 1 and the cell (*_trial*, *tracker*, i.e. the event number) of the corresponding Global matrix is set to a value equal to the corresponding Markov cycle. To illustrate how our framework works, an example is shown in Figure 31. Our hypothetical patient begins the process in the "AF-only" state. It is supposed that the patient suffers a temporary IS at stage two and so he moves to the state "temporary IS". At the end of cycle 2, the tracker variables that count the events "AF-only" and "temporary IS" will be then increased by 2 and 1 units, respectively and the associated global matrices will be set. At the start of each individual trial, all trackers are reset to 0. After the simulation is complete, the entire contents of the matrix can be dynamically saved to a text file or excel sheet for further elaborations.

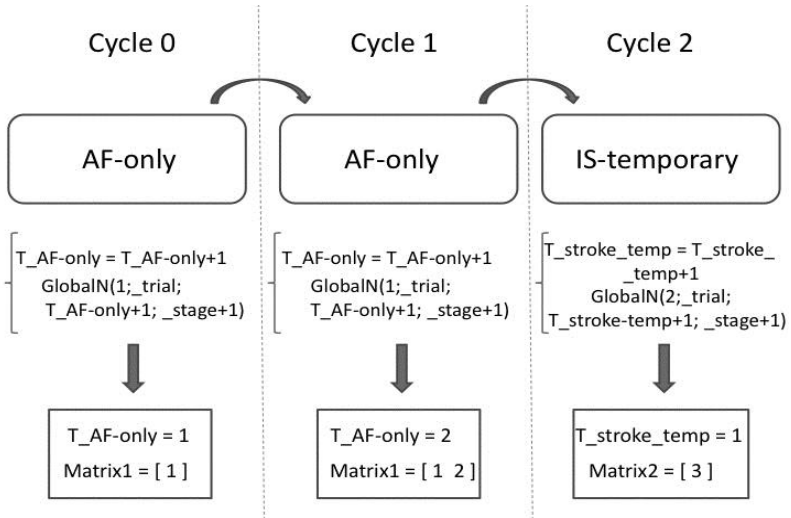


Figure 31 - Using tracker variables to keep trace of every event in a patient's life path.

In the following, we present some examples of the proposed visualizations that refer to the DT related to anticoagulant therapy selection in AF. For the sake of clarity only two of the four decision options, namely Dabigatran and no-treatment, are compared. Figure 32 depicts patients' lifetime paths using a stacked bar chart. Each one of the 100 bars displays the expected-life of a single patient. Each bar of the graph is divided into sections representing the different health states a patient goes through. Their position along the bar corresponds to the onset of the health state they represent. The length of each rectangle proportionally depicts the part of the patient's life spent in the corresponding state. Small triangles represent temporary events, such as temporary ischemic stroke (IS), intracranial hemorrhage (ICH) and extracranial bleedings. The color of both triangles and bars relates to the severity of the condition (darker color indicating more severe condition). This graphical form provides a valuable tool for integrating several information. In particular, for each decision option, it allows conveying information about the survival trends as well as the course of disease in terms of succession of different health states. Bars can be sorted according to different criteria but sorting according to the expected QALYs makes the survival trend more explicit, with some hints on the quality of life.

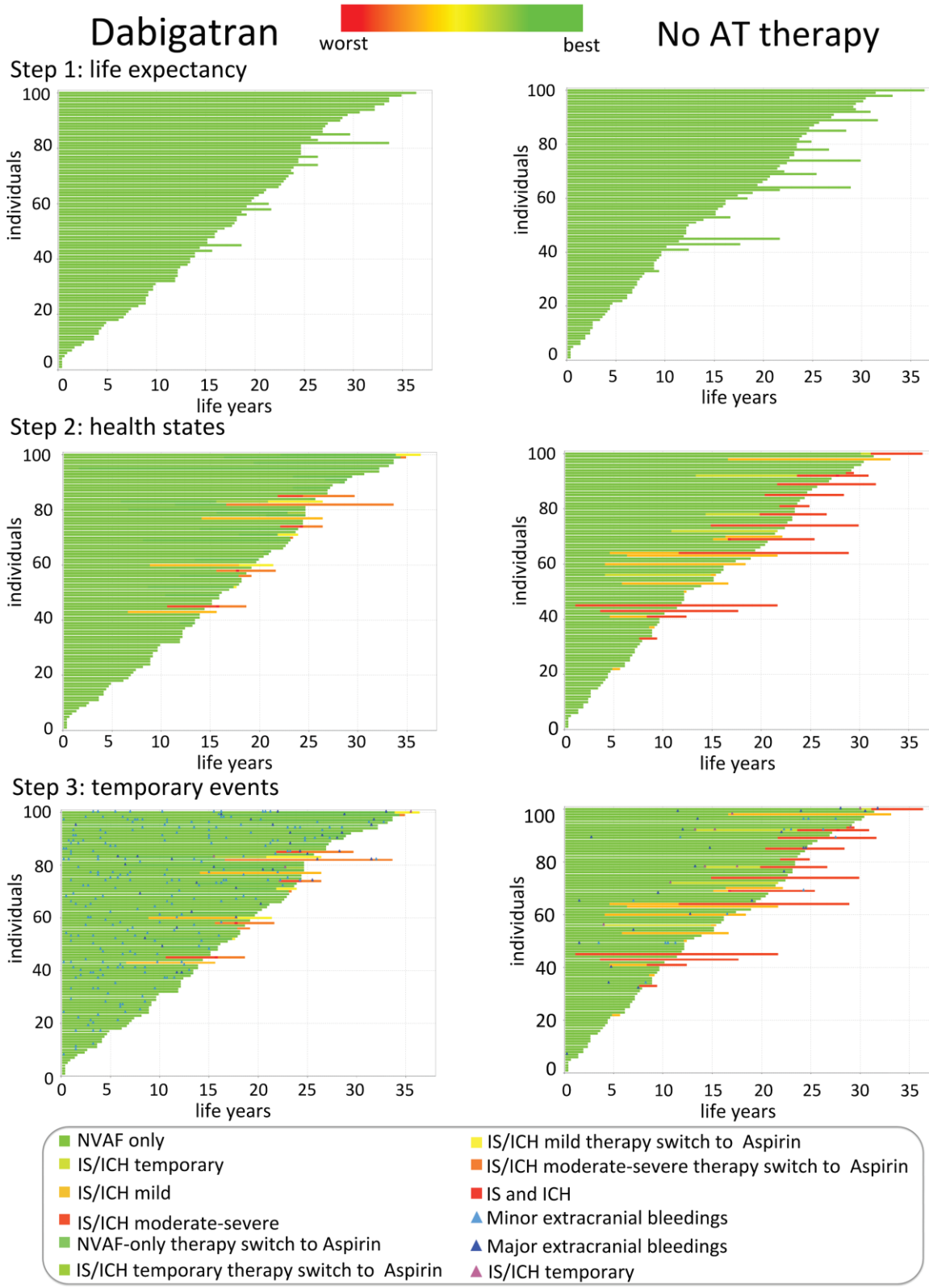


Figure 32 – Simulated health-paths of 100 patients treated with Dabigatran versus no antithrombotic therapy. Starting with life expectancy, the physician may gradually complicate the output (in steps) by adding further information about the related health states and then about temporary events.

5.2. Computerized approaches to health-related utility elicitation

In the context of the overall shared decision making framework presented in the previous sections one of the most challenging objective was patient preferences elicitation. In the MobiGuide system part of the patient-specific values that define the personal contexts (e.g. meal-times and preferences about receiving reminders or not) are acquired using the caregiver interface and stored in the PHR. A more subtle category of user preferences are those connected to health-related quality of life. Assigning a value to health states in terms of utility coefficients is an essential step for their use in the quantification of the DT models. Health-related utility elicitation requires a carefully designed process and very specific techniques. For this purpose two different approaches, which will be discussed in the following sections 5.2.1 and 5.2.2, were investigated: i) a web-based tool for direct utility elicitation and ii) a predictive model able to estimate utilities from other patients' preferences. This lead to a somewhat independent line of research (not necessarily confined to CDSS, and closer to quality of life studies, health-economics, and medical decision analysis) and ultimately a spin-off system able to work also in stand-alone mode.

5.2.1. Utility coefficients elicitation with UceWeb

Utility Coefficients Elicitation Web interface (UceWeb) has been developed to be a standard interactive tool to perform computer-assisted utility elicitation during shared decision encounters between patient and physician. Capitalizing on utility theory presented in chapter 3 the system has been designed to implement three main direct elicitation methods: standard gamble, time trade-off and rating scale. UceWeb however also has the secondary objective of systematically store elicited UCs to build a public repository to be further exploited in studies on specific target populations (e.g. cost/utility analyses).

A number of other efforts for the development of utility elicitation tools have been reported in the literature [153–156] but they all lacked some of the essential features of UceWeb. In particular, in respect to the existing systems, UceWeb is more general (it does not address a specific disease or class of patients), collaborative (different users including patients, physicians and researchers collectively participate in the growth of the QoL data repository through web-based collection of data on voluntary basis), and scalable (additional elicitation methods can be implemented without changing the system architecture). In other words, it is intended to use the UCs, elicited during face-to face clinical encounters, not only for the actual decisions, but also to build a UC repository, from which a researcher should be able to retrieve data. Every performed elicitation can indeed be immediately used in the quantification of DTs, but also contribute to the

growth of the UCs repository. The UceWeb platform, even when used in stand-alone mode, can thus become a unique source of data allowing researchers both to perform more reliable comparisons among healthcare interventions and build statistical models to gain deeper insight into quality of life data. As that repository grows, researchers will be more and more able to retrieve data close to their target population. This will solve the problem of several health economics studies that, in lack of such data, make use of UCs elicited from populations different from their target one.

5.2.1.1. Architecture

The functional architecture of UceWeb is represented in Figure 33. The web interface allows direct elicitation of UCs. This is mainly done during face to face encounters between a patient and a doctor, but also could be done by the patient alone when his doctor feels that, after extensive training, he may continue with self-elicitation. In the cases when UCs cannot be elicited directly from the patients (e.g., the patient is a child, or he is not able to answer do to the severity of his condition), an informal caregiver may replace the patient and answer to the questions on his behalf. However this eventuality should be avoided as much as possible since the elicitation process is formally supposed to be carried out by the patient himself. UCs are a formal representation of a patient very own preferences and quality of life perception which might be difficult to mediate through an interposed person (to face this issue a novel approach to estimate utilities based on the preferences of "similar" patients has been studied and will be presented in section 5.2.2). The result of each elicitation performed with UceWeb is stored in the system dedicated database. This acts as a collaborative repository that builds up in size and relevance as more elicitations are performed. Such a UCs repository provides a valuable source of data for the quantification of decision models, as well as for further research about UCE methods themselves.

In the effort of developing an integrated framework for shared decision a repository of decision models making will also be available to the users, with the possibility of personalizing them according to the specific patient's UCs. Currently, only the two models presented in section 5.1.2.2 are fully integrated with UceWeb. Figure 34 shows the architecture of the integrated components of MobiGuide shared decision framework.

Figure 35 shows the technical components of UceWeb. The tool has been developed as a web application mainly relying on Java technology and Model-View-Controller (MVC) architectural pattern. The java application core is linked to the jsp-based presentation layer using Struts 1.3 while the persistence layer is managed with Hibernate 3.2. Finally, the javascript library JQuery UI is used in the user accessible pages to guarantee richer user interaction capabilities.

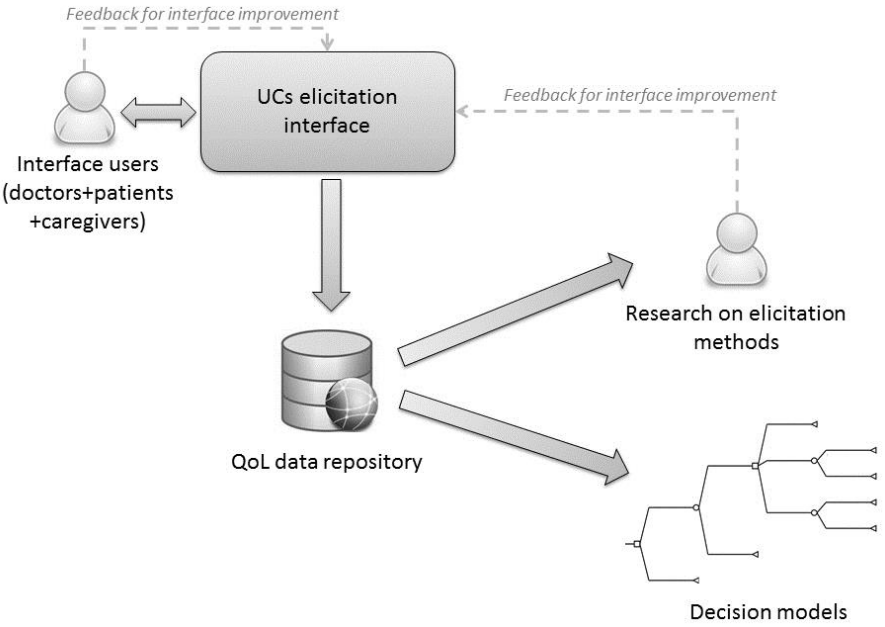


Figure 33 – Functional architecture of the UceWeb utility elicitation interface.

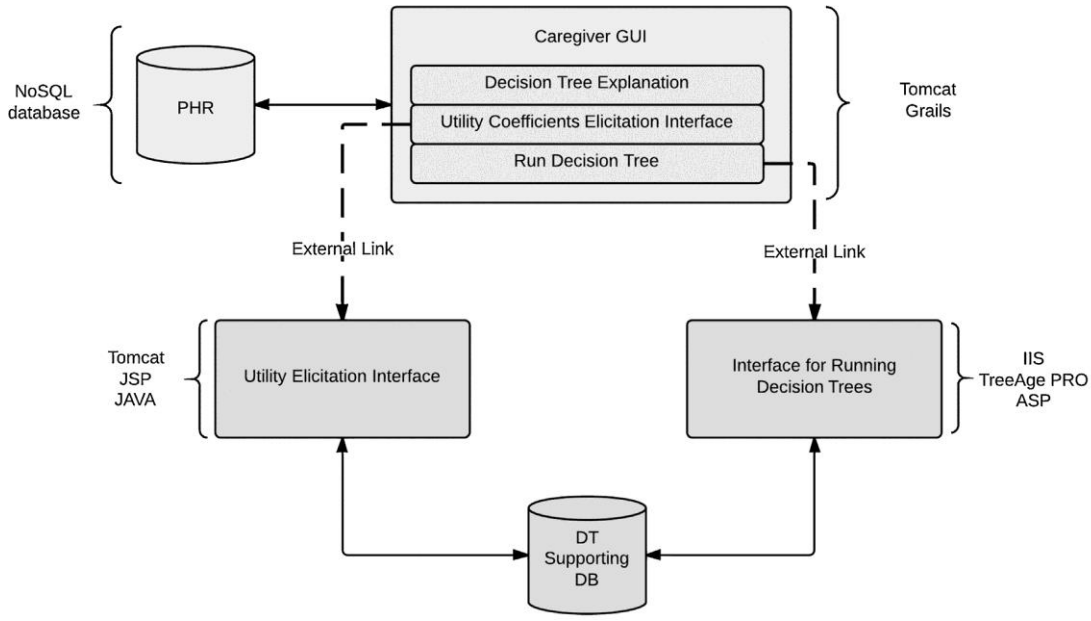


Figure 34 – Logical schema of the integrated framework of the shared decision making framework adopted in MobiGuide. Main supporting technologies are reported next to each component.

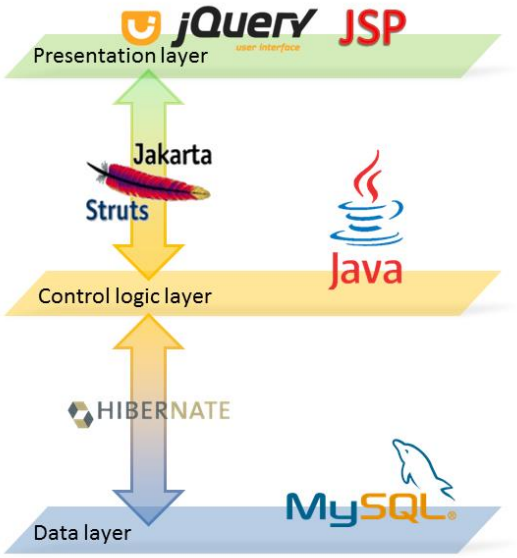


Figure 35 - Implementation technologies used in UceWeb. Separation of presentation, control and data layers comply to the MVC architectural pattern.

5.2.1.2. Data model

Figure 36 describes the logical data model of UceWeb in the Entity-Relationship diagram formalism. In the following we describe each entity in the effort to give more detail about the information collected and stored by the tool.

The *Patient* entity represents the subject to whom the UCs are referred. Although fully anonymized due to privacy restrictions, it is deeply characterized by a set of attributes that define a sort of patient profile. The information collected in the patient profile are: age, gender, race (essential to present a realistic estimate of life expectancy while asking the TTO question), education, marital status, job, computer literacy and geographical region. All this information can be useful for further statistical stratification. Names, surnames and other identifying information (social security number, phone numbers, email address, postal addresses, etc.) are never requested or stored. To be able to meet the strict requirements for fully de-identified medical data (e.g. the American HIPAA [157, 158]) all the ages over 89 are stored as “90 or older” and the country of residence is the only geographical information collected. The *Life expectancy* entity reports mortality tables. In the case of Italian population, data coming from national statistics institute ISTAT (www.istat.it) have been used. A non-Italian user may upload, in the same entity, a different table with different survival statistics. The *Utility coefficient* entity stores the UCs gathered with UceWeb. In addition to storing the elicitation date and the person who actually answered the elicitation questions (patient/relative/caregiver) it is related to *Health state*

and *Doctor*. The former represents the focus of the elicitation, and it is linked to a number of multimedia contents used to describe it to the patient. The latter administers the elicitation questions, and it is possibly characterized by his/her medical specialty. *Elicitation method* entity contains SG, TTO and RS. Weighted average values of SG and TTO may also be stored, as further detailed in the following.

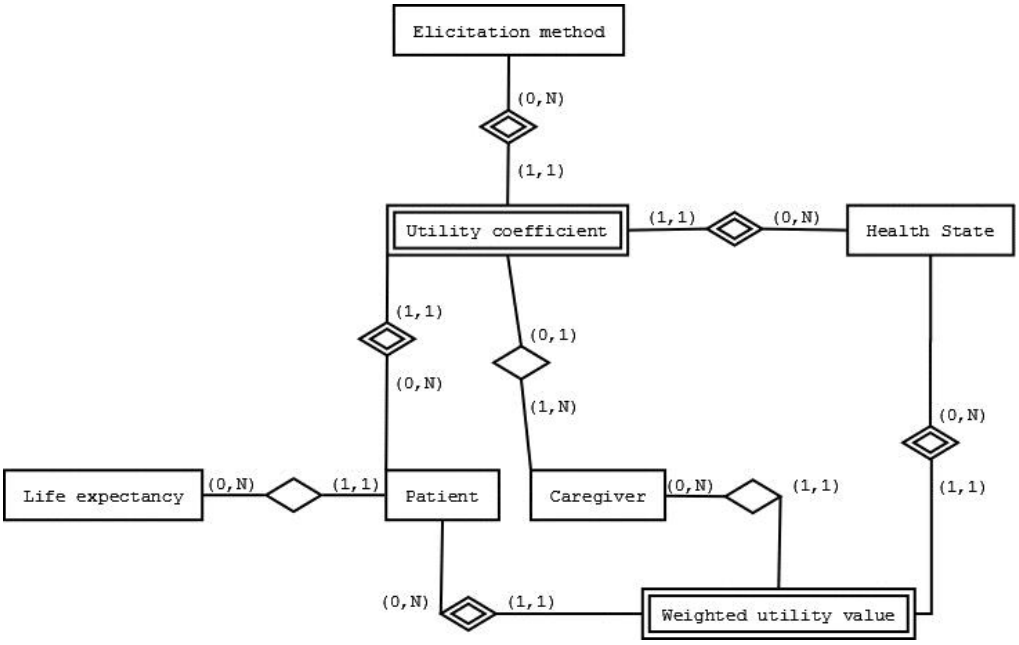


Figure 36 - ER diagram of the data model underlying UceWeb.

5.2.1.3. A typical elicitation session

A typical elicitation session starts with a Rating Scale (RS) elicitation (Figure 37) that provides UC_{RS} values. Since the output of RS is not a proper UC, as highlighted in the methods chapter, this first step is not mandatory but provides a useful starting point to the other methods and helps avoiding bias and anchoring effects.

After RS has been used, a Time Trade-off (TTO) elicitation follows. In the TTO method (Figure 38) the first question is dynamically generated starting from the patient profile to calculate life expectancy LE, and from UC_{RS} , if available, to assess the proposed time to give up, calculated as LE* UC_{RS}. When UC_{RS} is not available, LE*0.5 is used. The patient can operate a slider, changing the amount of time to trade off in order to heal completely. Note that the question is asked in both positive version (amount of time to be lived) and negative version (amount of time to give up) to avoid bias effects.

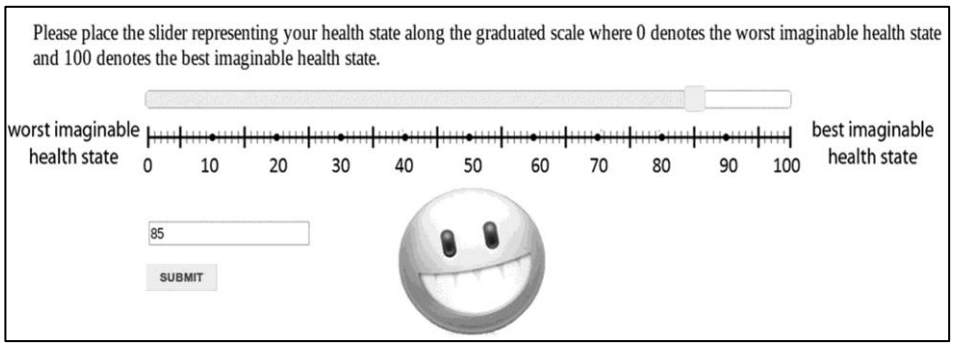


Figure 37 – User interface for RS elicitation. The use case is a patient with Spinal Cord Injury (SCI). The slider is operated by the patient and placed in on proper position on the numerical scale. A graphical visual aid (which can be shown or not, depending on configuration options) helps the patient in the process.

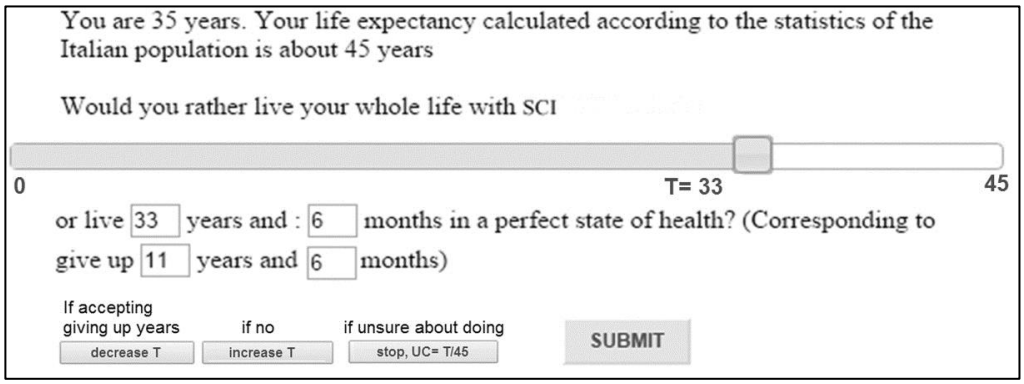


Figure 38 - User interface for TTO elicitation. Expected life years used in the method are calculated using statistical data for the Italian population. Life expectancy data for different countries can be uploaded and used by the tool engine.

Finally Standard Gamble (SG) elicitation starts with one basic question to check the patient's risk aversion (Figure 37). In case of negative answer (i.e. the patient does not accept any risk), the elicitation stops and the utility coefficient of the health state is set to 1. In the other case, the patient is presented with a first risk value, calculated as $SG_0 = 1 - UC_{RS}$ or $SG_0 = 0.5$ if UC_{RS} is not available. The proposed risk for the gamble is selected with an iterative process driven by a bisection algorithm until the indifference point SG_{final} is reached.

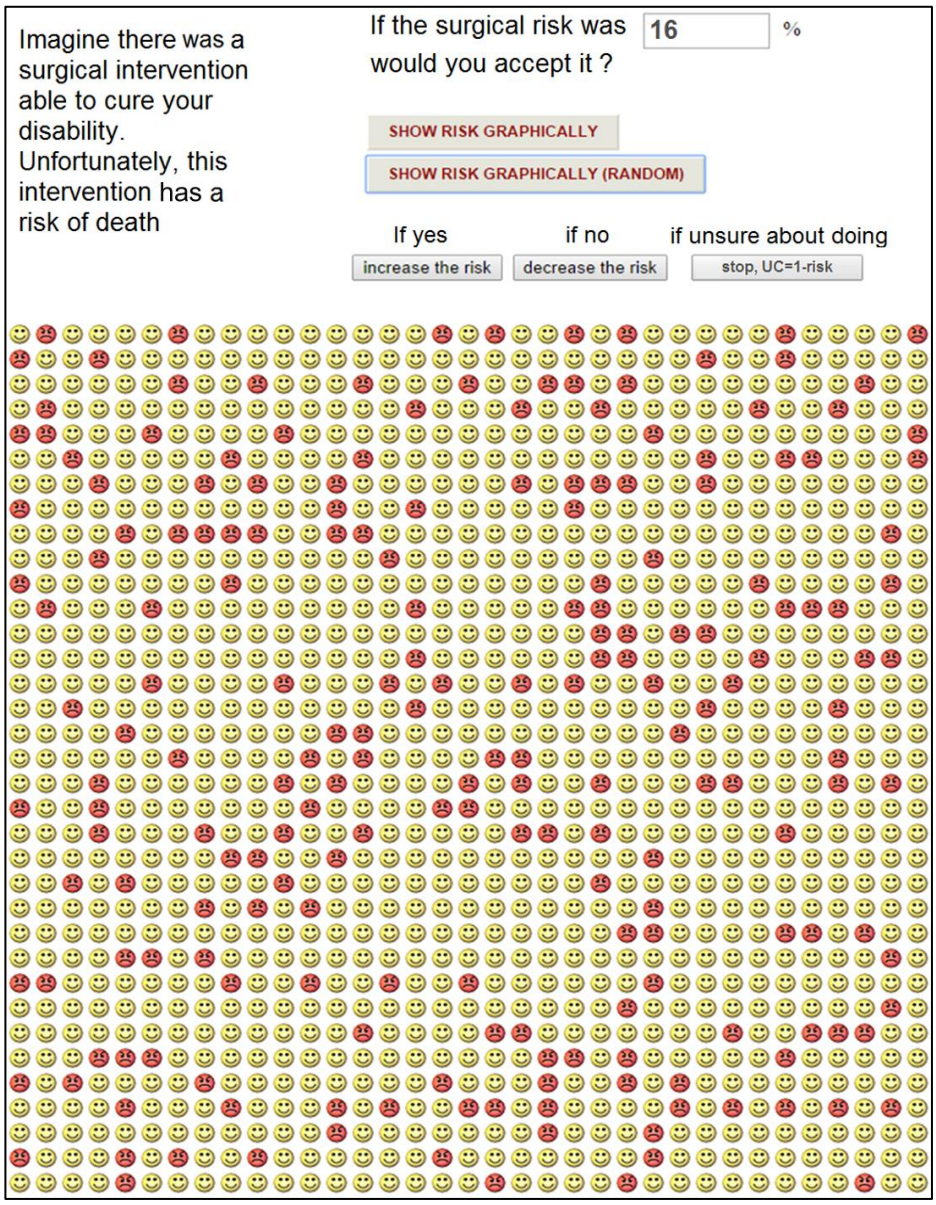


Figure 39 - Interface for SG elicitation. A grid of icons helps the patient to visualize the current risk of death (16% in the reported example). Red icons represent patients that die during the hypothetical procedure while yellow ones stand for those who live and heal completely. The random placement of red and yellow icons in the grid helps conveying the concept of chance. An alternative visualization with all the red icons displayed first is also available.

5.2.1.4. Facing the challenges of utility elicitation

In this section we provide a list of the main challenges faced during the design of UceWeb along with the measures adopted to overcome the major limitations of computer-assisted elicitation:

- Difficulty in understanding questions and elicitation methods

Utility elicitation methods can be difficult to understand for patients. This is one of the reasons why elicitation is usually assisted by a trained professional. However, some expedients can be adopted also in the design of the computerized elicitation tool to facilitate patients. For example, during RS elicitation, if a patient needs support beyond the numerical representation of values, a visual aid in the form of different smiles can be enabled (Figure 37). Another example is the graphical help for giving patients a better intuition of the "risk" concept they might be not familiar with. Numerical representations alone can be difficult to interpret, particularly when dealing with age-related impairment or stress. When presenting risk of death in SG elicitation, UceWeb can show a grid of smiles where the number of red icons reflects the portion of patients that would die according to the currently set risk of death (Figure 39). Icons can be alternatively placed all in a row since previous literature suggests that the estimated risk tends to be more accurate with this arrangement [159], or randomly positioned on the grid thus giving a better representation of the chance surrounding the risk concept.

- Dealing with hypothetical scenarios

UCs can be elicited for health states the patient is not experiencing at the moment of the elicitation or he has never experienced in his life. As stated in the previous sections, in a shared-decision making framework there is often the need to represent possible future health states in decision models [72]. The architectural design of the UceWeb tool allows to provide an effective explanation of the health state to the care professional with rich multimedia informative material including photos, videos and text documents. This contents can also be tailored to the specific patient, according to his attitude to look at stronger (e.g. very sick people, real patients movies) or softer (e.g. cartoons, text) material.

- Bias and anchoring

The value used in the initial question of SG ("...would you take a 5% risk of death?...") and TTO methods plays a major role in the elicitation process. Indeed it may lead responses to cluster around that value (anchoring), and it could be influenced by the physician's knowledge on the actual surgical risk of an intervention (bias). In practice, the choice of this initial value has traditionally been delegated to the professional's expertise. To deal with these issues, UceWeb exploits the values elicited with the RS method. Starting the patient's interview from the RS allows to use the elicited values as starting points for TTO and SG, because they should be reasonably close to the actual patient's preferences. This approach is in line with the suggestion of Torrance et al [82] who, recognizing the limits of RS, suggests not to deprecate its use, but applying

it together with additional, sounder, methods. Finally, to avoid framing bias effects, questions like "...would you take a 5% risk of death?" are also be presented by the application under the opposite perspective "…would a 95% chance of living be acceptable to you?".

- Patient fatigue

Multiple iterations on SG and TTO questions can lead to fatigue and boredom that could considerably alter the elicitation results. The choice of the next question is usually left to the doctor/psychologist. On the other hand UceWeb implements a bisection algorithm that automatically chooses the next question to optimize the tradeoff between the number of questions and the accuracy of the elicited UC. To illustrate the algorithm, we rely on an example applied to the SG elicitation method. In the first step of the iterative process a first risk value SG_0 is presented to the patient. The initial possible boundaries for the acceptable risk are $Boundary_{high} = 1$ and $Boundary_{low} = 0$. If the patient declares that SG_0 is an acceptable risk, the lower boundary gets updated to $Boundary_{low} = SG_0$, and a new risk $SG_1 = SG_0 + (Boundary_{high} - Boundary_{low}) / 2$ is proposed to him. On the other hand, if he declares the risk SG_0 to be unacceptable, the upper boundary is updated to $Boundary_{high} = SG_0$ and the next risk proposed to the patient is $SG_1 = SG_0 - (Boundary_{high} - Boundary_{low}) / 2$. The dialog continues in an iterative fashion that, at each step, shrinks the range where the utility value is known to be contained (i.e. $SG_{indifference}$ is contained in the interval $[Boundary_{low} , Boundary_{high}]$) until the indifference point is reached. In UceWeb the UCs are elicited with a 2 decimal digits granularity, resulting in 101 possible different values of UCs (ranging from 0 to 1, with 0.01 increment steps). The bisection algorithm described can be considered an implementation of a binary search tree over a sorted array of 101 elements. In general, with an array of N elements, the average successful search would take $\log_2(N)-1$ trials, with a worst case complexity of $\log_2(N)$ [160]. With N=101, this leads to approximately 5.66 questions to be asked. If increased precision is needed for UC (i.e. a third decimal place), the size of the searchable array is N = 1001, with approximately 8,97 questions needed. In practice, however, the iterative process can end earlier, if the patient declares to have reached his indifference point before restricting the interval $[Boundary_{low} , Boundary_{high}]$ to a single value.

5.2.1.5. Pilot experiments in the atrial fibrillation scenario

As a first application of the developed tool, UceWeb was used to assess quality of life related to atrial fibrillation health state that was of particular interest for the MobiGuide project. This assessment involved 20 patients and had the secondary purpose of investigating the correlation between UCs elicited with our tool and UCs derived from two validated instruments: namely the EuroQoL-5D [78] and AFEQT [161]

questionnaires. EuroQol-5D is a standardized measure of health status developed by the EuroQol Group to provide a simple and generic measure of health for clinical and economic assessment. It consists of 5 questions addressing mobility, self-care, usual activity, pain and discomfort, and anxiety/depression. Each question has 3 possible answers, namely: no problems, some problems, extreme problems. A mapping between the answers and a utility coefficient is available. The AFEQT questionnaire is an AF-specific questionnaire designed to assess the impact of AF on patients' quality of life. AFEQT consists of 20 questions, divided into 4 sections, namely Symptoms, Daily Activities, Treatment Concerns, and Treatment Satisfaction (see Appendix A). The responses to each AFEQT question are scored on a 1 to 7 Likert scale. Questions 19-20 are related to patients' satisfaction with treatment and are not included in the overall score of AFEQT, which is instead computed on the rest of the questions. The raw scores are transformed to a 0 to 100 scale, where a score of 0 indicates the most severe symptoms or disability and a score of 100 indicates no limitation or disability.

Twenty volunteer subjects were recruited among the admitted patients of the division of cardiac rehabilitation (n=18) or seen at the outpatient clinic (n=2) at Fondazione Salvatore Maugeri (FSM) hospital in Pavia between April and July 2013. The patient set included 10 males and 10 females, of ages between 34 and 79 years (average 66.2 years) with various diagnosis of atrial fibrillation (4 paroxysmal AF, 10 persistent AF, and 5 permanent AF). Each patient carried out one complete elicitation session including all the three methods while assisted by his/her doctor. Table 6 reports a summary of the UCs for the three UceWeb methods and the scores of the two questionnaires.

Table 6 - Mean and standard deviation (mean ± SD) of the scores obtained on the 20 AF patients.

AFEQT					EuroQol	U_{TTO}	U_{SG}	U_{RS}
Overall score	Symptoms	Daily activ.	Treatment concern	Treatment satisf.				
67.71 ±19.02	76.09 ±20.31	66.67 ±25.17	63.58 ±22.44	67.13 ±20.10	0.586 ±0.369	0.979 ±0.058	0.977 ±0.219	0.669 ±0.196

Since the questionnaires used are validated tools to assess quality of life, the correlations among their scores and the UCs elicited with the three implemented direct methods were investigated. The correlation matrix summarizing the results of this analysis is provided in Table 7.

Table 7 - Correlation coefficients (p-values) between the quality of life values elicited with different methods. (ns) indicates non-significant correlation.

	AFEQT	EuroQOL	SG	TTO	RS
EuroQOL	0.32(ns)				
SG	Overall:0.32(ns) Symptoms:0.49(0.04) Daily activity: 0.43(0.07)	0.58(0.02)			
TTO	0.56(0.015)	-0.18(ns)	0.26(ns)		
RS	0.24(ns)	0.35(ns)	-0.04(ns)	-0.07(ns)	
Avg (TTO,SG)	Overall: 0.56(0.02) Symptoms: 0.67(0.002) Daily activity: 0.63(0.005)	-0.1(ns)	-	-	-0.14(ns)

Significant correlations were obtained between the overall AFEQT score and the TTO method and between the AFEQT symptoms subsection score and the SG method. However, the most interesting finding regards the relationship between some AFEQT scores and the mean of SG and TTO utility coefficients. For the symptoms subscore, we found a correlation coefficient of 0.67 ($p<0.002$) (see Figure 40 for a graphical representation), whereas for the daily activities we found a correlation coefficient of 0.63 ($p=0.005$). These results suggest that SG and TTO probably capture different aspects of quality of life, and must be jointly considered when used in a shared decision making scenario to have a better picture of the patient's perspective.

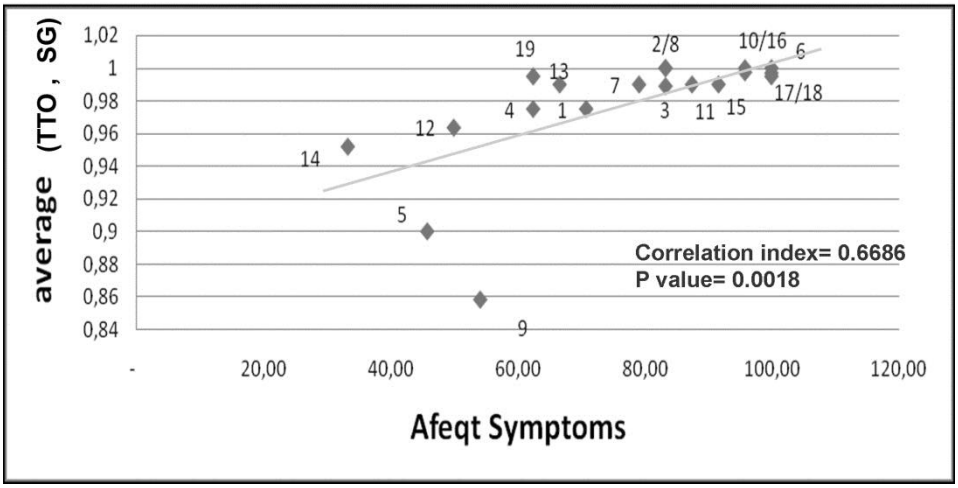


Figure 40 - Correlation of the AFEQT symptoms score with the average between TTO and SG values.

5.2.1.6. Application to spinal cord injury patients

After the first experience with AF patients, UceWeb has also been successfully used to assess quality of life related to Spinal Cord Injury (SCI). SCI is a damage to the spinal cord that results in a loss of function such as mobility or sensation. The most common causes of SCI are trauma, such as traffic accidents, gunshot injuries, knife injuries, falls and sports injuries or disease (e.g. poliomyelitis and spina bifida). The outcomes of SCI depend on the type and level of the injury. Patients usually have permanent and often devastating neurologic deficits and disability, showing not only damage to independence and physical function, but also many visceral complications such as bladder and bowel dysfunction. Work disability and productivity loss also bring psychosocial and economic burden. Understanding the factors that can predict higher QoL is important to build an individual rehabilitation plan for patients with SCI where the ultimate goal of rehabilitation is indeed to enhance QoL.

Fifty patients were recruited consecutively from October to December 2014 from the FSM clinic. SCI were classified as complete or incomplete using the American Spinal Injury Association (ASIA) classification [162]. The ASIA scale grades patients based on their functional impairment as a result of the injury, from A (worst state) to E (best state). The sample included 28 male and 22 female patients with a mean age of 54.5 (±15.5) years, ASIA classes between A and C (18A, 12B, 20C) and various causes of injury (17 traumatic, 8 vascular, 10 compression, 3 neoplasm, 8 myelitis, 2 cauda and 2 spina bifida). UceWeb has been used to administer RS, TTO and SG to the patients who also answered the SF36 questionnaire [80], from which additional UCs were calculated using Brazier's formula [81] for comparison.

The results (see Figure 41) pointed out that patients seem to understand RS and TTO methods slightly better than SG, for which two patients were not able to answer. While showing high variability, UCs elicited with all the different methods were significantly correlated. Moreover according to past literature, RS values are significantly lower ($p<0.01$) than TTO and SG ones, that incorporate the concept of risk in their definition. Since TTO and SG results are in general similar but not identical (and according to the findings of the previous experiment on AF), we decided to use their average value for further analysis. A first observation regards gender differences. SG method gives significantly different results showing lower UCs for males (**Figure 42**). Given the type of question that characterizes the method, this highlights a higher risk aversion in females (note that the majority of road accidents occur to males). Regarding differences among ASIA categories, ASIA C patients have higher UCs than A and B, but only when considering chronic phase. This could be explained by the fact that during acute phase patients are hospitalized, thus living in a safe environment (**Figure 43**). Finally, a subset of patients of particular interest for the study of QoL are those where the SCI caused an abrupt change in daily life (e.g. traumatic or neoplastic causes). In this sub-population, a

trend implying that chronic phase yields lower UCs than acute is visible (**Figure 44**), probably for burden of reintegration in daily life. In the same group, patients with low education had significantly higher ($p<0.038$) UCs than those who had higher education (**Figure 44**). This could be explained by higher adaptation capabilities of patients with low-education.

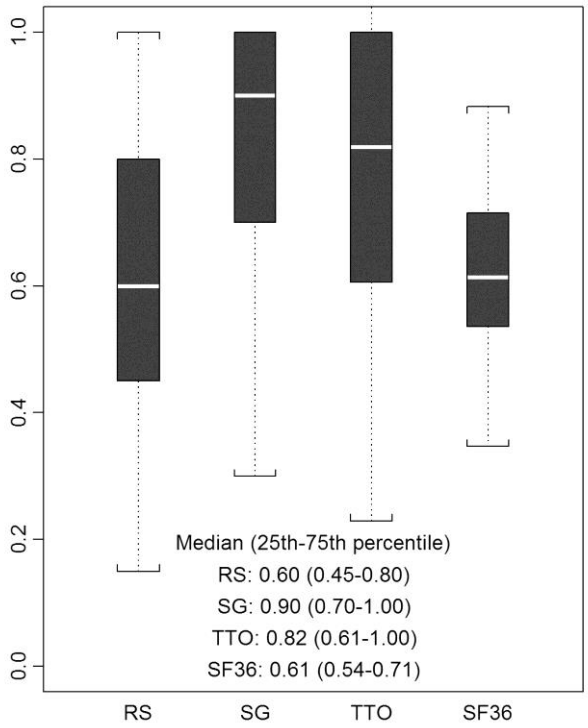


Figure 41 - Utility coefficients for SPI elicited with the 3 UceWeb methods and the SF36 questionnaire.

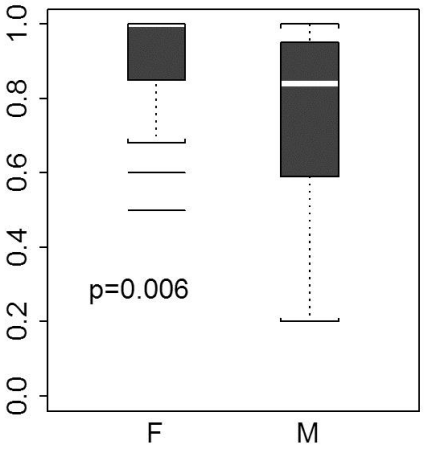


Figure 42 – UCs elicited with the SG method show significantly lower values for male patients.

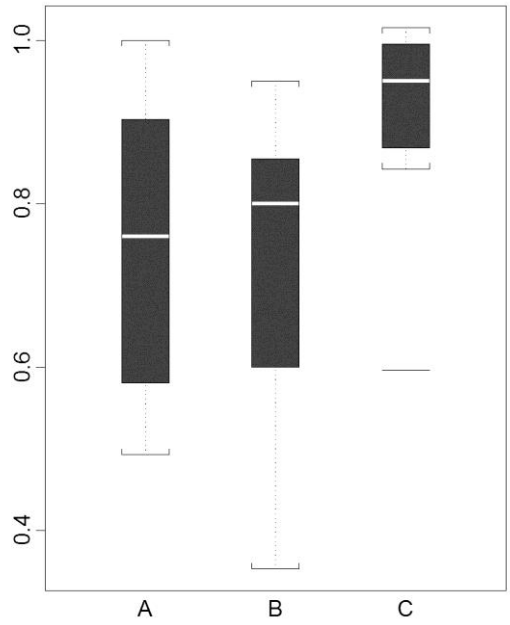


Figure 43 - UCs for patients in chronic phase with different ASIA classes.

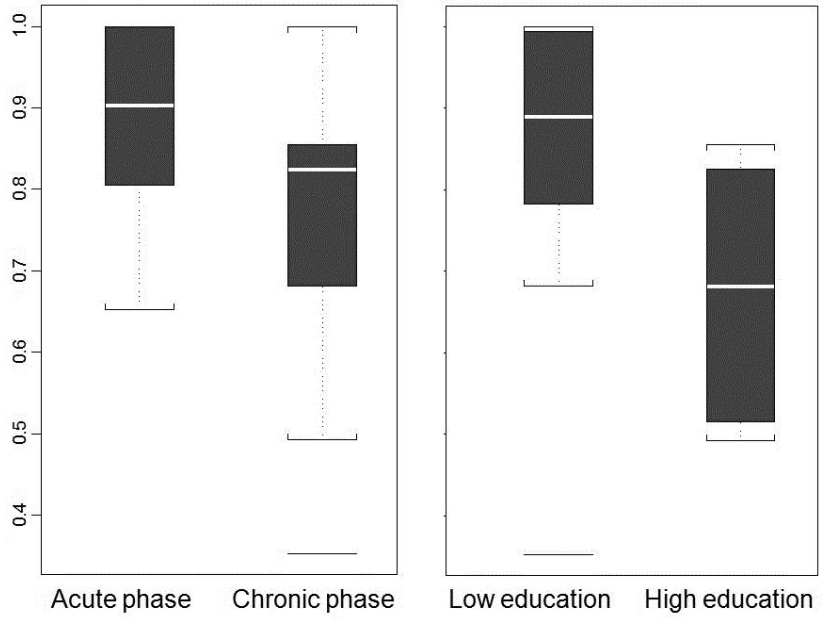


Figure 44 - UCs elicited for patients who suffered from a sudden injury. In this sub-population differences between acute/chronic phase and low/high education are observable.

5.2.2. Recommender system for utilities

In the previous section 5.2.1 we presented a tool specifically developed to assist computerized utility elicitation from patients. However direct utility elicitation methods like SG and TTO are still characterized by a number of unresolved challenges that leave space for further research. In this section we explore an alternative approach to health-related preferences elicitation able to predict UCs capitalizing on the preferences of similar individuals.

5.2.2.1. Motivation

While describing UceWeb [163] we stressed the significant effort devoted to carefully design the tool and achieve the double goal of interface usability and accuracy/reliability of the resulting UCs. However some limitations of the current direct elicitation approaches are still hard to overcome. Utility elicitation methods like SG or TTO, albeit theoretically sound, are often difficult to understand for patients. SG asks the patient to evaluate a hypothetical risk of death, which some patients with severe conditions might not be willing to reason upon. Similarly, TTO asks patients to evaluate the possibility of giving up part of their life but living better. Since this could not be sensible for patients with mild impairment, many different variants of TTO have been developed to try to overcome these limitations [84]. Moreover ability to correctly understand the questions is essential to guarantee the quality of the elicitation results. This is one of the reasons why elicitation is usually assisted by a trained professional (physician or a psychologist). However the presence of a human interviewer can add some significant bias and anchoring effects. For example the value chosen for the initial question ("...would you take a 5% risk of death?...") in the SG method can indeed lead responses to cluster around that value (anchoring), and it could be influenced by the physician's knowledge of the actual surgical risk of a real-world intervention the patient could undergo (bias). Another factor that can impact the effectiveness of direct elicitation is the need for the patients to evaluate unfamiliar health states. As a matter of fact, in a shared decision framework, decision models often represent possible future health states that can occur as consequences of the different treatment options. Patients can have a hard time answering the elicitation questions for health states they are not experiencing at the moment or have never experienced in their lives. Finally, all the issues that affect direct elicitation can be even more evident in particular groups of patients [164, 165] like the elderly or those having cognitive impairment. In extreme albeit not so rare cases, direct elicitation might also result impossible to perform.

In these cases a traditional approach consists in doctors taking over the whole responsibility of the clinical decision, even if some involvement of the patient would still have been appropriate. Another often applied possibility is (e.g. in cost/utility analysis studies) to use UCs available in

the literature to quantify decision models, disregarding that they could have been derived from a different population. However, it is nowadays acknowledged that environmental data play an important role in quality of life assessment [72, 166].

Applying one of these generic approaches inevitably results in the loss of personalization of the decision process (see [167] for a case study in AF anticoagulant therapy selection) which, on the other hand, is one of the focal points of the shared decision theory and of this dissertation. In order to address these challenges we developed a system based on collaborative filtering that aims at predicting the values of UCs for a patient, based on the preferences of similar individuals.

5.2.2.2. High-level architecture

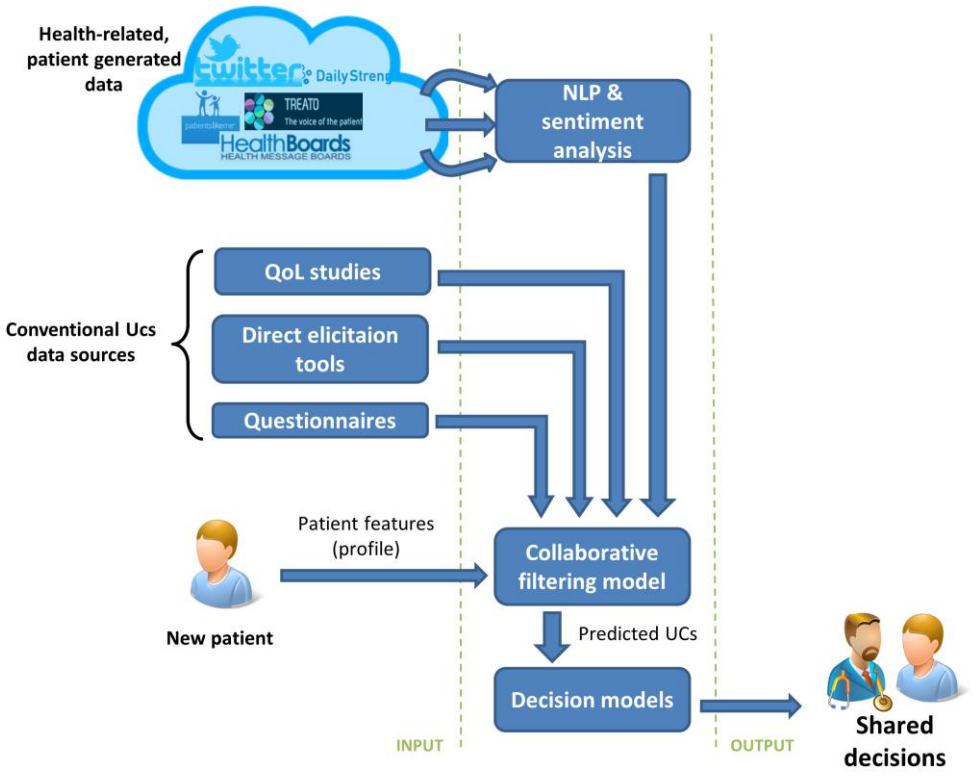


Figure 45 - Logical architecture of the UCs recommender system and its integration in the shared decision framework.

The basic idea consists in building a recommender system able to give an estimate of the preferences (in the form of UCs) of a patient regarding a set of health states of interest. Such a system, taking advantage of data regarding preferences of other patients, could be able to still ensure better personalization of the decision task than using general-purpose non-personalized decision models.

The core of the system is represented by a collaborative filtering (CF) model (Figure 45) that collects UCs from a variety of sources including: i) QoL studies published in the literature (note that, apart from the average population-level UCs, several studies also report the complete dataset of the elicited UCs); ii) UCs elicited using direct elicitation tools (e.g. UceWeb); iii) UCs elicited by the use of questionnaires like EuroQoL [78] or SF-36 [80]. In addition to these conventional UCs data sources we also explored the possibility of including unstructured, patient-generated data in our recommender system approach. In fact availability and relevance of raw data about patients' preferences, QoL and other health-related information is growing. More and more often patients report their status, share their experiences and discuss their health in discussion boards and other social platforms [168]. As we reported in chapter 3, when we introduced recommender systems, this kind of user-generated data has been widely used in the industry to gain valuable knowledge about people preferences. More details about the data sources used to train the CF model will be given in the next section 5.2.2.3, with a particular focus on the NLP module in section 5.2.2.4.

Once properly populated, the CF model would then be able to estimate the value of a UC (or a set of UCs) for a new patient given a set of features that define his profile.

The output of the collaborative filtering component can finally be used to quantify the parameters of formal decision models (such as one of the decision previously presented in section 5.1.2.2) thus contributing to the personalization of the decision process and ultimately facilitating shared decision making. In the following of this chapter we will focus on the application of the architecture we just presented to the implementation of a prototype of a recommender system dedicate to the prediction of UCs for the AF health state.

5.2.2.3. Data sources

One of the most important preliminary steps for the implementation of a recommender system is to populate the utility matrix, i.e. the database collecting the preferences of a set of individuals (see chapter 3 for further details). After that, the recommender system will be able to predict missing values in the utility matrix based on the preferences of similar individuals. In our case, UCs will be predicted for health states for which the currently active patient wasn't able to provide personal values.

The first data source that was used as an input to build such a database is the Euro Hearth Survey (EHS). EHS is a multi-center prospective observational study among cardiology practices in ESC (European Society of Cardiology) member countries aimed at describing current AF management practices and verifying them against guidelines [169]. The EHS included data on 5333 inpatients or outpatients, aged more than 18 years, who were referred to 182 university, non-university, and specialized

hospitals among 35 member countries of the ESC, with an ECG or holter-proven diagnosis of atrial fibrillation during the qualifying admission visit or in the preceding year. Among enrolled patients, 978 only had a single detected episode of AF (i.e. a more precise diagnosis was yet to be provided), 1517 had a diagnosis of paroxysmal AF, 1167 persistent AF, and 1547 permanent AF. Data from the EHS have proven to be useful to investigate the QoL related to the condition of AF [170] and to evaluate the trade-off between risk of stroke and risk of bleeding in the selection of oral anticoagulant treatment in AF patients [171]. Both of these objectives combine in one of the shared decision implementations we developed in MobiGuide (see section 5.1.2.2) and thus make EHS data particularly relevant to our preferences estimation task.

A second source of data comes from the MobiGuide pilot trial on AF patients. During the 3 months of duration of the study enrolled patients were asked to provide data about their QoL periodically answering the EuroQoL (at the enrollment and unenrollment, as well as once a month during the main phase of the trial) and AFEQT (only at enrollment and end of study) questionnaires. EuroQoL scores can be transformed to proper UCs through validated procedures, and thus provide a good source of additional data about AF-related patient preferences. A subset of the patients enrolled in the MobiGuide study also participated in the pilot experiments we reported in section 5.2.1.5 for the validation of the UceWeb tool. The elicited UCs were also collected to populate the recommender system database.

Finally we also explored the possibility of taking into account less structured data sources in the form of free-text patient-generated content reporting personal experiences and opinions. Several different data sources can potentially provide data about patient experiences and QoL (e.g. Twitter and other social networks, patient communities implementing a forum) and many of those have been used in published works [91, 93, 94, 114]. The following section is dedicated to the use of this kind of unstructured data for the purposes of the UCs recommender system dedicated to AF.

5.2.2.4. From text to UCs

We already pointed out in chapter 3 how health information seeking on online review sites, communities, social networks and discussion boards is nowadays a common behavior for people facing medical decisions. The availability and relevance of this kind of patient-generated information provides a unique opportunity also for QoL and shared decision making research to gain further insights on patient preferences and behaviors. For these reasons a NLP and sentiment analysis module (Figure 46) was developed to take advantage of these data sources in the task of building a recommender system for health-related utilities.

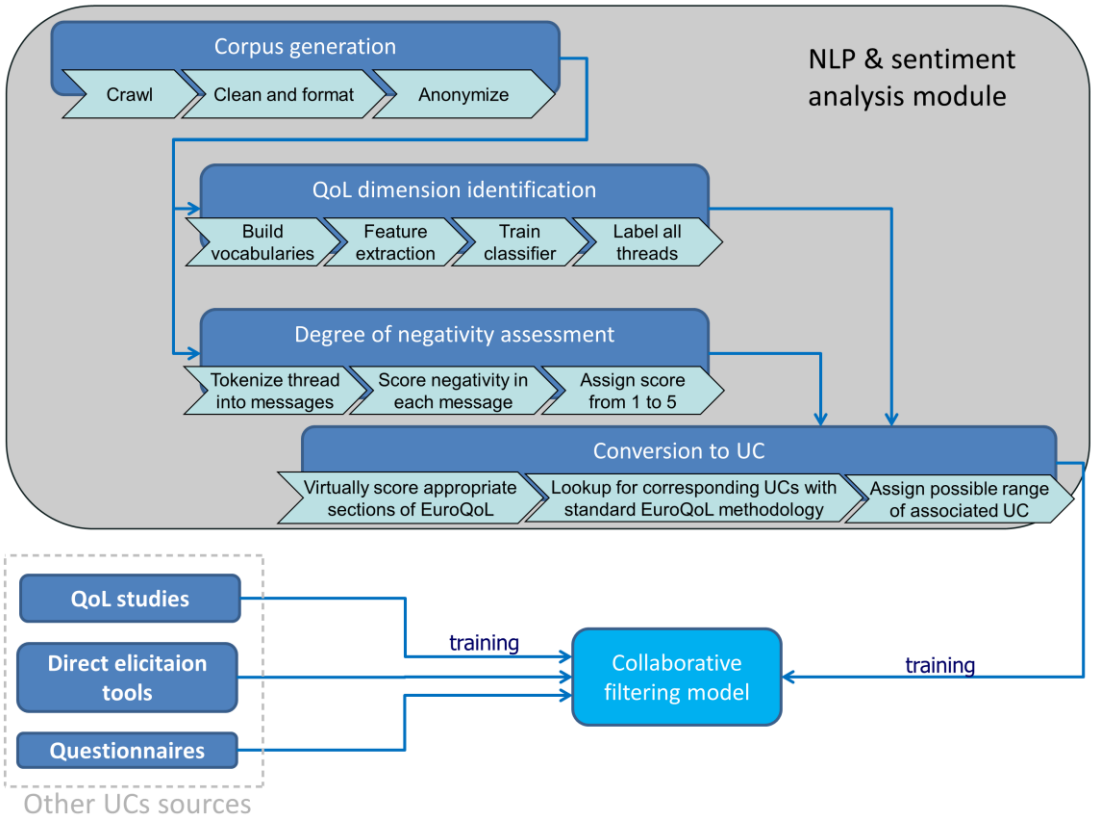


Figure 46 – Details of the logical building blocks of the NLP and sentiment analysis module. This module is responsible for converting free-text reports written by patients to proper UCs that can be exploited by the collaborative filtering module (adapted from [172]).

Figure 46 shows the functional processing steps we designed to extract UCs starting from patient experiences in the form of natural language texts. Even though several different data sources can potentially be mined to collect data about patients' experiences and quality of life (QoL), here we limit our scope to disease-specific discussion boards. The first step of the pipeline is dedicated to building a corpus which consists in a collection of AF-related discussion threads. In the chosen use-case regarding AF, the Medpie framework [173] was used to collect a de-identified corpus of 3757 AF-related threads for a total of approximately 30000 messages. Each thread then undergoes two processing steps:

- QoL dimension identification: intended to identify those threads where at least one of the 5 QoL dimensions considered in the EuroQol questionnaire is present. In fact EuroQoL evaluates the impact on the QoL experienced by the patient of five relevant dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) and, given a negativity score of 1-5 to each of them, is able to output a UC with a validated formula [78]. A custom Python information extraction module based on nltk (Natural

Language Toolkit) has been developed to mine the corpus and identify threads related to the different QoL dimensions. A mixed approach based on text classification algorithms and regular expressions has been used to cope with the very low signal-to-noise ratio present in the corpus.

- Degree of negativity assessment: whose purpose is to assess, by the means of sentiment analysis techniques, the degree of negativity associated to each of the identified QoL dimensions. For the sentiment analysis of messages we rely on Sentistrength [174, 175], a popular lexicon-based sentiment analysis tool developed in Java. Sentistrength has the ability to assess positivity and negativity on a 1 to 5 intensity scale (which is a natural match for the EuroQoL scores) and to effectively cope with the particular lexicon used in web-based conversation. The tool also features a set of optimization procedures [176] that make it easily customizable to the specific medical domain we are considering.

The two information gathered from the steps described above, when combined, allow to virtually score the corresponding EuroQol questionnaire and finally perform the conversion to a proper UC using the standard EuroQol index value calculation algorithm. It is obviously unlikely that the system would be able to score all the 5 dimensions of the EuroQol questionnaire for each single patient only looking at what he reported on social media [168]. However, scoring the questionnaire even partially will allow us to derive at least the boundaries of the interval where the actual UC is contained. These observations, although less precise than UCs directly elicited from a patient, will still be a valuable resource to enrich the database of preferences that the recommender system uses.

5.2.2.5. Collaborative filtering model

In chapter 3 we presented a number of different approaches that are nowadays used for the implementation of recommender systems. This section presents the application of one of them, namely neighborhood based collaborative filtering, to our use case of building a recommender system able to predict health-related utilities. Some approaches to the task of estimating utilities using predictive models have been described in previously published literature [177]. However, to our best knowledge, this is the first attempt to build a recommender system for utilities based on a collaborative filtering methodology.

The first prerequisite for designing such a system is to repurpose the utility matrix, usually featuring items and users, to include patients and health states. Table 8 shows the solution to this first task: users (on columns) can be, quite straightforwardly, transformed to patients while items (on rows) are substituted by health states. This second substitution in particular, although less intuitive than the user-patient one, is an essential

step of the proposed methodology. In fact, having now patients and health states defining the structure of the utility matrix, we can assign each value $r(i,j)$ of the matrix to be equal to the UC that user j associates to health state i. Note that, with this approach, we are still respecting the basic assumptions behind the utility matrix concept, i.e. it should store the user-item (in our case patient-state) preferences.

Table 8 - Adaptation of the utility matrix to our UCs prediction problem. Each column represents a single patient while rows represent different health states. Values contained in the matrix are previously known UCs for each patient-state pair.

Health state	Patient 1	Patient 2	Patient 3
AF	0.8	0.9	?
Stroke mild	0.6	?	0.6
Stroke severe	0.4	0.1	?

Having the utility matrix in place allows us to now move on to the problem of selecting a suitable approach for building the actual recommendation algorithm. Considered the difficulties of defining a proper set of features that could appropriately represent each of the health states in terms of content, following a content-based approach would have been very challenging. On the other hand a CF approach, where we could disregard the need of defining the content of each health state as a finite-length feature vector, has been selected as a better fit for our use case. In particular a neighborhood-based approach where the prediction is calculated using the preferences of a set of the K most similar patients was chosen.

One of the most important steps for the identification of the K-most-similar patients is the definition of a proper similarity measure. In most traditional approaches to CF (see chapter 3) two similar users are defined as users that have similar preferences, i.e. two users that have rated a set of corresponding items with similar scores. This is the preferred approach when the only available data about each user is the content of the utility matrix itself. However in the case of medical applications we can assume that, apart from data regarding UCs, a wider range of clinically relevant variables are also collected for each patient. Availability of this data to the CF model allows to define a richer similarity function that also takes into account the clinical context of each patient. In particular a recent study [170] conducted on patients with AF investigated the influence of a wide set of clinical variables on perceived QoL. For each patient j we thus define a profile vector ϑ_j that includes a set of demographic and clinical variables (refer to Table 9 for the full list) that have been demonstrated to be relevant to QoL. Finally, the similarity between two patients has been defined using the cosine distance calculated between the profile vectors:

$$similarity(patient\ a, patient\ b) = \frac{\vartheta_a \cdot \vartheta_b}{\|\vartheta_a\| \|\vartheta_b\|} \quad (1)$$

Table 9 - Features of the profile vector ϑ used to characterize patients and compute their similarity. Correlation coefficients between each variable and UCs for AF can be found in Berg et al [170].

Feature	Description
Age	Age of the patient at the time of QoL assessment (in years)
Gender	Male/female
Geographical region	Western Europe/Mediterranean/Central Europe
Regular exercise	Never exercise, occasionally exercise, exercise regularly < 3 times/week, exercise regularly 3-5 times/week, exercise > 5 times/week
BMI	Body Mass Index
Underlying heart diseases	Number of cardiac co-morbidities (miocardial infarction, angina pectoris, valvular heart disease, congestive heart failure, cardiomyopathy, congenital heart disease, sick sinus syndrome, sustained ventricular tachicardia, ventricular fibrillation)
Co-morbidities	Number of relevant co-morbidities (COPD, Malignacy, Peripheral vascular disease, Chronic renal failure, Hypo/Hyperthyroidism)
Diabetes	Diagnosis of Diabetes Mellitus
Stroke	Previous stroke episodes (hemorragic or ischaemic)
Current symptoms	Palpitations, chest pain, shortness of breath, syncope, dizziness, fatigue, other
AF type	Paroxysmal, Persistent, Permanent or considered cured
Medical treatment	Divided by drug class: OAT, antiplatelet drugs, heparin, renin-angiotensin-aldosteron system blockers, Class I antiarrhythmics, Class II antiarrhythmics, rate control drugs)
Interventions	Number of interventions (including ablations, ICD implantation, CABG, PCI, Pacemaker implantation, valve replacement and other cardiac interventions)

It is worth mentioning that the just described approach for the definition of a similarity measure assumes that a wide range of clinical and demographic variables can be collected. However this is not always the case for the data coming from the unstructured sources processed by the NLP module described in section 5.2.2.4. Data gathered from various social platforms like medical discussion boards are far less complete than those coming from clinical trials or EHRs. For example, user profiles collecting demographic and general information are completed by patients themselves when registering to the platform, which often results in blank or incomplete data. Moreover, when collecting a corpus from these data sources for research purposes, strict privacy regulations have to be applied [157] and de-identification of the data often prevents the effective mining of patient

profiles. However a number of specific techniques based on network analysis have been developed to address the problem of calculating similarity among the users of health-related online communities [178, 179]. Integration of these alternative similarity calculation methodologies into our prototype of the system could be an interesting future development of the current approach.

After having defined the similarity measure the final step is to create the actual recommendation algorithm. As we previously mentioned we chose a neighborhood based CF approach where the prediction is based on the preferences of the set of the K most similar patients. In particular, for our target patient p and health state h, we define the predicted value of $UC_{h,p}$ as the weighted mean (where the weight is the similarity between patient p and patient j, as defined in equation (1)) of the UCs given to h by his K most similar patients:

$$UC_{h,p} = \frac{\sum_{j=1}^{K} sim(p,j)\, UC_{h,j}}{\sum_{j=1}^{K} sim(p,j)}$$

5.2.2.6. Evaluation on AF patients

This section reports some results from a proof-of-concept implementation of the recommender system described in the previous sections. Data regarding QoL gathered during the MobiGuide pilot trial (using EuroQoL) is used as the gold standard against which we validate our approach. This validation effort, albeit limited in terms of number of patients, aims at demonstrating the feasibility of our UC recommender system in a real-world application involving patients with AF. We used the EHS data to populate the utility matrix and used the previously described similarity measure to determine the 50 most similar patients to each of the available MobiGuide patients. Eight out of the 10 patients enrolled in MobiGuide had the correct inclusion criteria to be used in this validation (of the 2 excluded one never measured QoL during the study and the other did not actually have AF but was instead diagnosed with supraventricular tachycardia during the pilot trial). For each of them the CF algorithm described in section 5.2.2.5 was applied to predict the UC this patient would assign to the AF health state. Finally, in the effort to evaluate the accuracy of the estimate, we compared the predicted value with the actual UC elicited from the patients using EuroQoL. Table 10 summarises the results of this analysis.

One first comment to the evaluation results is that the CF approach can't be capable of capturing the high variability of a UC rapidly evolving over time. Some cases highlight how UCs for the same health state and for the same patient can vary significantly over a relatively short period (e.g. patient 1a3f had a UC of 0,768 at the enrollment while 0,423 at the end of the study). In fact the task of predicting a UC using CF can only leverage

on the historical data collected in the utility matrix and does not have access to detailed data about the specific moment the patient is experiencing (e.g. a high level of anxiety, even if temporary, can cause the UC measured with EuroQoL to drop significantly). Thus the fact that the model would be able to effectively follow these kind of variations is an incorrect assumption. Instead, the predicted UC has to be regarded as a value that would be close to the average UC assigned by the patient to that health state, with several elicitations over time. For these reasons the UCs predicted by the CF model have been compared to the average of the UCs elicited by each patient (UC avg). The caption of Table 10 reports that the median[2] of these errors is smaller than the inter-quartile range calculated on the UCs elicited using EuroQoL. The sample size on which this evaluation was performed is rather small including only 8 patients and further evaluation efforts are needed to assess the accuracy of the prediction. However these preliminary results confirm that CF might be a feasible approach to elicit UCs when traditional elicitation methods fail.

Table 10 – Results of the evaluation of the CF algorithm predicting UC for AF. The MGID column reports (part of) the id of the patient in the MobiGuide system. Median of the errors is 0,128 while inter-quartile range of the average UCs is 0,241.

	MobiGuide				
MGID	UC baseline	UC end-of-study	UC avg	Predicted UC	Error (avg-predicted)
8c85…	1	1	1	0,902	0,098
4751…	0,2	0,391	0,2955	0,796	0,501
7815…	1	0,9	0,95	0,778	0,172
d177…	0,837	1	0,9185	0,856	0,063
dd74…	0,627	0,803	0,715	0,857	0,142
012e…	0,837	0,757	0,797	0,913	0,116
1a3f…	0,768	0,423	0,5955	0,798	0,202
562e…	0,837	n.a.	0,837	0,825	0,012

[2] The median was preferred to the mean as an overall evaluation metric to avoid the effect of the outliers (e.g. patient 4751), given the low number of patients in the validation sample.

Chapter 6

Conclusions and future work

The aim of the work described in this dissertation was to investigate possible solutions for the management and effective use of the great amount and variety of data about patients available nowadays with the final goal of providing a personalized guidance to patients and patient-specific decision support for their physicians. Clinical decision support systems (CDSSs) have been identified as some of the most promising tools able to facilitate the implementation of these objectives both in the clinical practice and in the home settings of the patients. The main challenges to be addressed in pursuing such goals are the heterogeneity of the data sources to be considered by the CDSS and the important requirement of a "personalized" decision support that properly fits the preferences, lifestyle, personal and clinical characteristics of each patient.

To face the first challenge different data-integration methodologies for clinical data were considered and a personal health record data model based on the HL7 virtual medical record standard was developed. On the other hand, regarding the challenge of providing personalized decision support, a shared decision making framework was designed along with techniques that enable the customization of CDSSs to a specific patient context and preferences.

The abovementioned methodologies have been applied to the development of a ubiquitous, guideline-based, patient guidance system in the context of a 4-year European project named MobiGuide. In particular the vMR-based data model has been successfully used to build an integrated personal health record (PHR) which collects data coming from the hospital EMRs, patient inputs, data coming from patient operated sensors and the outputs of the guideline execution engine in the form of clinical recommendations. The PHR constitutes one of the fundamental building blocks of the entire MobiGuide system and its vMR-based design has proven to be an effective solution to provide support for some important characteristics of the final system: flexibility of the data model to

support different clinical domains, possibility to easily integrate different data sources, support for standard medical ontologies, possibility to extend the standard model to represent entities that are peculiar to the MobiGuide system (without impacting the overall standard paradigm of vMR) and support for a distributed architecture and workflow.

The shared decision making framework was also integrated in the atrial fibrillation (AF) domain of the MobiGuide system, choosing decision analysis as a methodological framework for the implementation. Two shared decision points regarding oral anticoagulant therapy and ablation were identified in the AF clinical guideline and the corresponding decision tree models were developed. The decision models were also developed in a way such that patient-specific personalization of the decision analysis was possible. In particular QALYs were chosen as the outcome to consider, with the final goal to maximize survival while also considering quality of life, when choosing among different treatment options.

To this end a specific tool for the computer-assisted elicitation of utility coefficients from patients was developed as a by-product of the shared decision making implementation. Elicitation of utilities from patients and their use in the decision tree models proved to be an effective way of personalizing the otherwise generic decision models and to make them a valuable tool to take better informed decision even in single-patient-level scenarios (since decision trees are most popular in population-level studies) where shared decision making is desirable.

Finally further research was conducted to face the still unresolved challenges that arise when direct utility elicitation fails to perform well. For these purpose a recommender system able to predict the value of utilities given a specific patient profile was designed. A prototype implementation of this system based on collaborative filtering has been presented in the dissertation. In particular the system capitalizes on the preferences of other similar patients to predict an expected value for the still unknown utility coefficients of a new patient. Both the utility elicitation tool and the recommender system for utilities have been evaluated using a set of AF patients enrolled in the MobiGuide pilot trial, in the effort to demonstrate the validity of the proposed approaches.

Future directions of the research described in this dissertation include continuing to develop the UceWeb tool, increasing the size of collected data, to build a collaborative repository of utility coefficients to be used in quality of life research and cost-utility studies. Moreover it would be useful to incorporate a more extensive catalogue of decision models that represent specific clinical decision problems in addition to the two decision trees developed in this work. On the recommender system side, a more comprehensive validation of the performance of the single components would be desirable since the evaluation performed so far was more of a proof-of-concept validation of the overall approach than a proper performance assessment. Finally, it would be interesting to investigate proper ways to integrate one or more knowledge representation formalisms

into the PHR. This would allow to store in a single centralized repository not only the clinical data of the patients but also the formalization of the clinical knowledge (e.g. guidelines) needed by the CDSS, limiting labor-intensive knowledge-to-data mapping processes and ultimately facilitating integration and extensibility of the system. One of the most promising approaches to this challenge consists in the openEHR family of standards and, in particular, its guideline definition language (GDL) whose specification is currently in trial state [180].

References

[1] B. M. Welch and K. Kawamoto, "Clinical decision support for genetically guided personalized medicine: a systematic review," *Journal of the American Medical Informatics Association*, vol. 20, no. 2, pp. 388–400, Mar. 2013.

[2] E. A. Balas and S. A. Boren, "Managing clinical knowledge for health care improvement," *Yearbook of medical informatics*, vol. 2000, no. 2000, pp. 65–70, 2000.

[3] R. A. Greenes, "Chapter 1 - Definition, Scope, and Challenges," in *Clinical Decision Support (Second Edition)*, R. A. Greenes, Ed. Oxford: Academic Press, 2014, pp. 3–47.

[4] J. Osheroff, J. Teich, D. Levick, L. Saldana, F. Velasco, D. Sittig, K. Rogers, and R. Jenders, *Improving Outcomes with Clinical Decision Support: An Implementer's Guide, Second Edition*, 2 edition. Chicago, IL: CRC Press, 2012.

[5] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," *BMJ*, vol. 330, no. 7494, p. 765, Mar. 2005.

[6] K. Cresswell, A. Majeed, D. W. Bates, and A. Sheikh, "Computerised decision support systems for healthcare professionals: an interpretative review," *Inform Prim Care*, vol. 20, no. 2, pp. 115–128, 2012.

[7] P. S. Roshanov, N. Fernandes, J. M. Wilczynski, B. J. Hemens, J. J. You, S. M. Handler, R. Nieuwlaat, N. M. Souza, J. Beyene, H. G. C. Van Spall, A. X. Garg, and R. B. Haynes, "Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials," *BMJ*, vol. 346, p. f657, 2013.

[8] M. W. M. Jaspers, M. Smeulers, H. Vermeulen, and L. W. Peute, "Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings," *J Am Med Inform Assoc*, vol. 18, no. 3, pp. 327–334, May 2011.

[9] T. J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R. R. Coeytaux, G. Samsa, V. Hasselblad, J. W. Williams, M. D. Musty, L. Wing, A. S. Kendrick, G. D. Sanders, and D. Lobach, "Effect of clinical decision-support systems: a systematic review," *Ann. Intern. Med.*, vol. 157, no. 1, pp. 29–43, Jul. 2012.

[10] Garg AX, Adhikari NJ, McDonald H, and et al, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review," *JAMA*, vol. 293, no. 10, pp. 1223–1238, Mar. 2005.

[11] E. H. Shortliffe, R. Davis, S. G. Axline, B. G. Buchanan, C. C. Green, and S. N. Cohen, "Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system," *Comput. Biomed. Res.*, vol. 8, no. 4, pp. 303–320, Aug. 1975.

[12] S. G. Pauker, G. A. Gorry, J. P. Kassirer, and W. B. Schwartz, "Towards the simulation of clinical cognition. Taking a present illness by computer," *Am. J. Med.*, vol. 60, no. 7, pp. 981–996, Jun. 1976.

[13] C. A. Kulikowski and S. M. Weiss, "Representation of expert knowledge for consultation: the CASNET and EXPERT projects," *Artificial Intelligence in medicine*, vol. 51, 1982.

[14] P. Gooch and A. Roudsari, "Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems," *J Am Med Inform Assoc*, vol. 18, no. 6, pp. 738–748, Dec. 2011.

[15] "Clinical Practice Guidelines We Can Trust," *Institute of Medicine*. Available at: http://iom.nationalacademies.org /Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx. [Accessed: 22-Aug-2015].

[16] A. Latoszek-Berendsen, H. Tange, H. J. van den Herik, and A. Hasman, "From Clinical Practice Guidelines to Computer-interpretable Guidelines: A Literature Overview," *Methods of Information in Medicine*, vol. 49, no. 6, pp. 550–570, Nov. 2010.

[17] M. Peleg, "Computer-interpretable clinical guidelines: A methodological review," *Journal of Biomedical Informatics*, vol. 46, no. 4, pp. 744–763, Aug. 2013.

[18] E. Abrahams and M. Silver, "The History of Personalized Medicine," in *Integrative Neuroscience and Personalized Medicine*, E. Gordon, MD, PhD and S. Koslow, PhD, Eds. Oxford University Press, 2010, pp. 3–16.

[19] FDA, "Paving the Way for Personalized Medicine," *www.fda.gov*, Oct-2013. Available at: http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/PersonalizedMedicine/UCM372421.pdf. [Accessed: 25-Nov-2015].

[20] T. M. Burton, J. D. Rockoff, and R. Winslow, "Obama Announces $215 Million Precision-Medicine Genetic Plan," *Wall Street Journal*, 31-Jan-2015.

[21] "FACT SHEET: President Obama's Precision Medicine Initiative," *whitehouse.gov*. Available at: https://www.whitehouse.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative. [Accessed: 13-Jul-2015].

[22] F. S. Collins and H. Varmus, "A New Initiative on Precision Medicine," *New England Journal of Medicine*, vol. 372, no. 9, pp. 793–795, Feb. 2015.

[23] "Precision Medicine Initiative," *The White House*. Available at: https://www.whitehouse.gov/node/333101. [Accessed: 13-Jul-2015].

[24] National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease, *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington (DC): National Academies Press (US), 2011.

[25] J. L. Jameson and D. L. Longo, "Precision medicine--personalized, problematic, and promising," *N. Engl. J. Med.*, vol. 372, no. 23, pp. 2229–2234, Jun. 2015.

[26] M. Soda, Y. L. Choi, M. Enomoto, S. Takada, Y. Yamashita, S. Ishikawa, S. Fujiwara, H. Watanabe, K. Kurashina, H. Hatanaka, M. Bando, S. Ohno, Y. Ishikawa, H. Aburatani, T. Niki, Y. Sohara, Y. Sugiyama, and H. Mano, "Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer," *Nature*, vol. 448, no. 7153, pp. 561–566, Aug. 2007.

[27] A. T. Shaw, B. Y. Yeap, M. Mino-Kenudson, S. R. Digumarthy, D. B. Costa, R. S. Heist, B. Solomon, H. Stubbs, S. Admane, U. McDermott, J. Settleman, S. Kobayashi, E. J. Mark, S. J. Rodig, L. R. Chirieac, E. L. Kwak, T. J. Lynch, and A. J. Iafrate, "Clinical features and outcome of patients with non-small-cell lung cancer who harbor EML4-ALK," *J. Clin. Oncol.*, vol. 27, no. 26, pp. 4247–4253, Sep. 2009.

[28] "All About The Human Genome Project (HGP)." Available at: http://www.genome.gov/10001772. [Accessed: 13-Jul-2015].

[29] E. G. Cerami, B. E. Gross, E. Demir, I. Rodchenkov, Ö. Babur, N. Anwar, N. Schultz, G. D. Bader, and C. Sander, "Pathway Commons, a web resource for biological pathway data," *Nucl. Acids Res.*, vol. 39, no. suppl 1, pp. D685–D690, Jan. 2011.

[30] "BD2K Home Page | Data Science at NIH." Available at: https://datascience.nih.gov/bd2k. [Accessed: 13-Jul-2015].

[31] "History of AMA Ethics." Available at: http://www.ama-assn.org/ama/pub/about-ama/our-history/history-ama-ethics.page? [Accessed: 30-Aug-2015].

[32] E. Topol, *The Patient Will See You Now: The Future of Medicine is in Your Hands*. New York: Basic Books, 2015.

[33] A. Kessler, "The Cheap-Smartphone Revolution," *Wall Street Journal*, 12-May-2014.

[34] "The rise of the cheap smartphone," *The Economist*, Apr-2014.

[35] J. Brubaker, C. Lustig, and G. Hayes, "PatientsLikeMe: empowerment and representation in a patient-centered social network," in *CSCW'10; Workshop on Research in Healthcare: Past, Present, and Future*, 2010.

[36] "WebMD - Better information. Better health.," *WebMD*. Available at: http://www.webmd.com/default.htm. [Accessed: 30-Aug-2015].

[37] D. L. Frosch, S. G. May, K. A. S. Rendle, C. Tietbohl, and G. Elwyn, "Authoritarian physicians and patients' fear of being labeled 'difficult' among key obstacles to shared decision making," *Health Aff (Millwood)*, vol. 31, no. 5, pp. 1030–1038, May 2012.

[38] "Practicing defensive medicine—Not good for patients or physicians." Available at: http://www.aaos.org/news/bulletin/janfeb07/clinical2.asp. [Accessed: 30-Aug-2015].

[39] 2013 By David M. Cutler on September 20, "Why Medicine Will Be More Like Walmart," *MIT Technology Review*. Available at: http://www.technologyreview.com/news/518906/why-medicine-will-be-more-like-walmart/. [Accessed: 30-Aug-2015].

[40] E. Hill, "Smart patients," *The Lancet Oncology*, vol. 15, no. 2, pp. 140–141, Feb. 2014.

[41] S. Quaglini, L. Sacchi, G. Lanzola, and N. Viani, "Personalization and Patient Involvement in Decision Support Systems: Current Trends," *Yearb Med Inform*, vol. 10, no. 1, pp. 106–118, Aug. 2015.

[42] G. J. Downing, S. N. Boyle, K. M. Brinner, and J. A. Osheroff, "Information management to enable personalized medicine: stakeholder roles in building clinical decision support," *BMC Medical Informatics and Decision Making*, vol. 9, no. 1, p. 44, Oct. 2009.

[43] D. Stacey, C. L. Bennett, M. J. Barry, N. F. Col, K. B. Eden, M. Holmes-Rovner, H. Llewellyn-Thomas, A. Lyddiatt, F. Légaré, and R. Thomson, "Decision aids for people facing health treatment or screening decisions," *Cochrane Database Syst Rev*, no. 10, p. CD001431, 2011.

[44] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, and P. C. Tang, "Clinical Decision Support Systems for the Practice of Evidence-based Medicine," *Journal of the American Medical Informatics Association*, vol. 8, no. 6, pp. 527–534, Nov. 2001.

[45] E. Shalom, Y. Shahar, Y. Parmet, and E. Lunenfeld, "A multiple-scenario assessment of the effect of a continuous-care, guideline-based decision support system on clinicians' compliance to clinical guidelines," *International Journal of Medical Informatics*, vol. 84, no. 4, pp. 248–262, Apr. 2015.

[46] L. Ahmadian, M. van Engen-Verheul, F. Bakhshi-Raiez, N. Peek, R. Cornet, and N. F. de Keizer, "The role of standardized data and terminological systems in computerized clinical decision support systems: Literature review and survey," *International Journal of Medical Informatics*, vol. 80, no. 2, pp. 81–93, Feb. 2011.

[47] "epSOS: Home." Available at: http://www.epsos.eu/. [Accessed: 02-Jul-2015].

[48] "Virtual Medical Record - HL7Wiki." Available at: http://wiki.hl7.org/index.php?title=Virtual_Medical_Record. [Accessed: 01-Jul-2015].

[49] P. D. Johnson, S. W. Tu, M. A. Musen, and I. Purves, "A virtual medical record for guideline-based decision support.," *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pp. 294–298, 2001.

[50] K. Kawamoto, G. Del Fiol, D. F. Lobach, and R. A. Jenders, "Standards for scalable clinical decision support: need, current and emerging standards, gaps, and proposal for progress," *Open Med Inform J*, vol. 4, pp. 235–244, 2010.

[51] K. Kawamoto, G. Del Fiol, H. R. Strasberg, N. Hulse, C. Curtis, J. J. Cimino, B. H. Rocha, S. Maviglia, E. Fry, H. J. Scherpbier, V. Huser, P. K. Redington, D. K. Vawdrey, J.-C. Dufour, M. Price, J. H. Weber, T. White, K. S. Hughes, J. C. McClay, C. Wood, K. Eckert, S. Bolte, D. Shields, P. R. Tattam, P. Scott, Z. Liu, and A. K. McIntyre, "Multi-National, Multi-Institutional Analysis of Clinical Decision Support Data Needs to Inform Development of the HL7 Virtual Medical Record Standard," *AMIA Annu Symp Proc*, vol. 2010, pp. 377–381, 2010.

[52] F. A. Sonnenberg and C. G. Hagerty, "Computer-interpretable clinical practice guidelines. Where are we and where are we going ?," *Yearb Med Inform*, pp. 145–158, 2006.

[53] R. A. Jenders and B. Dasgupta, “Challenges in implementing a knowledge editor for the Arden Syntax: knowledge base maintenance and standardization of database linkages,” *Proc AMIA Symp*, pp. 355–359, 2002.

[54] “Standards & Interoperability (S&I) Framework - Health eDecisions Homepage.” Available at: http://wiki.siframework.org/Health+eDecisions+Homepage. [Accessed: 02-Jul-2015].

[55] “QDM | eCQI Resource Center.” Available at: https://ecqi.healthit.gov/qdm. [Accessed: 02-Jul-2015].

[56] “HL7 CDS Standards - HL7Wiki.” Available at: http://wiki.hl7.org/index.php?title=HL7_CDS_Standards. [Accessed: 02-Jul-2015].

[57] D. Kalra, T. Beale, and S. Heard, “The openEHR Foundation,” *Stud Health Technol Inform*, vol. 115, pp. 153–173, 2005.

[58] S. Garde, P. Knaup, E. Hovenga, and S. Herd, “Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes,” *Methods of Information in Medicine*, 2007.

[59] T. Beale, “Archetypes: Constraint-based domain models for future-proof information systems,” in *OOPSLA 2002 workshop on behavioural semantics*, 2002, vol. 105.

[60] “Clinical Knowledge Manager.” Available at: http://www.openehr.org/ckm/. [Accessed: 06-Jul-2015].

[61] P. Muñoz, J. Trigo, I. Martínez, A. Muñoz, J. Escayola, and J. García, “The ISO/EN 13606 Standard for the Interoperable Exchange of Electronic Health Records,” *Journal of Healthcare Engineering*, vol. 2, no. 1, pp. 1–24, Mar. 2011.

[62] P. Schloeffel, T. Beale, G. Hayworth, S. Heard, and H. Leslie, “The Relationship between CEN 13606, HL7, and OpenEHR,” 2006. Available at: http://search.informit.com.au /documentSummary;dn=950616334398351;res=IELHEA. [Accessed: 17-Jul-2015].

[63] M. Eichelberg, T. Aden, J. Riesmeier, A. Dogac, and G. B. Laleci, “A Survey and Analysis of Electronic Healthcare Record Standards,” *ACM Comput. Surv.*, vol. 37, no. 4, pp. 277–315, Dec. 2005.

[64] S. Quaglini, Y. Shahar, M. Peleg, S. Miksch, C. Napolitano, M. Rigla, A. Pallàs, E. Parimbelli, and L. Sacchi, "Supporting shared decision making within the MobiGuide project," in *AMIA Annual Symposium proceedings*, 2013, vol. 2013, pp. 1175–1184.

[65] N. Moumjid, A. Gafni, A. Brémond, and M.-O. Carrère, "Shared decision making in the medical encounter: are we all talking about the same thing?," *Med Decis Making*, vol. 27, no. 5, pp. 539–546, Oct. 2007.

[66] D. Stacey, M. A. Murray, F. Légaré, D. Sandy, P. Menard, and A. O'Connor, "Decision coaching to support shared decision making: a framework, evidence, and implications for nursing practice, education, and policy," *Worldviews Evid Based Nurs*, vol. 5, no. 1, pp. 25–35, 2008.

[67] B. J. McNeil, R. Weichselbaum, and S. G. Pauker, "Speech and survival: tradeoffs between quality and quantity of life in laryngeal cancer," *N. Engl. J. Med.*, vol. 305, no. 17, pp. 982–987, Oct. 1981.

[68] L. Doyal, "Informed consent: moral necessity or illusion?," *Qual Health Care*, vol. 10 Suppl 1, pp. i29–33, Sep. 2001.

[69] R. A. Greenes, Ed., "Section I. Computer-based Clinical Decision Support: Overview, Status, and Challenges," in *Clinical Decision Support (Second Edition)*, Oxford: Academic Press, 2014, pp. 1–2.

[70] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton university press, 2007.

[71] M. C. Weinstein, *Clinical Decision Analysis*, First Edition edition. Philadelphia: W B Saunders Co, 1980.

[72] H. F. Ryder, C. McDonough, A. N. A. Tosteson, and J. D. Lurie, "Decision Analysis and Cost-effectiveness Analysis," *Semin Spine Surg*, vol. 21, no. 4, pp. 216–222, Dec. 2009.

[73] A. S. Detsky, G. Naglie, M. D. Krahn, D. A. Redelmeier, and D. Naimark, "Primer on medical decision analysis: Part 2--Building a tree," *Med Decis Making*, vol. 17, no. 2, pp. 126–135, Jun. 1997.

[74] F. R. Johnson, “Editorial: Moving the QALY forward or just stuck in traffic?,” *Value Health*, vol. 12 Suppl 1, pp. S38–39, Mar. 2009.

[75] M. R. Gold, D. Stevenson, and D. G. Fryback, “HALYs and QALYs and DALYs, Oh My: Similarities and Differences in Summary Measures of Population Health,” *Annual Review of Public Health*, vol. 23, no. 1, pp. 115–134, 2002.

[76] P. Kind, J. E. Lafata, K. Matuszewski, and D. Raisch, “The use of QALYs in clinical and patient decision-making: issues and prospects,” *Value Health*, vol. 12 Suppl 1, pp. S27–30, Mar. 2009.

[77] G. W. Torrance, “Measurement of health state utilities for economic appraisal: A review,” *Journal of Health Economics*, vol. 5, no. 1, pp. 1–30, Mar. 1986.

[78] R. Rabin and F. de Charro, “EQ-SD: a measure of health status from the EuroQol Group,” *Ann Med*, vol. 33, no. 5, pp. 337–343, Jan. 2001.

[79] “EuroQol - Home.” Available at: http://www.euroqol.org/. [Accessed: 07-Aug-2015].

[80] C. Jenkinson, A. Coulter, and L. Wright, “Short form 36 (SF36) health survey questionnaire: normative data for adults of working age.,” *BMJ*, vol. 306, no. 6890, pp. 1437–1440, May 1993.

[81] J. Brazier, J. Roberts, and M. Deverill, “The estimation of a preference-based measure of health from the SF-36,” *J Health Econ*, vol. 21, no. 2, pp. 271–292, Mar. 2002.

[82] G. W. Torrance, D. Feeny, and W. Furlong, “Visual analog scales: do they have a role in the measurement of preferences for health states?,” *Med Decis Making*, vol. 21, no. 4, pp. 329–334, Aug. 2001.

[83] A. Gafni, “The standard gamble method: what is being measured and how it is interpreted.,” *Health Serv Res*, vol. 29, no. 2, pp. 207–224, Jun. 1994.

[84] J. K. Buckingham, J. Birdsall, and J. G. Douglas, “Comparing three versions of the time tradeoff: time for a change?,” *Med Decis Making*, vol. 16, no. 4, pp. 335–347, Dec. 1996.

[85] F. A. Sonnenberg and J. R. Beck, “Markov models in medical decision making: a practical guide,” *Med Decis Making*, vol. 13, no. 4, pp. 322–338, Dec. 1993.

[86] L. Eaton, “Europeans and Americans turn to internet for health information,” *BMJ: British Medical Journal*, vol. 325, no. 7371, p. 989, 2002.

[87] C. A. Brownstein, J. S. Brownstein, D. S. Williams, P. Wicks, and J. A. Heywood, “The power of social networking in medicine,” *Nat Biotech*, vol. 27, no. 10, pp. 888–890, Oct. 2009.

[88] “Live better, together! | PatientsLikeMe.” Available at: http://www.patientslikeme.com/. [Accessed: 01-May-2014].

[89] “Treato Pharma: Brand intelligence from patient online discussions.” Available at: http://treato.com/apps/pharma/treato-service.htm. [Accessed: 01-May-2014].

[90] “Online Support Groups and Forums at DailyStrength.” Available at: http://www.dailystrength.org/. [Accessed: 07-Oct-2015].

[91] G. Eysenbach, “Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness,” *J Med Internet Res*, vol. 10, no. 3, Aug. 2008.

[92] F. J. Grajales 3rd, S. Sheps, K. Ho, H. Novak-Lauscher, and G. Eysenbach, “Social media: a review and tutorial of applications in medicine and health care,” *J. Med. Internet Res.*, vol. 16, no. 2, p. e13, 2014.

[93] N. Kamal, S. Fels, and K. Ho, “Online Social Networks for Personal Informatics to Promote Positive Health Behavior,” in *Proceedings of Second ACM SIGMM Workshop on Social Media*, New York, NY, USA, 2010, pp. 47–52.

[94] J. H. Frost and M. P. Massagli, “Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another’s data,” *J. Med. Internet Res.*, vol. 10, no. 3, p. e15, 2008.

[95] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Foundations and trends in information retrieval*, vol. 2, no. 1–2, pp. 1–135, Jan. 2008.

[96] S. Das and M. Chen, “Yahoo! for Amazon: Extracting market sentiment from stock message boards,” in *Proceedings of the Asia Pacific finance association annual conference (APFA)*, 2001, vol. 35, p. 43.

[97] R. M. Tong, “An operational system for detecting and tracking opinions in on-line discussion,” in *Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification*, 2001, vol. 1, p. 6.

[98] K. Dave, S. Lawrence, and D. M. Pennock, “Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews,” in *Proceedings of the 12th International Conference on World Wide Web*, New York, NY, USA, 2003, pp. 519–528.

[99] T. Mullen and R. Malouf, “A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse.,” in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2006, pp. 159–162.

[100] L. Rainie and J. Horrigan, “Election 2006 online,” *Pew Internet & American Life Project Report*, 2007.

[101] E. X. Zhang, Y. Yang, R. Di Shang, J. J. P. Simons, B. K. Quek, X. F. Yin, W. See, O. S. H. Oh, K. S. T. Nandar, V. R. Y. Ling, P. P. Chan, Z. Wang, R. S. M. Goh, L. James, and J. S. H. Tey, “Leveraging social networking sites for disease surveillance and public sensing: the case of the 2013 avian influenza A(H7N9) outbreak in China,” *Western Pac Surveill Response J*, vol. 6, no. 2, pp. 66–72, Jun. 2015.

[102] T. H. McCoy, V. M. Castro, A. Cagan, A. M. Roberson, I. S. Kohane, and R. H. Perlis, “Sentiment Measured in Hospital Discharge Notes Is Associated with Readmission and Mortality Risk: An Electronic Health Record Study,” *PLoS ONE*, vol. 10, no. 8, p. e0136341, 2015.

[103] S. Noferesti and M. Shamsfard, “Using Linked Data for polarity classification of patients’ experiences,” *J Biomed Inform*, Jul. 2015.

[104] H. Cole-Lewis, A. Varghese, A. Sanders, M. Schwarz, J. Pugatch, and E. Augustson, “Assessing Electronic Cigarette-Related Tweets for Sentiment and Content Using Supervised Machine Learning,” *J. Med. Internet Res.*, vol. 17, no. 8, p. e208, 2015.

[105] P. Cavazos-Rehg, M. Krauss, R. Grucza, and L. Bierut, "Characterizing the followers and tweets of a marijuana-focused Twitter handle," *J. Med. Internet Res.*, vol. 16, no. 6, p. e157, 2014.

[106] N. Jindal and B. Liu, "Opinion spam and analysis," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*, 2008, pp. 219–230.

[107] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[108] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th annual meeting on association for computational linguistics*, 2002, pp. 417–424.

[109] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.," in *LREC*, 2010, vol. 10, pp. 2200–2204.

[110] R. Agerri and A. García-Serrano, "Q-WordNet: Extracting Polarity from WordNet Senses.," in *LREC*, 2010.

[111] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[112] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "SentiFul: A Lexicon for Sentiment Analysis," *IEEE Transactions on Affective Computing*, vol. 2, no. 1, pp. 22–36, Jan. 2011.

[113] N. Grabar and L. Dumonet, "Automatic Computing of Global Emotional Polarity in French Health Forum Messages," in *Artificial Intelligence in Medicine*, Cham, 2015, vol. 9105, pp. 243–248.

[114] A. K. Godea, C. Caragea, F. A. Bulgarov, and S. Ramisetty-Mikler, "An Analysis of Twitter Data on E-cigarette Sentiments and Promotion," in *Artificial Intelligence in Medicine*, Cham, 2015, vol. 9105, pp. 205–215.

[115] M. Araújo, P. Gonçalves, M. Cha, and F. Benevenuto, "iFeel: A System That Compares and Combines Sentiment Analysis Methods," in *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, Republic and Canton of Geneva, Switzerland, 2014, pp. 75–78.

[116] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds., *Recommender Systems Handbook*. Boston, MA: Springer US, 2011.

[117] X. Amatriain, "Recommender Systems," presented at the Machine Learning Summer School 2014 @ CMU, Carnegie Mellon University, 2014.

[118] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[119] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence*, vol. 2009, pp. 1–19, 2009.

[120] Y. Ren, G. Li, J. Zhang, and W. Zhou, "Lazy Collaborative Filtering for Data Sets With Missing Values," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1822–1834, Dec. 2013.

[121] R. M. Bell and Y. Koren, "Lessons from the Netflix Prize Challenge," *SIGKDD Explor. Newsl.*, vol. 9, no. 2, pp. 75–79, Dec. 2007.

[122] "MobiGuide - Home." Available at: http://www.mobiguide-project.eu/. [Accessed: 04-May-2014].

[123] M. Peleg, Y. Shahar, and S. Quaglini, "Making healthcare more accessible, better, faster, and cheaper: the MobiGuide Project," *Eur. J. ePractice Issue Mob. eHealth*, vol. 20, pp. 5–20, 2014.

[124] D. Dabelea, "The predisposition to obesity and diabetes in offspring of diabetic mothers," *Diabetes Care*, vol. 30 Suppl 2, pp. S169–174, Jul. 2007.

[125] K. Wasmer and L. Eckardt, "Management of atrial fibrillation around the world: a comparison of current ACCF/AHA/HRS, CCS, and ESC guidelines," *Europace*, vol. 13, no. 10, pp. 1368–1374, Oct. 2011.

[126] E. Shalom, Y. Shahar, M. Taieb-Maimon, G. Bar, A. Yarkoni, O. Young, S. B. Martins, L. Vaszar, M. K. Goldstein, Y. Liel, A. Leibowitz, T. Marom, and E. Lunenfeld, "A quantitative assessment of a methodology for collaborative specification and evaluation of clinical guidelines," *Journal of Biomedical Informatics*, vol. 41, no. 6, pp. 889–903, Dec. 2008.

[127] O. Young, Y. Shahar, Y. Liel, E. Lunenfeld, G. Bar, E. Shalom, S. B. Martins, L. T. Vaszar, T. Marom, and M. K. Goldstein, "Runtime application of Hybrid-Asbru clinical guidelines," *Journal of Biomedical Informatics*, vol. 40, no. 5, pp. 507–526, Oct. 2007.

[128] Y. Shahar, O. Young, E. Shalom, M. Galperin, A. Mayaffit, R. Moskovitch, and A. Hessing, "A framework for a distributed, hybrid, multiple-ontology clinical-guideline library, and automated guideline-support tools," *J Biomed Inform*, vol. 37, no. 5, pp. 325–344, Oct. 2004.

[129] A. Hatsek, Y. Shahar, M. Taieb-Maimon, E. Shalom, D. Klimov, and E. Lunenfeld, "A scalable architecture for incremental specification and maintenance of procedural and declarative clinical decision-support knowledge," *Open Med Inform J*, vol. 4, pp. 255–277, 2010.

[130] D. Boaz and Y. Shahar, "A framework for distributed mediation of temporal-abstraction queries to clinical databases," *Artif Intell Med*, vol. 34, no. 1, pp. 3–24, May 2005.

[131] M. Peleg, S. Keren, and Y. Denekamp, "Mapping computerized clinical guidelines to electronic medical records: Knowledge-data ontological mapper (KDOM)," *Journal of Biomedical Informatics*, vol. 41, no. 1, pp. 180–201, Feb. 2008.

[132] C. Marcos, A. González-Ferrer, M. Peleg, and C. Cavero, "Solving the interoperability challenge of a distributed complex patient guidance system: A data integrator based on HL7's Virtual Medical Record standard," *Journal of the American Medical Informatics Association*, p. ocv003, Apr. 2015.

[133] N. Larburu, R. G. A. Bults, I. Widya, and H. J. Hermens, "Quality of data computational models and telemedicine treatment effects," in *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2014, pp. 364–369.

[134] "OpenAM - Access Management & SSO - Project Home Page," *ForgeRock Community*. .

[135] D. Detmer, M. Bloomrosen, B. Raymond, and P. Tang, "Integrated Personal Health Records: Transformative Tools for Consumer-Centric Care," *BMC Med Inform Decis Mak*, vol. 8, p. 45, Oct. 2008.

[136] A. González-Ferrer, M. Peleg, B. Verhees, J.-M. Verlinden, and C. Marcos, "Data Integration for Clinical Decision Support Based on openEHR Archetypes and HL7 Virtual Medical Record," in *Process Support and Knowledge Representation in Health Care*, vol. 7738, R. Lenz, S. Miksch, M. Peleg, M. Reichert, D. Riaño, and A. ten Teije, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 71–84.

[137] A. González-Ferrer and M. Peleg, "Understanding requirements of clinical data standards for developing interoperable knowledge-based DSS: A case study," *Computer Standards & Interfaces*, vol. 42, pp. 125–136, Nov. 2015.

[138] Arturo González-Ferrer, Mor Peleg, Enea Parimbelli, Erez Shalom, Carlos Marcos Lagunar, Guy Klebanov, Iñaki Martínez-Sarriegui, Nick Lik San Fung, and Tom Broens, "Use of the Virtual Medical Record Data Model for Communication among Components of a Distributed Decision-Support System," in *Proceedings of 2nd IEEE Biomedical and Health Informatics International Conference*, Valencia, 2014.

[139] J. Pathak, J. Wang, S. Kashyap, M. Basford, R. Li, D. R. Masys, and C. G. Chute, "Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience," *J Am Med Inform Assoc*, vol. 18, no. 4, pp. 376–386, Aug. 2011.

[140] E. Shalom, Y. Shahar, A. Goldstein, E. Ariel, S. Quaglini, L. Sacchi, N. Fung, V. Jones, T. Broens, G. García-Sáez, and others, "Enhancing guideline-based decision support with distributed computation through local mobile application," in *Business Process Management Workshops*, 2014, pp. 53–58.

[141] L. Sacchi, A. Fux, C. Napolitano, S. Panzarasa, M. Peleg, S. Quaglini, E. Shalom, P. Soffer, and P. Tormene, "Patient-tailored workflow patterns from clinical practice guidelines recommendations," *Stud Health Technol Inform*, vol. 192, pp. 392–396, 2013.

[142] Y. Shahar, "Dynamic temporal interpretation contexts for temporal abstraction," *Annals of Mathematics and Artificial Intelligence*, vol. 22, no. 1–2, pp. 159–192, Feb. 1998.

[143] B. Chewning, C. Bylund, B. Shah, N. K. Arora, J. A. Gueguen, and G. Makoul, "Patient preferences for shared decisions: A systematic review," *Patient Educ Couns*, vol. 86, no. 1, pp. 9–18, Jan. 2012.

[144] "TreeAge," *TreeAge Software*. Available at: https://www.treeage.com/. [Accessed: 02-Sep-2015].

[145] V. Fuster, L. E. Rydén, D. S. Cannom, H. J. Crijns, A. B. Curtis, K. A. Ellenbogen, J. L. Halperin, G. N. Kay, J.-Y. Le Huezey, J. E. Lowe, S. B. Olsson, E. N. Prystowsky, J. L. Tamargo, L. S. Wann, S. C. Smith, S. G. Priori, N. A. M. Estes, M. D. Ezekowitz, W. M. Jackman, C. T. January, J. E. Lowe, R. L. Page, D. J. Slotwiner, W. G. Stevenson, C. M. Tracy, A. K. Jacobs, J. L. Anderson, N. Albert, C. E. Buller, M. A. Creager, S. M. Ettinger, R. A. Guyton, J. L. Halperin, J. S. Hochman, F. G. Kushner, E. M. Ohman, W. G. Stevenson, L. G. Tarkington, C. W. Yancy, and American College of Cardiology Foundation/American Heart Association Task Force, "2011 ACCF/AHA/HRS focused updates incorporated into the ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines," *Circulation*, vol. 123, no. 10, pp. e269–367, Mar. 2011.

[146] Authors/Task Force Members, A. J. Camm, G. Y. H. Lip, R. De Caterina, I. Savelieva, D. Atar, S. H. Hohnloser, G. Hindricks, P. Kirchhof, ESC Committee for Practice Guidelines (CPG), J. J. Bax, H. Baumgartner, C. Ceconi, V. Dean, C. Deaton, R. Fagard, C. Funck-Brentano, D. Hasdai, A. Hoes, P. Kirchhof, J. Knuuti, P. Kolh, T. McDonagh, C. Moulin, B. A. Popescu, Z. Reiner, U. Sechtem, P. A. Sirnes, M. Tendera, A. Torbicki, A. Vahanian, S. Windecker, Document Reviewers, P. Vardas, N. Al-Attar, O. Alfieri, A. Angelini, C. Blomstrom-Lundqvist, P. Colonna, J. De Sutter, S. Ernst, A. Goette, B. Gorenek, R. Hatala, H. Heidbuchel, M. Heldal, S. D. Kristensen, P. Kolh, J.-Y. Le Heuzey, H. Mavrakis, L. Mont, P. P. Filardi, P. Ponikowski, B. Prendergast, F. H. Rutten, U. Schotten, I. C. Van Gelder, and F. W. A. Verheugt, "2012 focused update of the ESC Guidelines for the management of atrial fibrillation: An update of the 2010 ESC Guidelines for the management of atrial fibrillation * Developed with the special contribution of the European Heart Rhythm Association," *European Heart Journal*, vol. 33, no. 21, pp. 2719–2747, Nov. 2012.

[147] R. W. Smalling, S. J. Zieman, G. N. Levine, P. R. Liebson, D. Mukherjee, E. D. Peterson, M. S. Sabatine, D. R. H. Ganiats Jr, A. S. Jaffe, H. Jneid, and others, "2014 AHA/ACC Guideline for the Management of Patients With Atrial Fibrillation," 2014.

[148] J. L. Anderson, J. L. Halperin, N. M. Albert, B. Bozkurt, R. G. Brindis, L. H. Curtis, D. DeMets, R. A. Guyton, J. S. Hochman, R. J. Kovacs, E. M. Ohman, S. J. Pressler, F. W. Sellke, and W.-K. Shen, "Management of Patients With Atrial Fibrillation (Compilation of 2006 ACCF/AHA/ESC and 2011 ACCF/AHA/HRS Recommendations) A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines," *Circulation*, vol. 127, no. 18, pp. 1916–1926, Jul. 2013.

[149] M. R. Reynolds, P. Zimetbaum, M. E. Josephson, E. Ellis, T. Danilov, and D. J. Cohen, "Cost-effectiveness of radiofrequency catheter ablation compared with antiarrhythmic drug therapy for paroxysmal atrial fibrillation," *Circ Arrhythm Electrophysiol*, vol. 2, no. 4, pp. 362–369, Aug. 2009.

[150] A. Robinson, R. Thomson, D. Parkin, M. Sudlow, and M. Eccles, "How patients with atrial fibrillation value different health outcomes: a standard gamble study," *J Health Serv Res Policy*, vol. 6, no. 2, pp. 92–98, Apr. 2001.

[151] G. J. Povar, "Medical Decision Making: A Physician's Guide," *Annals of Internal Medicine*, vol. 150, no. 5, pp. 364–364, 2009.

[152] S. Rubrichi, C. Rognoni, L. Sacchi, E. Parimbelli, C. Napolitano, A. Mazzanti, and S. Quaglini, "Graphical Representation of Life Paths to Better Convey Results of Decision Models to Patients," *Medical Decision Making*, Jan. 2015.

[153] W. Sumner, R. Nease, and B. Littenberg, "U-titer: a utility assessment tool," *Proc Annu Symp Comput Appl Med Care*, pp. 701–705, 1991.

[154] T. L. Sims, A. M. Garber, D. E. Miller, P. T. Mahlow, D. M. Bravata, and M. K. Goldstein, "Multimedia quality of life assessment: advances with FLAIR," *AMIA Annu Symp Proc*, pp. 694–698, 2005.

[155] L. A. Lenert, A. Sturley, and M. E. Watson, "iMPACT3: Internet-based development and administration of utility elicitation protocols," *Med Decis Making*, vol. 22, no. 6, pp. 464–474, Dec. 2002.

[156] A. Bayoumi and W. Dale, "ProSPEQT: A new program for computer-assisted utility elicitations," in *Poster presented at: Annual Meeting of Society for Medical Decision Making*, 2004.

[157] Mitka M, "NEw hipaa rule aims to improve privacy and security of patient records," *JAMA*, vol. 309, no. 9, pp. 861–862, Mar. 2013.

[158] Office for Civil Rights, HHS, "Standards for privacy of individually identifiable health information. Final rule," *Fed Regist*, vol. 67, no. 157, pp. 53181–53273, Aug. 2002.

[159] J. S. Ancker, E. U. Weber, and R. Kukafka, "Effect of arrangement of stick figures on estimates of proportion in risk graphics," *Med Decis Making*, vol. 31, no. 1, pp. 143–150, Feb. 2011.

[160] T. H. Cormen, *Introduction to algorithms*. MIT press, 2009.

[161] J. Spertus, P. Dorian, R. Bubien, S. Lewis, D. Godejohn, M. R. Reynolds, D. R. Lakkireddy, A. P. Wimmer, A. Bhandari, and C. Burk, "Development and validation of the Atrial Fibrillation Effect on QualiTy-of-Life (AFEQT) Questionnaire in patients with atrial fibrillation," *Circ Arrhythm Electrophysiol*, vol. 4, no. 1, pp. 15–25, Feb. 2011.

[162] S. C. Kirshblum, W. Waring, F. Biering-Sorensen, S. P. Burns, M. Johansen, M. Schmidt-Read, W. Donovan, D. Graves, A. Jha, L. Jones, M. J. Mulcahey, and A. Krassioukov, "Reference for the 2011 revision of the International Standards for Neurological Classification of Spinal Cord Injury," *J Spinal Cord Med*, vol. 34, no. 6, pp. 547–554, Nov. 2011.

[163] E. Parimbelli, L. Sacchi, S. Rubrichi, A. Mazzanti, and S. Quaglini, "UceWeb: a Web-based Collaborative Tool for Collecting and Sharing Quality of Life Data," *Methods Inf Med*, vol. 53, no. 6, Nov. 2014.

[164] F. E. van Nooten, X. Koolman, and W. B. F. Brouwer, "The influence of subjective life expectancy on health state valuations using a 10 year TTO," *Health Econ.*, vol. 18, no. 5, pp. 549–558, May 2009.

[165] L. S. Matza, K. S. Boye, D. H. Feeny, J. A. Johnston, L. Bowman, and J. B. Jordan, "Impact of caregiver and parenting status on time trade-off and standard gamble utility scores for health state descriptions," *Health Qual Life Outcomes*, vol. 12, p. 48, Apr. 2014.

[166] F. E. van Nooten, N. J. A. van Exel, X. Koolman, and W. B. F. Brouwer, "'Married with children' the influence of significant others in TTO exercises," *Health and Quality of Life Outcomes*, vol. 13, no. 1, Dec. 2015.

[167] J. Protheroe, T. Fahey, A. A. Montgomery, and T. J. Peters, "Effects of patients' preferences on the treatment of atrial fibrillation," *West J Med*, vol. 174, no. 5, pp. 311–315, May 2001.

[168] E. Parimbelli, S. Quaglini, C. Napolitano, S. Priori, R. Bellazzi, and J. H. Holmes, "Use of Patient Generated Data from Social Media and Collaborative Filtering for Preferences Elicitation in Shared Decision Making," in *AAAI Fall Symposium Series*, 2014.

[169] R. Nieuwlaat, A. Capucci, A. J. Camm, S. B. Olsson, D. Andresen, D. W. Davies, S. Cobbe, G. Breithardt, J.-Y. L. Heuzey, M. H. Prins, S. Lévy, and H. J. G. M. Crijns, "Atrial fibrillation management: a prospective survey in ESC Member Countries The Euro Heart Survey on Atrial Fibrillation," *Eur Heart J*, vol. 26, no. 22, pp. 2422–2434, Jan. 2005.

[170] J. Berg, P. Lindgren, R. Nieuwlaat, O. Bouin, and H. Crijns, "Factors determining utility measured with the EQ-5D in patients with atrial fibrillation," *Qual Life Res*, vol. 19, no. 3, pp. 381–390, Apr. 2010.

[171] M. Marcucci, G. Y. H. Lip, R. Nieuwlaat, R. Pisters, H. J. G. M. Crijns, and A. Iorio, "Stroke and Bleeding Risk Co-distribution in Real-world Patients with Atrial Fibrillation: The Euro Heart Survey," *The American Journal of Medicine*, vol. 127, no. 10, pp. 979–986.e2, Oct. 2014.

[172] E. Parimbelli, S. Quaglini, R. Bellazzi, and J. H. Holmes, "Collaborative Filtering for Estimating Health Related Utilities in Decision Support Systems," in *Artificial Intelligence in Medicine*, 2015, pp. 106–110.

[173] A. Benton, J. H. Holmes, S. Hill, A. Chung, and L. Ungar, "medpie: an information extraction package for medical message board posts.," *Bioinformatics (Oxford, England)*, vol. 28, no. 5, pp. 743–4, 2012.

[174] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *J. Am. Soc. Inf. Sci.*, vol. 61, no. 12, pp. 2544–2558, Dec. 2010.

[175] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," *J. Am. Soc. Inf. Sci.*, vol. 63, no. 1, pp. 163–173, Jan. 2012.

[176] M. Thelwall, "Heart and soul: Sentiment strength detection in the social web with sentistrength," *Proceedings of the CyberEmotions*, pp. 1–14, 2013.

[177] U. Chajewska, L. Getoor, J. Norman, and Y. Shahar, "Utility Elicitation As a Classification Problem," in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, USA, 1998, pp. 79–88.

[178] L. Jiang and C. C. Yang, “Determining User Similarity in Healthcare Social Media Using Content Similarity and Structural Similarity,” in *Artificial Intelligence in Medicine*, 2015, pp. 216–226.

[179] M. Gan, “Walking on a user similarity network towards personalized recommendations,” *PLoS ONE*, vol. 9, no. 12, p. e114662, 2014.

[180] openEHR Foundation, “Guideline Definition Language (GDL).” Available at: http://www.openehr.org/releases/CDS/latest/docs/GDL/GDL.html. [Accessed: 04-Dec-2015].

Appendix A

Atrial Fibrillation Effect on QualiTy-of-life (AFEQT) Questionnaire

Section 1. Occurrence of atrial fibrillation **Name or ID:**_____

Are you currently in atrial fibrillation? ☐ Yes ☐ No

If **No**, when was the last time you were aware of having had an episode of atrial fibrillation? (Please check one answer which best describes your situation)

__earlier today
__within the past week
__within the past month
__1 month to 1 year ago
__ more than 1 year ago
__I was never aware of having atrial fibrillation

Section 2. The following questions refer to how atrial fibrillation affects your quality of life.

On a scale of 1 to 7, over the past 4 weeks, as a result of your atrial fibrillation, how much were you bothered by:
(Please circle one number which best describes your situation)

	Not at all bothered Or I did not have this symptom	Hardly bothered	A little bothered	Moderately bothered	Quite a bit bothered	Very bothered	Extremely bothered
1. Palpitations: Heart fluttering, skipping or racing	1	2	3	4	5	6	7
2. Irregular heart beat	1	2	3	4	5	6	7
3. A pause in heart activity	1	2	3	4	5	6	7
4. Lightheadedness or dizziness	1	2	3	4	5	6	7

On a scale of 1 to 7, over the past 4 weeks, have you been limited by your atrial fibrillation in your:
(Please circle one number which best describes your situation)

	Not at all limited	Hardly limited	A little limited	Moderately limited	Quite a bit limited	Very limited	Extremely limited
5. Ability to have recreational pastimes, sports, and hobbies	1	2	3	4	5	6	7
6. Ability to have a relationship and do things with friends and family	1	2	3	4	5	6	7

On a scale of 1 to 7, over the past 4 weeks, as a result of your atrial fibrillation, how much difficulty have you had in:
(Please circle one number which best describes your situation)

	No difficulty at all	Hardly any difficulty	A little difficulty	Moderate difficulty	Quite a bit of difficulty	A lot of difficulty	Extreme difficulty
7. Doing any activity because you felt tired, fatigued, or low on energy	1	2	3	4	5	6	7
8. Doing physical activity because of shortness of breath	1	2	3	4	5	6	7
9. Exercising	1	2	3	4	5	6	7
10. Walking briskly	1	2	3	4	5	6	7
11. Walking briskly uphill or carrying groceries or other items, up a flight of stairs without stopping	1	2	3	4	5	6	7
12. Doing vigorous activities such as lifting or moving heavy furniture, running, or participating in strenuous sports like tennis or racquetball	1	2	3	4	5	6	7

Developed by AFEQT Core Team-John Spertus, MD, Mid America Heart Institute, Kansas City, MO; Paul Dorian, MD, St. Michaels Hospital, Toronto, ON; Rosemary Bubien, RN, University of Alabama, Birmingham, AL; Caroline Burk, Pharm D. M.S; Steven Lewis, PhD; Donna Godejohn, BSN, St. Jude Medical, St. Paul, MN.

Atrial Fibrillation Effect on QualiTy-of-life (AFEQT) Questionnaire

On a scale of 1 to 7, over the past 4 weeks as a result of your atrial fibrillation, how much did the feelings below bother you? (Please circle one number which best describes your situation)

	Not at all Bothered	Hardly bothered	A little bothered	Moderately bothered	Quite a bit bothered	Very bothered	Extremely bothered
13. Feeling worried or anxious that your atrial fibrillation can start anytime	1	2	3	4	5	6	7
14. Feeling worried that atrial fibrillation may worsen other medical conditions in the long run	1	2	3	4	5	6	7

On a scale of 1 to 7, over the past 4 weeks, as a result of your atrial fibrillation treatment, how much were you bothered by: (Please circle one number which best describes your situation)

	Not at all bothered	Hardly bothered	A little bothered	Moderately bothered	Quite a bit bothered	Very bothered	Extremely bothered
15. Worrying about the treatment side effects from medications	1	2	3	4	5	6	7
16. Worrying about complications or side effects from procedures like catheter ablation, surgery, or pacemakers therapy	1	2	3	4	5	6	7
17. Worrying about side effects of blood thinners such as nosebleeds, bleeding gums when brushing teeth, heavy bleeding from cuts, or bruising.	1	2	3	4	5	6	7
18. Worrying or feeling anxious that your treatment interferes with your daily activities	1	2	3	4	5	6	7

On a scale of 1 to 7, overall, how satisfied are you **at the present time** with:
(Please circle one number which best describes your situation)

	Extremely satisfied	Very satisfied	Somewhat satisfied	Mixed with satisfied and dissatisfied	Somewhat dissatisfied	Very dissatisfied	Extremely dissatisfied
19. How well your current treatment controls your atrial fibrillation?	1	2	3	4	5	6	7
20. The extent to which treatment has relieved your symptoms of atrial fibrillation?	1	2	3	4	5	6	7

Name or ID:_____

Developed by AFEQT Core Team-John Spertus, MD, Mid America Heart Institute, Kansas City, MO; Paul Dorian, MD, St. Michaels Hospital, Toronto, ON; Rosemary Bubien, RN, University of Alabama, Birmingham, AL; Caroline Burk, Pharm D. M.S; Steven Lewis, PhD; Donna Godejohn, BSN, St. Jude Medical, St. Paul, MN.