

This is a post-print of the accepted version.

## UceWeb: a web-based collaborative tool for collecting and sharing quality of life data

Enea Parimbelli<sup>1</sup>, Lucia Sacchi<sup>1</sup>, Stefania Rubrichi<sup>1</sup>, Andrea Mazzanti<sup>2</sup>, Silvana Quaglini<sup>1</sup>

<sup>1</sup>Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy

<sup>2</sup>IRCCS Salvatore Maugeri Foundation, Pavia, Italy

### Summary

*Objectives* This work aims at building a platform where quality-of-life data, namely utility coefficients, can be elicited not only for immediate use, but also systematically stored together with patient profiles to build a public repository to be further exploited in studies on specific target populations (e.g. cost/utility analyses).

*Methods* We capitalized on utility theory and previous experience to define a set of desirable features such a tool should show to facilitate sound elicitation of quality of life. A set of visualization tools and algorithms has been developed to this purpose. To make it easily accessible for potential users, the software has been designed as a web application. A pilot validation study has been performed on 20 atrial fibrillation patients.

*Results* A collaborative platform, UceWeb, has been developed and tested. It implements the standard gamble, time trade-off and rating-scale utility elicitation methods. It allows doctors and patients to choose the mode of interaction to maximize patients' comfort in answering difficult questions. Every utility elicitation may contribute to the growth of the repository.

*Conclusion* UceWeb can 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.

**Keywords:** utility elicitation, quality of life, web-application, collaborative platform, repository

### Introduction

Although they have been criticized by a number of authors and some alternative proposals have arisen in the literature [1, 2, 3], Quality Adjusted Life Years (QALYs) [4] 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 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. Thus, they can be proficiently used in decision problems focused on a patient's perspective, for instance within a shared-decision making framework. In the EU-funded project MobiGuide, we use UCs to incorporate patients' preferences into decision models as a way to personalize the process of guideline-based clinical decision making [5]. However, 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. Given the above mentioned subjectivity of UCs, population studies require the definition of summary values derived by appropriately synthesizing the coefficients collected on a population. In general, collecting UCs on a sufficient number of patients is very time consuming. Thus, to quantify their models, authors of cost/utility analyses often use coefficients available in the literature [6], disregarding that these could have been derived from a population different from theirs, whereas it is nowadays acknowledged that environmental data play an important role in quality of life (QoL) assessment (see [7], and the EuroQOL questionnaire -[www.euroqol.org](http://www.euroqol.org) - which uses different weights in its algorithm, for different European countries).

Using reliable UCs is very important, as healthcare policy makers often use cost/utility analysis to judge whether a new program (e.g. a new drug, a new instrument, etc.) is worth to be introduced in the medical practice, basing the decision on some upper thresholds for ICUR suggested by international organizations (e.g. the World Health Organisation, for most of EU countries, suggests a threshold of 30,439\$ for QALY gained). It is clear that UC values may have a key role in positioning ICUR with respect to that threshold. Using (even slightly) incorrect values, for example values referring to another population, may bias the final decision.

Originally developed within the MobiGuide project to interact with atrial fibrillation patients, our web-based tool for utility elicitation, UceWeb, has currently become a generic, multilingual tool able to support UC elicitation (UCE) for any health state. UceWeb is therefore proposed as an independent research tool to collect and share data and results about quality of life. We recommend it as a facility both for people who want to directly elicit UCs from an individual patient or a set of patients, and for those interested in further basic research on utility elicitation methods. UceWeb provides a structured methodology for systematic collection of utility values, a shared database of elicitation results and a multimedia repository of health states descriptions including text, pictures and videos.

Wide use of UceWeb will lead to a UC repository that could be exploited by authors of cost-utility analyses in the future, providing them with more specific, population-targeted values.

## **Background**

### ***Utility elicitation methods***

UCE methods can be divided into direct and indirect methods. While indirect methods are usually based on questionnaires [8], here we will focus on the three main direct methods, namely Standard Gamble (SG), Time Trade-off (TTO), and Rating Scale (RS).

SG [9] has its theoretical basis in the axioms of Von Neumann-Morgenstern utility theory [10]. 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  $p$ . 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  $(1-p)$ . In TTO, the patient is asked to choose between living his entire remaining life ( $t_1$ ) in the specific health state being evaluated or to live shorter ( $t_2 < t_1$ ) but in a perfect health state. 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  $t_2/t_1$ . Finally, 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 [11]) but rather what it is called a “value”, calculated as  $x/100$  where  $x$  is the marker position on the scale.

### ***Related works***

The idea of building a reference collection of UCs was first proposed in the Beaver Dam Health Outcomes Study [12] in the early 90s, where participant patients responded to several questions with both direct and indirect methods.

However, to our knowledge, after collecting data on about 1400 patients during the first years of the study, that repository was no longer maintained. In the same years, information technology started playing a role in UCE, also with the intention of facilitating the storage and retrieval of the elicited data. One of the first examples of computerized UCE, no longer available to the public, was U-titer [13], developed to achieve a utility assessment methodology as independent as possible from the specific human interviewer. It was a portable desktop application implementing SG, TTO, and RS, which allowed configuring some features like changing labels, setting an extreme health state different from death, and showing or not a graphical help for the user.

Functional Limitation And Independence Rating (FLAIR) is a Macromedia Shockwave web application [14] developed to elicit UCs from older adults suffering from some dependencies in the activities of daily living. Only the SG and Willingness To Pay<sup>1</sup> elicitation methods are implemented. Rich multimedia health state description including pictures and audio are provided to help patients understanding health states they have little personal experience with.

Internet Multimedia Preference Assessor Construction Tool (iMPACT) is a family of tools [15], whose latest product is iMPACT3. It helps researchers and care professionals to build their own utility elicitation program without requiring advanced software development skills. SG, TTO and RS methods are available as well as rich multimedia capabilities for description of health states.

Health Decision Strategy LLC, a healthcare private company, has developed a series of interactive software products ([www.healthstrategy.com](http://www.healthstrategy.com)) related to decision support and cost/effectiveness analysis, including tools for SG and TTO, but they only allow very basic interaction and do not support any multimedia capability.

ProSPEQT (Program to Survey Preferences and Evaluate Quality of life) is a general tool for UCE with the three classical methods, and has been developed with similar objectives to ours. However, it does not allow retrieving values for a target population [16].

With respect to those tools, 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, and will be, implemented without changing the system architecture). Moreover, it is intended to exploit 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 values by filtering on patients' profiles.

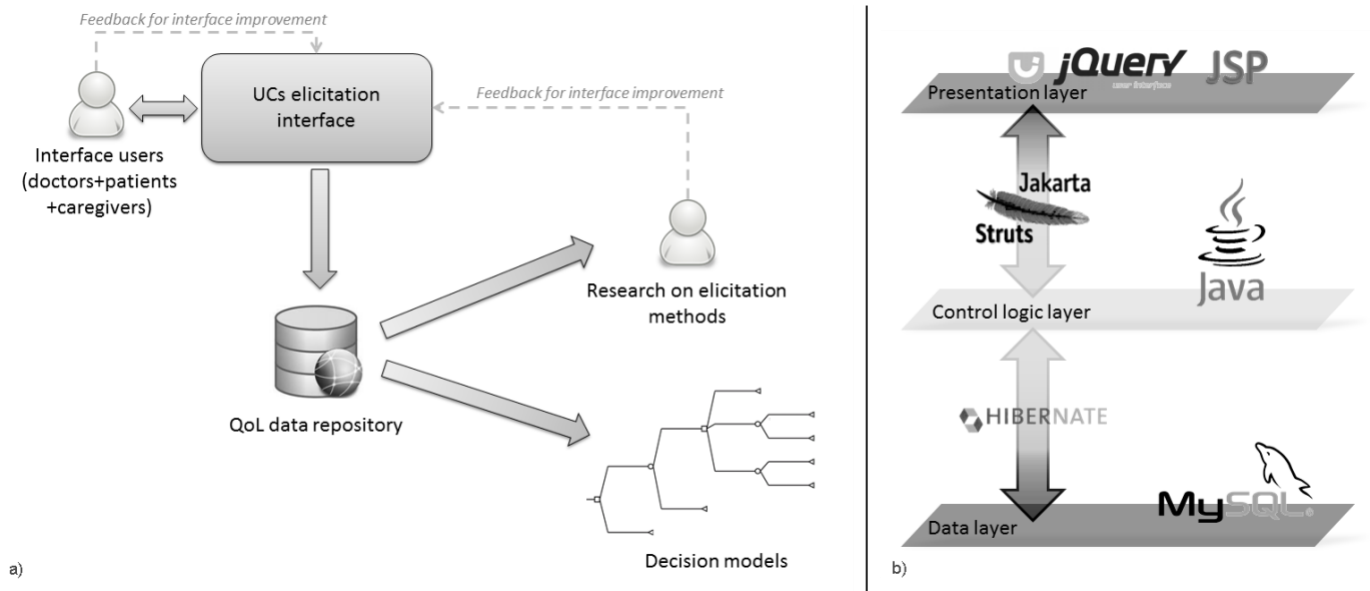
## **Methods**

### ***Architecture***

The functional architecture of UceWeb is represented in Figure 1a. 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. The result of each elicitation is stored in the UceWeb database. As this collaborative repository builds up in size and relevance, it will provide a valuable source of data for the quantification of decision models, as well as for further research about UCE methods themselves. A repository of decision models will be available to the users, with the possibility of personalizing them according to the specific patient's UCs. Currently, two such models have been built for the MobiGuide Project, using TreeAgePro ([www.treeage.com](http://www.treeage.com)) [5].

---

<sup>1</sup> Willingness To Pay is the maximum price a subject would pay to obtain perfect health. This amount is in general constrained with an upper threshold depending on the patient's income.



**Figure 1 - Functional architecture (a) and technical architecture (b) of UceWeb**

Figure 1b shows the technical components, mainly relying on Java technology and Model-View-Controller 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 for a richer interaction.

### **Data model**

In order to illustrate with sufficient detail the information collected through UceWeb, in the following we describe the entities of the implemented data model (see the Entity-Relationship diagram in the supplementary material, Suppl Figure 1). 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 HIPAA requirements for fully de-identified data [17] 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](http://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 (as mentioned, in special cases this could be a patient’s 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.

### **Design features**

Besides the data model, the main focus in the development of UceWeb has been on the design of an effective graphical user interface (GUI). The review of previous research about UCE and preliminary feedback received by our patients and care professionals during the development phase highlighted the main challenges encountered in computerized utility elicitation.

*Difficult questions/concepts.* UCE methods are often difficult to understand for a patient. This is one of the reasons why elicitation is usually assisted by a trained professional. However, some expedients can be adopted in the design of the elicitation tool to facilitate patients. For example, during RS elicitation, if a patient prefers to have a visual, instead of a numerical representation of values, UceWeb is able to dynamically switch to a set of smiles (see results section). 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 [18], particularly when dealing with age-related impairment or stress. Thus, 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 risk of death being presented to the patient. Icons can be alternatively placed all in a row (previous literature suggests that the estimated risk tends to be more accurate with this arrangement [19]) 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 a matter of fact, in a shared-decision making framework, there is often the need to represent possible future health states that can occur as consequences of the different treatments being evaluated (those states are then in general exploited in formal decision models). An effective explanation of the health state is essential for a sound UCE, and UceWeb provides 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.* As already explained, in TTO and SG the patient is iteratively asked a set of questions until the indifference point is reached and the UC value can be calculated. The value used in the initial question (“...would you take a 5% risk of death?...”) plays a major role in the elicitation process, as it may lead responses to cluster around that value (anchoring) [13], 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. We already pointed out that RS does not provide true UCs, but instead it provides values that can be used to rank the health states. Starting the patient's interview from the RS has the additional advantage of making the patient familiar with the tool using the simplest method. The elicited values can then be profitably used 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 [11] who, recognizing the limits of RS, suggests not to deprecate its use, but applying it together with additional, sounder, methods. Finally, questions like “...would you take a 5% risk of death?” can also be presented by the application under the opposite perspective “...would a 95% chance of living be acceptable to you?”. This simple feature can help to avoid framing bias effects [20].

*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. In contrast, 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  $\log_2(N)$  [21]. With  $N=101$ , this leads to approximately 5.66 questions to be asked. If increased precision is needed for UC (e.g., the 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.

*Summarizing heterogeneous utility values in a single meaningful score.* Once  $UC_{TTO}$  and  $UC_{SG}$  have been elicited, UceWeb also provides a functionality to calculate the average value. As shown in the validation section, this value is able to potentially better summarize the QoL experienced by the patient. Moreover, a weighted mean, giving more

importance to SG or to TTO, can also be calculated, according to the doctor's feeling about the patient's understanding of the two methods (motivating that feeling in a mandatory text field). This functionality is in line with findings reporting that some patients perceive the two methods with different difficulty and may misunderstand the questions or get confused with the required tasks [22].

*Data sharing, anonymous data and awareness of patient profile.* As already claimed, UceWeb has been designed to be a collaborative tool. Besides being a computerized tool to assess patient preferences, it also aims at being a unified source of data for the scientific community interested in QoL and in research about the elicitation methods. Since data collected with UceWeb will be publicly available, personally identifying information is never stored within the application. On the other hand, a rich patient profile (refer to the data model section for the full list of information collected) is still provided to include interesting metadata in the UCs repository. Such data enables more effective analysis tasks like clustering patients or searching for common features that influence the elicitation process.

## Results

The UCE process starts with the introduction of a new patient in the system together with his profile. The patient is assigned a unique anonymous identifier generated by the system, and from then on he will be shown in the list of patients available for utility elicitation. A simple menu allows the user to choose the health state (labeled according to the SNOMED standard terminology) and the elicitation method. Once the health state of interest has been identified, appropriate multimedia and textual information material is provided by UceWeb before the elicitation process starts. Elicitation sessions are usually carried out during face to face encounters between a patient and a doctor. However, as already mentioned, self-elicitation could be possible after extensive training.

A typical elicitation session starts with a RS elicitation (Figure 2a) that provides  $UC_{RS}$  values. This is not mandatory, but it can be helpful to proceed with the other methods, as pointed out in the previous section. In the TTO method (Figure 2b) 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. SG elicitation (Figure 2c and 2d) starts with one basic question to check the patient's risk aversion. 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  $UC_R$  is not available. The proposed risk for the gamble is selected with an iterative process driven by a bisection algorithm (see the methods section for a detailed description) until the indifference point  $SG_{final}$  is reached.

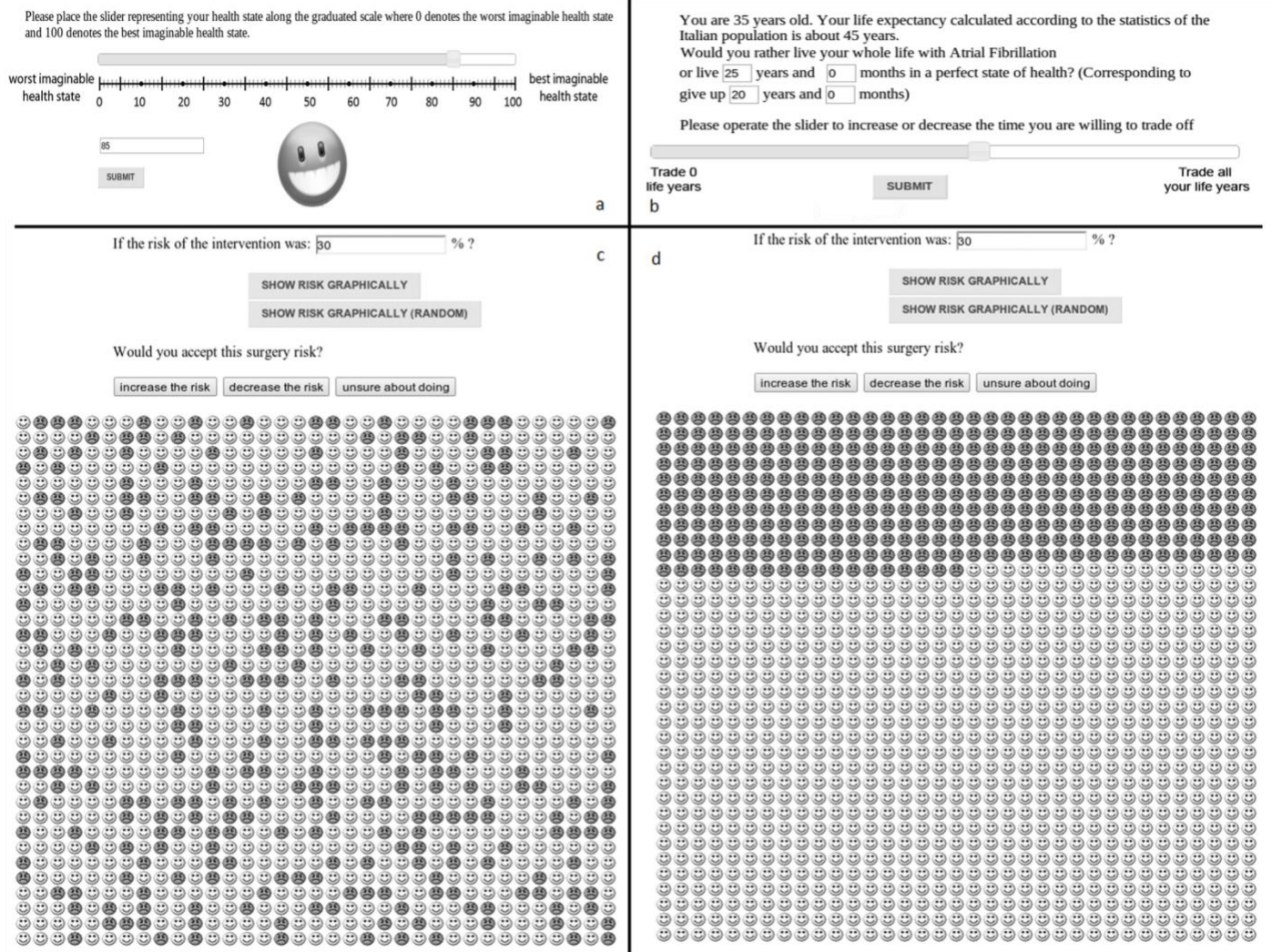


Figure 2 – UceWeb elicitation pages: a) RS: the smile under the bar turns to a more sad/happy look while operating the slider. The icon visualization can also be turned off according to the patient's preferences; b)TTO; c)SG with random risk visualization d)SG with linear risk visualization

Every elicitation ends with a report page reporting date of the elicitation, method, health state and elicited UC.

## Validation

UceWeb was tested on 20 volunteers (10 males and 10 females) with a diagnosis of atrial fibrillation (AF) (4 paroxysmal AF, 10 persistent AF, and 5 permanent AF) treated at the Fondazione Salvatore Maugeri (FSM) Hospital (Pavia, Italy) from April to July 2013. Mean age was 66.2 years (range 34-79); there were 19 Caucasian and one North African subjects. The average number of years of education was 10.35. Three patients regularly use computers, 4 occasionally, 3 often and 10 out of 20 never made use of any computer.

### Validation methods

Inclusion criteria were the diagnosis of AF (independently from other comorbidities) and the willingness to participate in the study. Subjects were either admitted to the division of cardiac rehabilitation (n=18) or seen at the outpatient clinic (n=2) at FSM. The acquisition of medical records was regulated by formal informed consent approved by FSM IRB. Each patient carried out one complete elicitation session including all the three methods while assisted by his/her doctor. We relied on the doctor's observation to gather some feedback about patients' reactions and ability to use the tool. This was possible because he had good knowledge of his patients even on a personal level, which made him a good observer for reactions to the usage of the tool. Patients were regularly asked if they understood the question of each elicitation method and they also gave a feedback about the user interface (whether they liked the icons, if they found them helpful or confusing, if they would have preferred different layouts for the GUI, etc.).

EuroQol-5D and AFEQT, two validated questionnaires for QoL assessment, were also administered to the same patients. We used a combination of a generic and disease-specific questionnaire to cover a broader range of QoL relevant aspects [23]. EuroQol-5D ([www.euroqol.org](http://www.euroqol.org)) 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 is made up 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 ([www.afeqt.org](http://www.afeqt.org)) is an AF-specific questionnaire designed to assess the impact of AF on patients' QoL and possibly monitor the effects of treatment changes. AFEQT questionnaire is made up of 20 questions, divided into 4 sections, namely Symptoms, Daily Activities, Treatment Concerns, and Treatment Satisfaction. The responses to each AFEQT question are scored on a 1 to 7 Likert scale, where 1 stands for "not at all" and 7 stands for "extremely". 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 [24]. 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.

### Validation results

Summary statistics (mean  $\pm$  standard deviation) of the scores for the two questionnaires (AFEQT and EuroQol) and for the three direct utility elicitation methods (SG, TTO and RS) are reported in Table 1. In addition to the overall score, the four AFEQT subscores are reported. Concerning the direct methods, our results are in agreement with the literature, reporting that RS usually provides lower and more variable values [25].

**Table 1- Mean and standard deviation (SD) of the scores obtained with the questionnaires and with the three direct elicitation methods**

	AFEQT					EuroQol	U <sub>TTO</sub>	U <sub>SG</sub>	U <sub>RS</sub>
	Overall score	Symptoms	Daily activities	Treatment concern	Treatment satisfaction				
<b>Mean</b>	67.71	76.09	66.67	63.58	67.13	0.586	0.979	0.977	0.669
<b><math>\pm</math> SD</b>	$\pm$ 19.02	$\pm$ 20.31	$\pm$ 25.17	$\pm$ 22.44	$\pm$ 20.10	$\pm$ 0.369	$\pm$ 0.058	$\pm$ 0.219	$\pm$ 0.196

Since the questionnaires we selected are validated tools to assess QoL, we expected to find correlations among their scores and the UCs elicited with the three implemented direct methods. To investigate this issue, we computed all the correlation coefficients (using the Pearson's method) between the scores obtained with the questionnaires and the scores obtained using TTO, SG and RS. The full correlation matrix summarizing this analysis is provided in the supplementary material (Suppl. Table 1). We herein report a summary of the most relevant results. We obtained significant correlations between the overall AFEQT score and the TTO method and between the AFEQT symptoms subsection score and the SG method. From this observation it emerges that the values elicited with the different methods are in general not well-correlated. This can be due to different reasons. First of all, since, as mentioned, the direct methods may be difficult to understand, some of the values provided by patients could be not reliable. Second, the two questionnaires are quite different, being the EuroQol very general, and the AFEQT specific for AF. An interesting result regards the AFEQT symptom and daily activities subscores. In this case, the most significant correlation coefficient was not found with one of the elicitation methods, but with the average value obtained from TTO and SG. In particular, for the symptoms subscore, we found a correlation coefficient of 0.67 ( $p < 0.002$ ) (see supplementary material, Suppl. Figure 2), 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 the QoL, and must be jointly considered to have a global picture of the patient's perspective. This is an example of how the collected data can be used to gain further insight into theoretical models of patients' preferences.

Another point we have evaluated is the time needed to administer the different methods. This data is relevant because visit duration is always a concern for physicians. As a consequence, it is important to inform them about how longer visits would be with the introduction of a new task. Average minutes necessary to administer the tools were: 11.7 for



AFEQT (range 5-20), 3.2 for EuroQol (range 2-5), 1.5 for RS (range 1-2), 3.2 for TTO (range 2-5) and 5.9 for SG (range 2-15).

Finally, we considered the patients' feedback. We registered a great variability among patients, especially regarding the GUI features. This was one of the main drivers that led to the decision to leave the tool quite configurable: for example, the user can decide to turn on or off the display of the smiles in each of the methods. After setting these preferences the first time, the tool is able to save this configuration for the next elicitation sessions for that patient.

## **Discussion and conclusion**

During development and first evaluation of UceWeb some limitations were highlighted. The first issue regards the hypothesis of using UceWeb for monitoring QoL in longitudinal studies. The data repository of UceWeb is designed to store only de-identified data. This doesn't allow physicians to easily retrieve all the data related to a specific patient by simply querying the system. Currently, the only solution is to take note of the id generated by the system for each patient at the first registration and locally keeping trace (offline) of the corresponding patients' identifiers.

Another critical success factor for long-term monitoring using our tool is patients' willingness to regularly assess QoL on their own. QoL assessment can be a valuable complement to patients' health evaluation [26]. Integration of UceWeb in the MobiGuide telemedicine environment for monitoring and guidance can be an effective way to improve patients' engagement, stressing the overall benefits of providing self-evaluations [27]. We plan to develop a mobile version of the tool for this purpose.

Furthermore, the use of UceWeb assumes that the initial elicitation encounters are carried out by doctors and patients together. This fact is important for ensuring a good level of understanding of the tool and elicitation methods themselves, which is crucial for the reliability and accuracy of the resulting UCs. On the one hand, close collaboration between doctor and patient would promote the adoption of a shared-decision making model of medicine. On the other, though, this might raise some issues. For example, if a wider adoption of the tool was planned, several healthcare providers and patients would need to be trained. This would also represent an economical cost and organizational burden for the structure, deserving further investigation for sustainability.

Sustainability is also related to the time effort needed by physicians when using the tool during control visits. Considering the limited amount of time physicians usually have, even 10 extra minutes could be a problem. The values we have presented in the results section include the time spent by the interviewer to explain the different methods to the patient, thus they overestimate the time effort for the following visits, when the patient is already trained. In addition, utility elicitation is not a routine task and physicians could plan it in advance according to their time constraints.

Another important aspect to consider is the fact that, in this paper, we have mainly stressed the collaborative intention of UceWeb and its features to facilitate the utility elicitation encounters, more than the variety of elicitation methods implemented. We are aware that a number of other methods are available (e.g. Willingness-To-Pay, questionnaires etc.) and we plan to extend the tool to include them.

The presented validation of the tool deserves some discussion, too. As described in the previous sections, an preliminary effort was made to evaluate the results obtained with UceWeb. In the selected patients' sample, however, we obtained less than expected significant correlations between UceWeb scores and QoL questionnaires. This fact could be partially due to the small size of the set of considered patients. To further clarify the nature of these results, additional investigations on larger and more homogeneous sets of patients will be performed.

Finally, UceWeb, although useful as a standalone tool, will benefit from the integration of a module able to run decision models directly exploiting the elicitation results (both for a single specific patient and for population studies). A repository of decision trees represented with the TreeAge tool is under development for this purpose.

## **Acknowledgements**

The authors thank Riccardo Gorrini for his help in the development of UceWeb prototype and Berit Brattheim for the fruitful discussions on the tool. This work has been funded by the MobiGuide project (FP7-287811).

## References

1. Marthe R. Gold, David Stevenson, Dennis G. Fryback. Similarities and Differences in Summary Measures of Population Health. *Annu. Rev. Public Health* (2002) 23:115–34
2. F. Reed Johnson. Editorial: Moving the QALY Forward or Just Stuck in Traffic? *Value In Health* (2009) Volume 12, Supplement 1, S27-S30
3. Paul Kind, Jennifer Elston Lafata, Karl Matuszewski, Dennis Raisch, BSPHarm. The Use of QALYs in Clinical and Patient Decision-Making: Issues and Prospects. *Value In Health* (2009) Volume 12, Supplement 1, S27-S30
4. Gold MR, Stevenson D, Fryback DG. HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annual Review of Public Health* (2002); 23(1):115–34.
5. L.Sacchi, S. Rubrichi, C. Rognoni et al. From Decision to Shared-Decision: Introducing Patients' Preferences in Clinical Decision Analysis - A Case Study in Thromboembolic Risk Prevention. *Proceedings of 14th Conference on Artificial Intelligence in Medicine*, Murcia, (2013)
6. Gerzeli S, Rognoni C, Quaglini S, Cavallo MC, Cremonesi G, Papi A. Cost-effectiveness and cost-utility of beclomethasone/formoterol versus fluticasone propionate/salmeterol in patients with moderate to severe asthma. *Clin Drug Investig.* (2012);32(4):253-65
7. Hilary F. Ryder, Christine McDonough, Anna N. A. Tosteson, Jon D. Lurie, Decision Analysis and Cost-effectiveness Analysis. *Semin Spine Surg.* (2009) December; 21(4): 216–222.
8. Prosser LA, Grosse SD, Wittenberg E. Health utility elicitation: is there still a role for direct methods? *Pharmacoeconomics.* (2012) Feb 1;30(2):83-6
9. Gafni A. The Standard Gamble method: what is being measured and how it is interpreted. *Health Services Research.* (1994);29:207–224
10. Von Neumann J, Morgenstern O. *The Theory of Games and Economic Behavior*. Princeton university press; 1947
11. Torrance, G. W., Feeny, D., & Furlong, W. Visual Analog Scales Do They Have a Role in the Measurement of Preferences for Health States? *Medical Decision Making* (2001);21(4), 329-334
12. Fryback DG, Dasbach EJ, Klein R, Klein BE, Dorn N, Peterson K, Martin PA. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. *Medical Decision Making* (1993) Apr-Jun;13(2):89-102
13. W. Sumner, R. Nease, B. Littenberg. U-tier: A utility assessment tool. *Proceedings of the Fifteenth Annual Symposium on Computer Applications in Medical Care*. Washington, DC (1991). Pp. 710–715
14. Garber, A. M., Sims, T. L., Miller, D. E., Mahlow, P. T., Bravata, D. M., & Goldstein, M. K. Multimedia quality of life assessment: advances with FLAIR. *Proceedings of the American Medical Informatics Association Annual Symposium* (2005)
15. L. A. Lenert, A. Sturley, M. E. Watson. iMPACT3: Internet-Based Development and Administration of Utility Elicitation Protocols. *Medical Decision Making* (2002);22:464–474
16. Bayoumi A., Dale W. ProSPEQT: a new program for computer-assisted utility elicitation. *Poster presented at: Annual Meeting of Society for Medical Decision Making* (2004)
17. Standards for privacy of individually identifiable health information. Office of the Assistant Secretary for Planning and Evaluation, DHHS. Final rule. *Fed Regist.* 2000 Dec 28;65(250):82462–829
18. Jessica S Ancker, Yalini Senathirajah, Rita Kukafka, et al. Design Features of Graphs in Health Risk Communication: A Systematic Review. *J Am Med Inform Assoc* (2006); 13:608-618
19. Ancker JS, Weber EU, Kukafka R. Effect of Arrangement of Stick Figures on Estimates of Proportion in Risk Graphics. *Medical Decision Making* (2011) Jan;31 (1):143–150
20. Blumenschein K, Johannesson M. An experimental test of question framing in health state utility assessment. *Health Policy* (1998);45:187–193
21. Cormen TH, Leiserson CE, Rivest RL, Stein C. *Introduction to algorithms*. MIT press Cambridge; 2001
22. M. Ryan, V. Watson, M. Amaya-Amaya. Methodological issues in the monetary valuation of benefits in healthcare. *Expert Reviews in Pharmacoeconomic Outcomes Research* (2003); 3(6), 717-727
23. Malý M, Vondra V. Generic versus disease-specific instruments in quality-of-life assessment of chronic obstructive pulmonary disease. *Methods Inf Med* (2006); 45(2):211–5
24. J. Spertus, P. Dorian, R. Buben et al. Development and validation of the Atrial Fibrillation Effect on Quality of Life (AFEQT) Questionnaire in patients with atrial fibrillation, *Circ Arrhythm Electrophysiol* (2011); 4:15–25
25. Morimoto T, Fukui T. Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals. *J Epidemiol* (2002) Mar;12(2):160-78

26. Iida N, Koyama W, Kohashi N, Hayashi T. Significance of measuring the quality of life in health evaluation. *Methods Inf Med* (2000) Aug;39(3):213–6
27. Li I, Medynskiy Y, Froehlich J, Larsen J. Personal Informatics in Practice: Improving Quality of Life Through Data. *CHI '12 Extended Abstracts on Human Factors in Computing Systems*. ACM; New York (2012)