Exploring Differences in Interpretation of Words Essential in Medical Expert-Patient Communication

J. Navarro[a], C. Wagner[a], U. Aickelin[b], L. Green[c] and R. Ashford[c]
[a]Lab for Uncertainty in Data and Decision Making (LUCID) and Horizon,
School of Computer Science, University of Nottingham, UK
[b]The University of Nottingham Ningbo
[c]Nottingham University Hospitals, UK
Email: {psxfjn,christian.wagner,uwe.aickelin}@nottingham.ac.uk
Email: lynsey.green@nuh.nhs.uk, rashford@nhs.net

Abstract—In the context of cancer treatment and surgery, quality of life assessment is a crucial part of determining treatment success and viability. In order to assess it, patient-completed questionnaires which employ words to capture aspects of patients well-being are the norm. As the results of these questionnaires are often used to assess patient progress and to determine future treatment options, it is important to establish that the words used are interpreted in the same way by both patients and medical professionals. In this paper, we capture and model patients perceptions and associated uncertainty about the words used to describe the level of their physical function used in the highly common (in Sarcoma Services) Toronto Extremity Salvage Score (TESS) questionnaire. The paper provides detail about the interval-valued data capture as well as the subsequent modelling of the data using fuzzy sets. Based on an initial sample of participants, we use Jaccard similarity on the resulting words models to show that there may be considerable differences in the interpretation of commonly used questionnaire terms, thus presenting a very real risk of miscommunication between patients and medical professionals as well as within the group of medical professionals.

Keywords—*TESS, Survey data, Computing with Words, Interval Agreement Approach, similarity, medicine, questionnaire.*

I. INTRODUCTION

Fuzzy Set Theory (introduced in [1]) has provided a framework for modelling of uncertainties in a vast field of applications through Fuzzy Sets (FSs). In particular, one of the most recent fields of study is a paradigm called Computing with Words (CW) [2], which has established a methodology where words are used in place of numbers for computing and reasoning.

As part of this recent interest, modelling the subjective meaning of words and perceptions of people has been investigated and a number of techniques have been proposed, such as the Interval Approach (IA) [3], the Enhanced Interval Approach (EIA) [4], [5], and the Interval Agreement Approach (IAA) [6]. Such techniques allow the creation of FS models (from data) for words and/or concepts in order to subsequently perform computation and reasoning. The potential applications are diverse such as: recommendation systems (e.g., [7]) and decision support systems (refer to [8] for an overview). In this

study, we use IAA since it avoids making assumptions about the distributions of the models generated.

In medicine, the capture of information from patients' and medical professionals' perceptions of easiness/difficulty to perform daily activities commonly involves uncertainty due to a number of reasons: variability in people's perceptions throughout the day, mood changes, experience, training, disagreement between groups of people, etc. Several studies have shown that the use of representations of medical data based on FSs (e.g., fuzzy Apgar score [9], Cadiag-2 [10]) is valuable when uncertainty is present.

In this paper, we present an initial study focusing on two aims: on the one hand, we analyse if the words commonly used in this specific area of medicine, specifically the linguistic descriptors used in the Toronto Extremity Salvage Score (TESS) have the same meaning to different groups of people involved (e.g., medical professionals and patients) by comparing the FS models generated with the Jaccard similarity measure [11]. On the other hand, we explore the practical application of the IAA [6], [12], [13] to capture interval-valued data and generate FS models associated with such words.

The paper is structured as follows: Section II provides background on TESS, the importance of having standard words in communication between doctors and patients, the IAA method and the Jaccard similarity measure. In Section III, we provide detail about the interval-valued data captured from 3 different groups of people surveyed (surgeons, physiotherapists and patients) as well as a demonstration of their processing through IAA. Models generated from the data, as well as comparisons between different sources are presented in Section IV. Finally, Section V provides a discussion of the results obtained and their implications both in technical and in application terms while Section VI presents the conclusions and challenges/directions for future work.

II. BACKGROUND

This section presents the TESS questionnaire and discusses the importance of words in medicine. Later, we introduce the IAA used to generate FSs from interval-valued responses as well as the similarity measure used to compare the resulting sets.

A. Toronto Extremity Salvage Score (TESS)

The Toronto Extremity Salvage Score (TESS) is a standard patient-completed questionnaire for assessment of function following sarcoma surgery [14]. It was developed to monitor the effects of therapeutic interventions on patients undergoing sarcoma surgery on extremities. TESS is commonly administered at four time points: the first session (which is commonly before surgery) and 12, 18 and 24 months from then on. The TESS consists of a 30-item lower limb and 29-item upper limb questionnaire that allows participants to pick a selection of responses representing their perceptions about the extent of difficulty to perform daily activities based on a Likert scale. In Fig. 1, we show a fragment of the TESS with two items taken from the lower and upper extremity questionnaires respectively.
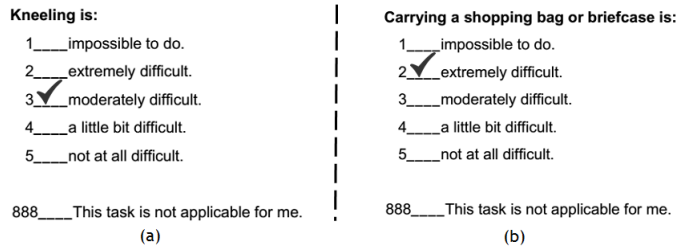

Fig. 1. Two sample TESS items: (a) item taken from the lower extremity questionnaire, (b) item taken from the upper extremity questionnaire.

Note that the questionnaires use the same linguistic descriptors e.g.: *Impossible to do* and *Extremely difficult*. Commonly, after having being completed by the patient, the whole set of answers is used to generate a standardized score ranging from 0 to 100. This evaluated TESS is finally analysed by surgeons/physiotherapists in order to measure changes in physical functions over time as well as evaluating the need for treatment intervention, assistive devices, job modifications, etc.

B. Importance of Words in Medicine

While words in questionnaires are convenient, the challenge of dealing with different interpretations of words ("words mean different things to different people") has previously been identified in Medicine in the context of communication with patients (e.g., patient consent, risk communication) [15]. Therefore, there is a motivation for reducing miscommunication since growing evidence indicates that errors in communication can give rise to an increase of risk in terms of clinical morbidity and mortality [16]. In the context of consent specifically, effective communication is the basis for informed patient consent for medical treatment. More specifically, doctors must empathise with the emotions of the patient and also, explain the possible outcomes and associated risks[17].

Moreover, also in the context of consent, the European Union suggests to use a standardised vocabulary ("very common", "common", "uncommon", "rare", and "very rare"). However, patients' interpretations of these terms do not seem to correlate with the probabilities that they were intended to convey [18] since abundant evidence points out that descriptive terms reflect the speaker's perspective, with the patient often understanding the risks to be of a totally different order of magnitude [15]. In addition, different countries probably bring different shades of meaning to various descriptions [18]. For these reasons, Paling [17] recommends to discuss with colleagues (at a local and national level) the use of a standardised vocabulary of descriptive words so that miscommunication is reduced.

While both the importance and the challenge around miscommunication of/through words has been explored in detail in the context of patient consent, the same is not the case in the context of patient treatment and quality of life assessment (i.e., as in the TESS). Thus, this paper employs recently developed data capture and modelling techniques to explore and numerically assess potential variations in meaning of key words by medical professionals and patients.

C. Interval Agreement Approach

In [6], the Interval Agreement Approach (IAA) is introduced as a method for generating FSs from interval-valued data representing uncertainty in people's opinions/perceptions. It is built on top of the work presented in [12], where an agreement-based method [19] of capturing interval-valued survey data is demonstrated. Also, in [13] its practical application along with the use of a similarity measures to relate attribute word models to concept models is explored.

The IAA considers two types of intervals in the process of capturing responses: crisp (no uncertainty about the interval endpoints) and uncertain (each endpoint modelled itself as a crisp interval). Also, it considers two types of uncertainty to be modelled through different dimensions of the resultant FSs, namely inter-source (variation among a group of participants) and intra-source (variation in the opinion of a particular participant). Depending on the data, the IAA can generate:

- Type-1 FSs (T1 FSs). In this case, crisp intervals and either inter- or intra-source uncertainty are modelled in the primary degree of membership y (or μ) by combining multiple intervals,

- Interval type-2 (IT2 FSs). In this case, uncertain intervals and also, either inter- or intra-source uncertainty is modelled in the primary degree of membership y (or μ) by combining multiple intervals,

- General type-2 FS based on zSlices [20]. In this case, both inter- and intra-source uncertainty are being modelled through the primary y (or u) and secondary z (or μ) degrees of membership.

In this paper, we conduct a single iteration of a survey with multiple participants. Thus, we are focusing on capturing inter-participant uncertainty through crisp intervals and will limit the further description of IAA to the case of T1 FS generation.
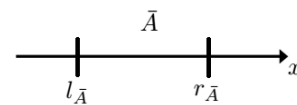

Fig. 2. Crisp interval $\bar{A}$.

Let $\bar{A}$ be a crisp interval with the left and right endpoints $l_{\bar{A}}$ and $r_{\bar{A}}$ (see Fig. 2a). For a given set of sources (e.g.,

a group of patients), a T1 FS is created on the basis of the provided crisp interval(s) (see Fig. 3) representing the agreement between different participants' opinions/responses. The degree of membership y of the set over the survey domain x captures the number of intervals overlapping at a particular point. Figures 2a and 2b show the case of generating a T1 FS for single and multiple intervals respectively using the IAA.
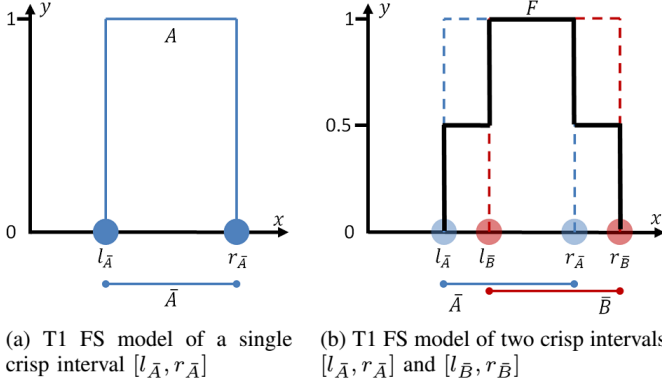

(a) T1 FS model of a single crisp interval $[l_{\bar{A}}, r_{\bar{A}}]$

(b) T1 FS model of two crisp intervals $[l_{\bar{A}}, r_{\bar{A}}]$ and $[l_{\bar{B}}, r_{\bar{B}}]$

Fig. 3. Generation of T1 FSs from crisp interval(s) using the IAA. Figure taken from [6].

Given a series of intervals $\bar{A}_n$ to be combined in a T1 FS A, $n \in \{1, ..., N\}$, where N is the number of intervals, the membership function of A (denoted by $\mu(A)$) is described as shown in (1).

$$\mu_A(x') = \frac{\left(\sum_{i=1}^{N} \mu_{\bar{A}_i}(x')\right)}{N},$$ (1)

where: $\mu_{\bar{A}_i}(x') = \begin{cases} 1 & l_{\bar{A}_i} \leq x' \leq r_{\bar{A}_i} \\ 0 & else \end{cases}$.

D. Similarity Measures

Similarity measures are functions used in (fuzzy) set theory to compare crisp and FSs. This section provides a brief overview of these methods focusing on the measure selected for this article.

A similarity measure $s : S(A, B) \rightarrow [0, 1]$ is a function which assigns a similarity value s to a pair of fuzzy sets (A, B) that indicates the degree to which the FSs A and B are similar [21]. One of the most used methods to measure similarity in both crisp and fuzzy set theory is the Jaccard similarity coefficient [11] which, in the case of type-1 fuzzy sets, can be expressed as:

$$S_J^{FS}(A, B) = \frac{\sum_{i=1}^{N} \min(\mu_A(x_i), \mu_B(x_i))}{\sum_{i=1}^{N} \max(\mu_A(x_i), \mu_B(x_i))},$$ (2)

where N is the total number of discretisations along the x-axis, and $x \in X$ is the domain of the membership functions associated with the FSs A and B. The result indicates how similar A is to B, with 1 indicating that both FSs are identical and 0 that they are disjoint.

III. COMPARING THE PERCEPTION OF LINGUISTICALLY EXPRESSED FUNCTIONAL OUTCOMES

In this section we describe the data collection conducted and provide an example of processing the interval-valued responses to generate FS models for later analysis in Section IV.

A. Overview

We surveyed thirty-six participants (12 sarcoma surgeons, 12 physiotherapists and 12 patients undergoing lower limb salvage surgery) on the 5 levels of each score from the TESS items: "*impossible to do*","*extremely difficult*", "*moderately difficult*", "*a little bit difficult*", and "*not at all difficult*". The sarcoma surgeons completed the questionnaires at the British Orthopaedic Association Annual Scientific Meeting and the physiotherapists at the Sarcoma Physiotherapist Networking Event, whereas the patients were given instructions during their respective medical consultations and completed the questionnaires prior to leaving the hospital following their consultation.

While the questionnaires were provided to all the participants within the same time frame, in the specific case of patients is noteworthy mentioning that their medical consultations were conducted at different points with regard to their surgery, i.e., some of them were completed before surgery intervention and some others after surgery (up to 18 months). Note that, as the questionnaire focused on the interpretation of the linguistic terms, and not on the patients' actual condition, the timing is not considered relevant.

B. Processing the real responses

As an initial part of this study, we developed a questionnaire with a continuous interval-valued scale for each item. We provided instructions to the participants to draw an ellipse around the appropriate extent of each of the 5 linguistic descriptors (words) used in the TESS score. The position and width of the ellipse on the scale indicated the difficulty level and the extent of the uncertainty perceived by the participant.

After collating the questionnaires, we proceeded to extract the interval-valued data and to model the inter-participant agreement from the different groups: patients, surgeons, physiotherapists and a combined model from both surgeons and physiotherapists representing the body of medical professionals. The rationale of doing so, was to analyse the variation in interpretation of TESS items amongst patients and medical professionals. This analysis is presented in Section IV.

C. Data Modelling Example

For a given subject/concept, consider two intervals $\bar{A}$ and $\bar{B}$ generated from the ellipses displayed in Fig. 4 where we are showing two possible responses from two different sources/participants ($N = 2$). Note that the wider the ellipse/interval, the higher the uncertainty in the answer. In other words, this difference in the width indicates how a participant might answer a question s/he is a) relatively certain or b) fairly uncertain of.

From the intervals $\bar{A} = [l_{\bar{A}}, r_{\bar{A}}] = [3, 5]$ and $\bar{B} = [l_{\bar{B}}, r_{\bar{B}}] = [4, 7]$ the FS C is generated using a discrete representation (see (4)). For space considerations, we only
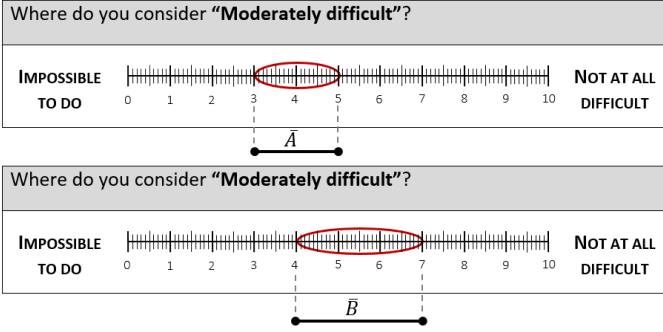
Fig. 4. Two sample ellipses from two sources. Each ellipse results in one interval.

present the calculations for two discrete values $x = 3$ and $x = 4$ (see (3)).

$$\mu_C\,(x = 3) = \frac{(1 + 0)}{2} = 0.5$$
$$\mu_C\,(x = 4) = \frac{(1 + 1)}{2} = 1$$

$$(3)$$

$$C = 0.5/3 + 1/4 + 1/5 + 0.5/6 + 0.5/7 \qquad (4)$$

Finally, the resulting FS C is depicted in Fig. 5. Note that a non-parametric FS model has been generated and the agreement between the two responses has been modelled in the primary degree of membership y (commonly called μ).
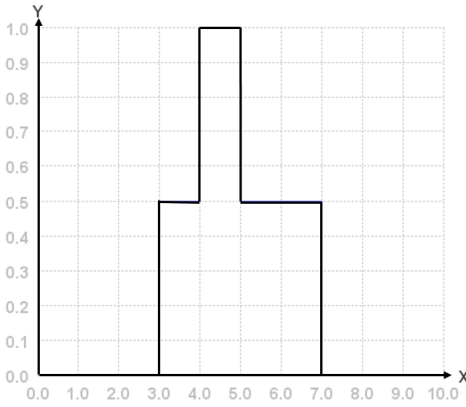

Fig. 5. FS model of two intervals $[l_{\bar{A}}, r_{\bar{A}}]$ and $[l_{\bar{B}}, r_{\bar{B}}]$ from different sources.

IV. RESULTS

In this section, we present the analysis of the different FSs created with the IAA capturing the agreement between four groups involved in the TESS application, namely: Patients, Physiotherapists, Surgeons and a fourth one created from the combined responses from both Physiotherapists and Surgeons (PS) which together represent the body of "medical profession-als" interacting with the patients. The rationale of modelling both groups is analysing the extent of agreement/disagreement within the overall group of medical professionals and to provide an overall comparison with the patients.

A. Fuzzy Sets generated

Figure 6a depicts the FSs for the inter-patient agreement for the 5 linguistic descriptions (words) present in the TESS. As can be seen, the generated FSs have intervals with perfect agreement ($y = 1$) for the descriptors *Impossible to do* and *Not at all difficult*, whereas for the remaining descriptors *Extremely difficult*, *Moderately difficult* and *A little bit difficult*, there are wider and shorter FSs as a result of lower levels of agreement (overlap) and higher levels of uncertainty.

Figure 6b shows the FS word models generated from the Physiotherapists' responses. It is noteworthy that similarly to the patients' case, these FSs have parts with high agreement such as the descriptors *Impossible to do*, *Moderately difficult*, and *Not at all difficult*. For the remaining descriptors *Extremely difficult* and *A little bit difficult*, it can be seen that there is more disagreement about the extent of such words on the scale. Interestingly, the model for *A little bit difficult* is quasi bi-modal, indicating two different interpretations of the term.

Figure 6c shows the FSs representing the word models generated from Surgeons' responses. For this group, the linguistic descriptor *Moderately difficult* presents the highest level of agreement overall whereas the linguistic descriptors for the extreme cases *Impossible to do* and *Not at all difficult* show wider FSs and lower agreement, representing higher uncertainty perceived by the Surgeons group when using those descriptions in contrast to the previous 2 groups.

In order to provide a comparison from the point of view of Patients vs Medical Professionals, we have created a set of models representing the Medical Professionals by using the responses of both Surgeons and Physiotherapists. The model is shown in Figure 6d. The model is more fine-grained as a result of having more intervals. Low agreement and high uncertainty still relates to the descriptor *A little bit difficult* as in the both original cases and *Moderately difficult* is still the descriptor with most agreement overall.

Overall, it is noteworthy that the model for *A little bit difficult* is the widest and lowest (least agreement) throughout, indicating the term is the least clear in the TESS vocabulary.

B. Comparisons between different groups of people

While the modelling of interval valued data with the IAA provides framework for visual interpretation of agreement between different groups of people, it is their analysis with similarity measures that provides a numerical one-to-one comparison. Table I contains all of the comparisons for the different linguistic descriptors based on the Jaccard Similarity measure (see (2)).

Summarizing Table I, we highlight the least similar FSs:

- For the linguistic descriptor *Impossible to do* (see also Fig. 7), the comparison Patients-Surgeons shows the least similar (0.670) models.

- For the linguistic descriptor *Extremely difficult* (see also Fig. 8), the comparison Physiotherapists-Surgeons shows lowest similarity (0.640).

- For the linguistic descriptor *Moderately difficult* (see also Fig. 9), the comparison Physiotherapists-Surgeons shows the lowest similarity (0.614).
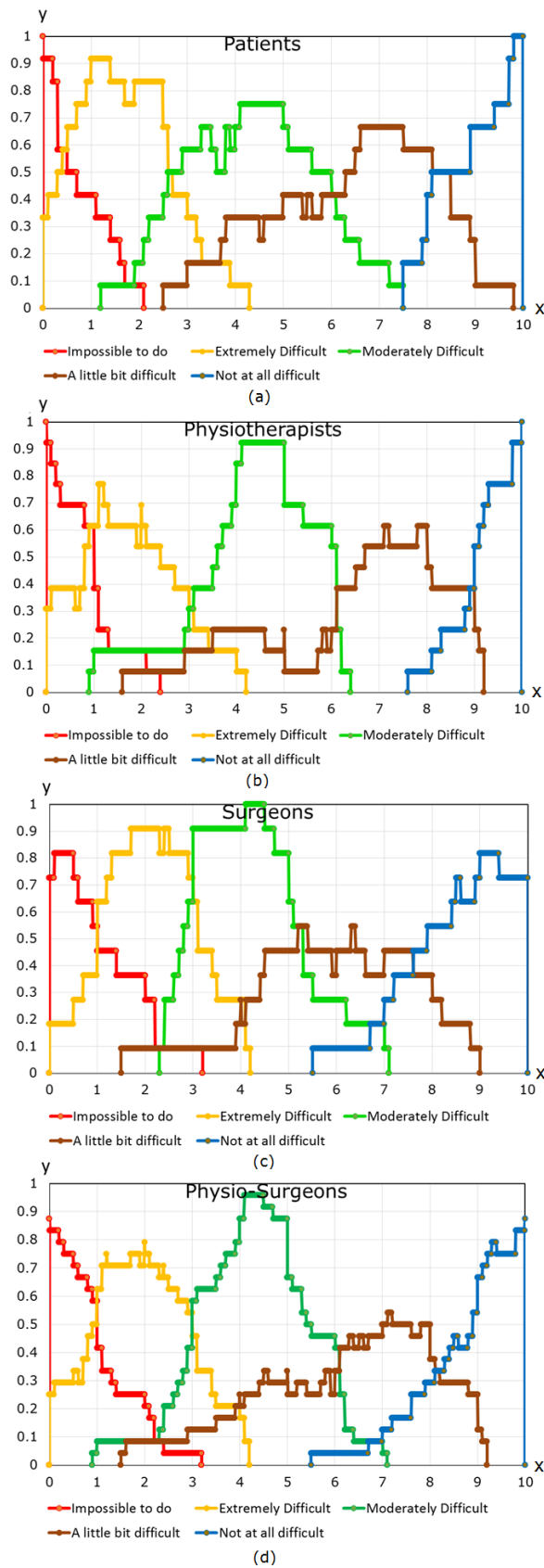
Fig. 6. FSs modelling the word concepts: *Impossible to do*, *Extremely difficult*, *Moderately difficult*, *A little bit difficult*, and *Not at all difficult*, generated from different sources: (a) is for patients. (b) is for Physiotherapists. (c) is for Surgeons. (d) is for the combined responses from both Physiotherapists and Surgeons.

TABLE I. SIMILARITIES BETWEEN DIFFERENT GROUPS FOR ALL WORDS

Impossible to do				
	Patient	Physio	Surgeon	PS
Patient	-	0.752	0.670	0.735
Physio	0.752	-	0.710	0.838
Surgeon	0.670	0.710	-	0.847
PS	0.735	0.838	0.847	-
Extremely difficult				
	Patient	Physio	Surgeon	PS
Patient	-	0.730	0.703	0.745
Physio	0.730	-	0.640	0.798
Surgeon	0.703	0.640	-	0.801
PS	0.745	0.798	0.801	-
Moderately difficult				
	Patient	Physio	Surgeon	PS
Patient	-	0.678	0.696	0.765
Physio	0.678	-	0.614	0.785
Surgeon	0.696	0.614	-	0.786
PS	0.765	0.785	0.786	-
A little bit difficult				
	Patient	Physio	Surgeon	PS
Patient	-	0.683	0.707	0.726
Physio	0.683	-	0.622	0.770
Surgeon	0.707	0.622	-	0.805
PS	0.726	0.770	0.805	-
Not at all difficult				
	Patient	Physio	Surgeon	PS
Patient	-	0.704	0.694	0.782
Physio	0.704	-	0.512	0.675
Surgeon	0.694	0.512	-	0.760
PS	0.782	0.675	0.760	-

- For the linguistic descriptor *A little bit difficult* (see also Fig. 10), comparisons between Physiotherapists-Surgeons and Physiotherapists-Patients show low similarity (0.622 and 0.683 respectively).

- For the linguistic descriptor *Not at all difficult* (see also Fig. 11), the comparison Physiotherapists-Surgeons shows the lowest similarity (0.512) amongst all the word models.
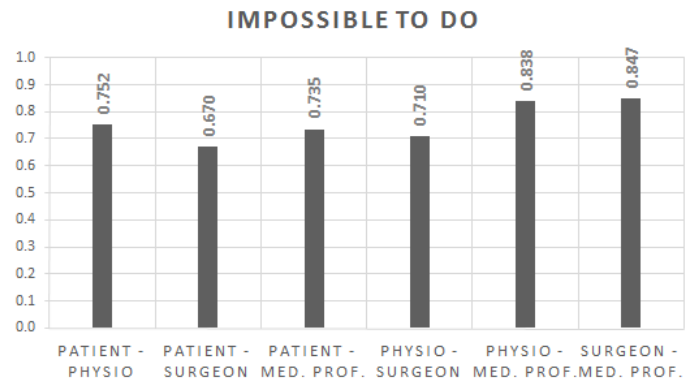

Fig. 7. FS similarities between different groups of people for the linguistic descriptor *Impossible to do*.

Table II and Figure 12 show the average of the similarities over the 5 words, thus providing an indication of how similar the interpretation of the TESS item vocabulary is by the key stakeholder groups: Patients, Physiotherapists and Surgeons.
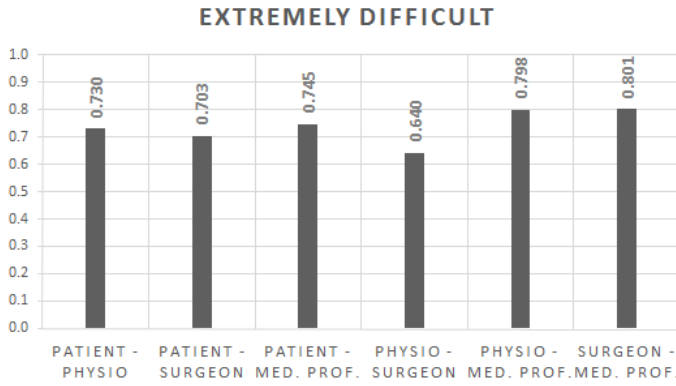
Fig. 8. FS similarities between different groups of people for the linguistic descriptor *Extremely difficult*.
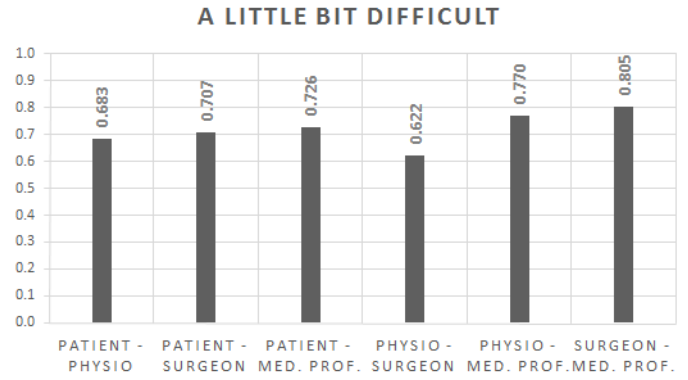

Fig. 10. FS similarities between different groups of people for the linguistic descriptor *A little bit difficult*.
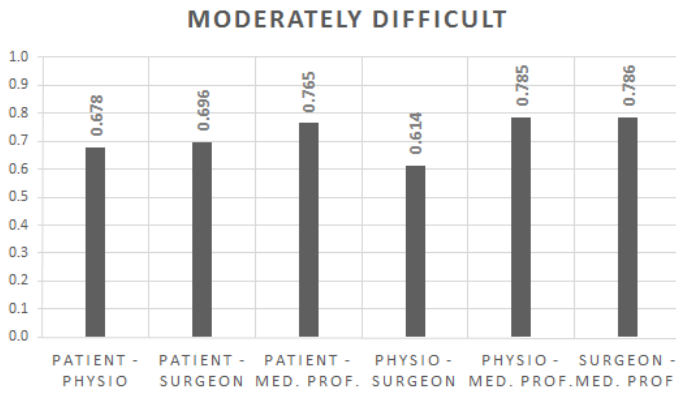

Fig. 9. FS similarities between different groups of people for the linguistic descriptor *Moderately difficult*.
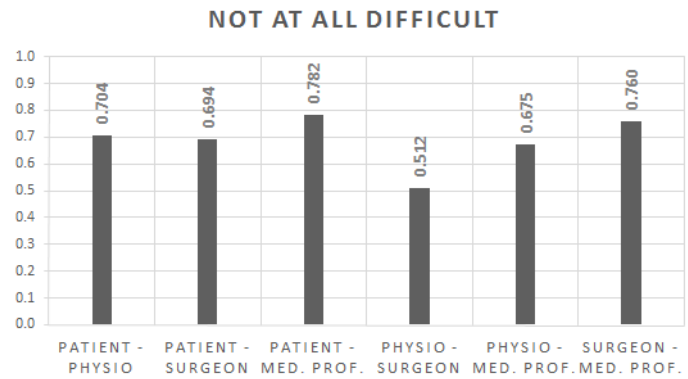

Fig. 11. FS similarities between different groups of people for the linguistic descriptor *Not at all difficult*.

While the sample used for this paper is too small to draw any conclusions, it is interesting to note that the understanding is best (most similar) between patients and physiotherapists. While the similarity in interpretation is lowest between physiotherapists and surgeons. Note that as here we are interested in comparing real-world stakeholder groups, the similarity to the generated group of medical professionals is omitted.

TABLE II. AVERAGE OF FS SIMILARITIES ACROSS ALL WORDS

	Patient	Physio	Surgeon	Physio-Surgeon
Patient	-	0.709	0.694	0.751
Physio	0.709	-	0.619	0.773
Surgeon	0.694	0.619	-	0.800

Table III shows the centroids of the FSs depicted in Fig. 6, while Table IV and Table V show their heights and the size of their support. The centroids, heights and support provide a basic numeric description of the FSs as a means of additional comparison allowing us to deduce the following:

- For the first linguistic descriptors *Impossible to do* and *Extremely difficult* it can be noted that Patients' and Physiotherapists' centroids are the closest among the groups whereas the Surgeons' centroid is higher considerably.

- For the linguistic descriptor *Moderately Difficult*, all

of the centroids are ≈ 4.3 which can be interpreted as a generalized agreement in the perception of the descriptor *Moderately Difficult*. Interestingly, it can be noted that the FSs are slightly "balanced" to the left side of the scale. This can indicate that the descriptor *Moderately Difficult* has not necessarily a neutral meaning as expected.

- For the linguistic descriptor *Not at all difficult*, the centroid associated to the Surgeons' model is consider-
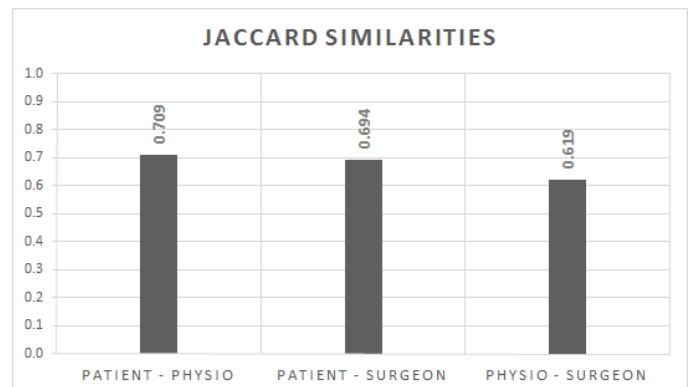

Fig. 12. Averages of FS similarities between different groups of people.

ably lower than the rest. This can be interpreted (along with the width of the FS) to a different perception of the linguistic descriptor for the Surgeons in which it can be used in more varied situations.

- The support and heights of the FS models related to the *Impossible to do* and *Not at all difficult* descriptors are the narrowest and highest amongst Patients and Physiotherapist, indicating such words are the less ambiguous for those groups.

- The size of support of the descriptor *Extremely Difficult* was almost identical for all groups of participants indicating that, despite the lack of a high agreement level in general (indicated by the height), the perception of such word can cover the same range of scenarios with all groups.

- For the case of the linguistic descriptor *A little bit difficult*, it can be seen that in general, all models have a low height (in comparison to other word models) and conversely, the supports are the widest overall covering the major part of the scale. These observations can suggest that the linguistic descriptor *A little bit difficult* does not describe in a satisfactory manner the intended level of function among our population of participants.

TABLE III.   Centroids of FS word models from different groups of people

	Patient	Physio	Surgeon	PS
Impossible to do	0.686	0.727	0.988	0.878
Extremely Difficult	1.711	1.767	2.085	1.954
Moderately Difficult	4.356	4.312	4.289	4.290
A little bit difficult	6.433	6.462	6.126	6.188
Not at all difficult	9.051	9.279	8.555	8.814
Overall (Average):	4.448	4.509	4.408	4.425

TABLE IV.   Heights of FS word models from different groups of people

	Patient	Physio	Surgeon	PS
Impossible to do	1.000	1.000	0.917	0.875
Extremely Difficult	0.917	0.769	0.917	0.792
Moderately Difficult	0.750	0.923	1.000	0.958
A little bit difficult	0.667	0.615	0.583	0.542
Not at all difficult	1.000	1.000	0.833	0.875

TABLE V.   Size of support of FSs related to word models from different groups of people

	Patient	Physio	Surgeon	PS
Impossible to do	2.1	2.4	3.2	3.2
Extremely Difficult	4.3	4.2	4.2	4.2
Moderately Difficult	6.3	5.5	4.8	6.2
A little bit difficult	7.3	7.6	8	7.7
Not at all difficult	2.5	2.4	4.5	4.5

V. Discussion

The IAA is designed to avoid assumptions about the data or FS shape [6], instead, it models the agreement/disagreement (in y) and the uncertainty associated (in x) from multiple intervals capturing a linguistic term. In this paper, we have used it to create FSs of the standardised vocabulary used in the TESS by three key stakeholder groups.

We began this study with the question if there are differences in the perception of the linguistic descriptors in the TESS that may influence the medical assessment and therapy outcomes. This analysis, despite being at a preliminary stage with limited data, suggests that each of the terms is not necessarily equidistant as assumed in the items of a Likert scale. Also, the similarity values indicate that for some stakeholders, the interpretations of linguistic descriptors vary considerably.

As expected, all FSs present the same order i.e., when comparing the "order of appearance" in each group of participants the first FS is *Impossible to do*, the second FS is *Extremely difficult* and so on. Therefore, the items (linguistic descriptors) do follow an ascending order in a ranking sense. However, by looking at the centroids of the FSs we can detect that they are not equidistant since their differences would be expected to be (at least) approximately equal. For example, the minimum and maximum differences between linguistic models created from Physiotherapists' responses are 1.041 (*Impossible to do* and *Extremely difficult*) and 2.818 (*A little bit difficult* and *Not at all difficult*). In addition, comparisons between the sizes of support help to show the differences in the spread of each word's interpretation model on the scale, where the size of the descriptor *A little bit difficult* is considerably wider than the rest of the words (across all groups).

In general, the linguistic descriptors found at the boundaries (i.e., *Impossible to do* and *Not at all difficult*) and in the middle of the scale (*Moderately difficult*) are more defined and have higher agreement whereas the remaining descriptors have more uncertain distributions. Here, *A little bit difficult* is the model reflecting the lowest level of agreement in all cases and the highest uncertainty, this, if confirmed with a more substantial dataset, can indicate that these terms lead to potential miscommunication between stakeholders of the TESS.

VI. Conclusions and Future Work

In this study, we have modelled the perceptions of the linguistic descriptors used as the standard vocabulary in the TESS questionnaire using the Interval Agreement Approach. We used an interval-valued scale to capture the inter-participant uncertainty about the extent of the given descriptors with 4 groups, namely: Patients, Surgeons, Physiotherapist and Surgeons-Physio, i.e., Health Professionals.

We performed comparisons between the Fuzzy Sets, showing that some words are clearer and more unanimously understood than others. In particular, we found very little similarity between Physiotherapists and Surgeons for the descriptors *Extremely difficult*, *Moderately difficult*, *A little bit difficult* and *Not at all difficult*. We also analysed the centroids of the models as a means of numeric description indicating that the items are not equidistant (as assumed for the Likert scale). In particular, for the specific case of the term *A little bit difficult* it is interesting how it stands out as the least well defined, with generally low agreement and at times, a quasi bi-modal model and interpretation.

While this study is at a preliminary stage with a small sample not large enough to support statistically relevant conclusions, the paper has shown that interval-valued data capture and subsequent modelling of the intervals using the IAA provides

a promising tool for analysing standardised vocabulary -for example, in medical-patient communication. Such analysis can be used to identify potential variations in meaning/perceptions of key words in medical treatment/diagnosis and thus support improved expert-patient communication which in turn, may lead to improve medical treatment.

In the future, we aim to collect a larger sample of the proposed data to validate the initial finding in this paper. More fundamentally, we are seeking to develop the approach used in this paper to experimentally assess the equidistant spacing of terms on ordinal scales while exploring the additional information provided by interval-valued scales. In relation to the latter, we are in particular looking to transform questionnaires such as the TESS to an interval-valued scale and to analyse the outcome and its potential for real-world application. Finally, we are currently developing further tools for the (statistical) comparison of data-driven fuzzy sets such as generated by the IAA.

ACKNOWLEDGEMENT

REFERENCES

[1] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] ——, "Fuzzy logic equals computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 2, pp. 103–111, 1996.

[3] L. Feilong and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *Fuzzy Systems, IEEE Transactions on*, vol. 16, no. 6, pp. 1503–1521, 2008.

[4] S. Coupland, J. M. Mendel, and D. Wu, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and convergence of the word fous," in *International Conference on Fuzzy Systems*. IEEE, 2010, pp. 1–8.

[5] D. Wu, J. M. Mendel, and S. Coupland, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis," *IEEE Transactions on Fuzzy Systems*, vol. 20, pp. 499–513, 2012.

[6] C. Wagner, S. Miller, J. Garibaldi, D. Anderson, and T. Havens, "From interval-valued data to general type-2 fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 23, pp. 248–269, 2014.

[7] J. Mcculloch, S. S. Member, C. Wagner, S. S. Member, K. Bachour, and T. Rodden, "give me what i want-enabling complex queries on rich multi-attribute data," in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, 2015, pp. 1–9.

[8] F. Herrera, S. Alonso, F. Chiclana, and E. Herrera-Viedma, "Computing with words in decision making: Foundations, trends and prospects," *Fuzzy Optimization and Decision Making*, vol. 8, pp. 337–364, 2009.

[9] H. Shono, M. Oga, K. Shimomura, M. Yamasaki, Y. Ito, M. Muro, and H. Sugimori, "Application of fuzzy logic to the apgar scoring system." *International journal of bio-medical computing*, vol. 30, no. 2, pp. 113–123, 1992.

[10] A. Ciabattoni, T. Vetterlein, and K.-P. Adlassnig, "A formal logical framework for cadiag-2." *Studies in health technology and informatics*, vol. 150, pp. 648–52, 2009.

[11] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bulletin de la Société vaudoise des sciences naturelles*, vol. Volume 5, no. 163, 1908.

[12] S. Miller, C. Wagner, J. M. Garibaldi, and S. Appleby, "Constructing general type-2 fuzzy sets from interval-valued data," *IEEE International Conference on Fuzzy Systems*, pp. 10–15, 2012.

[13] C. Wagner, "Juzzy - a java based toolkit for type-2 fuzzy logic," *Proceedings of the IEEE Symposium Series on Computational Intelligence, Singapore*, pp. 45–52, 2013.

[14] A. M. Davis, J. G. Wright, J. I. Williams, C. Bombardier, A. Griffin, and R. S. Bell, "Development of a measure of physical function for patients with bone and soft tissue sarcoma," *Quality of Life Research*, vol. 5, no. 5, pp. 508–516, 1996.

[15] J. Merz, M. Druzdzel, and D. Mazur, "Verbal expressions of probability in informed consent litigation," *Medical Decision Making*, vol. 11, no. 4, pp. 273–81, 1991.

[16] E. Coiera, "When conversation is better than computation." *Journal of the American Medical Informatics Association : JAMIA*, vol. 7, no. 3, pp. 277–86, 2000.

[17] J. Paling, "Strategies to help patients understand risks," *BMJ: British Medical Journal*, vol. 327, no. 7417, pp. 745–748, 2003.

[18] D. C. Berry, D. K. Raynor, P. Knapp, and E. Bersellini, "Patients' understanding of risk associated with medication use: impact of european commission guidelines and other risk scales." *Drug safety : an international journal of medical toxicology and drug experience*, vol. 26, no. 1, pp. 1–11, 2003.

[19] C. Wagner and H. Hagras, "Employing zslices based general type-2 fuzzy sets to model multi level agreement," *IEEE SSCI 2011: Symposium Series on Computational Intelligence - T2FUZZ 2011: 2011 IEEE Symposium on Advances in Type-2 Fuzzy Logic Systems*, pp. 50–57, 2011.

[20] ——, "Toward general type-2 fuzzy logic systems based on zslices," *Fuzzy Systems, IEEE Transactions on*, vol. 18, no. 4, pp. 637–660, 2010.

[21] M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification." *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 28, no. 3, pp. 376–86, 1998.