

A consensus model for Delphi processes with linguistic terms and its application to chronic pain in neonates definition

Núria Agell¹, Christ Jan van Ganzewinkel², Mónica Sánchez³, Llorenç Roselló³,
Francesc Prats³, Peter Andriessen²

¹ *Department of Information Systems Management and GREC Group, Esade - Universitat Ramon Llull, Barcelona, Spain (email: nuria.agell@esade.edu).*

² *Department of Neonatology, Máxima Medical Center, The Netherlands (email: {p.andriessen;C.vanGanzewinkel}@mmc.nl).*

³ *Department of Applied Mathematics 2 and GREC Group, Technical University of Catalonia, UPC-BarcelonaTech, Barcelona, Spain (email: {monica.sanchez;llorenç.rosello;francesc.prats}@upc.edu).*

A consensus model for Delphi processes with linguistic terms and its application to chronic pain in neonates definition

Núria Agell¹, Christ Jan van Ganzewinkel², Mónica Sánchez³, Llorenç Roselló³,
Francesc Prats³, Peter Andriessen²

¹ *Department of Information Systems Management and GREC Group, Esade - Universitat Ramon Llull, Barcelona, Spain (email: nuria.agell@esade.edu).*

² *Department of Neonatology, Máxima Medical Center, The Netherlands (email: {p.andriessen;C.vanGanzewinkel}@mmc.nl).*

³ *Department of Applied Mathematics 2 and GREC Group, Technical University of Catalonia, UPC-BarcelonaTech, Barcelona, Spain (email: {monica.sanchez;llorenç.rosello;francesc.prats}@upc.edu).*

Abstract

This paper proposes a new model of consensus based on linguistic terms to be implemented in Delphi processes. The model of consensus involves qualitative reasoning techniques and is based on the concept of entropy. The proposed model has the ability to reach consensus automatically without the need for either a moderator or a final interaction among panelists. In addition, it permits panelists to answer with different levels of precision depending on their knowledge on each question. The model defined has been used to establish the relevant features for the definition of a type of chronic disease. A real-case application conducted in the Department of Neonatology of Máxima Medical Center in The Netherlands is presented. This application considers the opinions of stakeholders of neonate health-care in order to reach a final consensual definition of chronic pain in neonates.

Keywords: Qualitative Reasoning, Group Decision systems, Consensus measures, Knowledge Management, Knowledge acquisition, Delphi technique,

1. Introduction

Delphi technique is a well-known group decision-making method involving a structured interaction among a panel of experts or stakeholders, which anonymously tries to reach consensus on the significant features of a certain topic [25]. Since its introduction in the 1960s, it has been used to attain convergence of expert opinion in a variety of fields of knowledge such as program planning, needs assessment, policy determination and resource utilization [11, 19, 22, 35].

The Delphi method has proved to have some functional advantages over other consensus-building methods, such as brainstorming, dialectical inquiry and nominal group [14, 22, 24, 25, 35]. The moderator in these group decision-making techniques conducts the group communication and consensus processes through several rounds. However handling uncertainty and linguistic terms in group assessments is one of the main problems of this type of methods.

To handle the uncertainty and linguistic information inherent to human consensus processes, many group decision-making techniques have been developed and are available in the academic literature [2, 7, 13, 14, 16, 17]. In [18] a review of consensus models in a fuzzy environment can be found. There is, nowadays, a wide range of areas of application for these methods, from managerial to medical or engineering [6, 8, 10, 30]. In particular, some fuzzy Delphi approaches have been proposed to deal with uncertainty and linguistic information [9, 12, 27]. Although, through these approaches, participants use a set of ordered linguistic labels, they are unable to use different levels of precision in their assessments. In

addition, these fuzzy Delphi approaches share with original Delphi technique the absence of a definition of a degree of consensus. These have been considered as significant drawbacks by Delphi technique users.

The new approach to the Delphi method developed in this paper, not only includes the use of linguistic information, with different levels of precision, but also computes a degree of consensus in each round of the Delphi process. It permits each participant to utilize linguistic terms that reflect more adequately the level of uncertainty intrinsic to his evaluation, and to be dynamically aware of their agreement in each round.

To this end, the new Delphi approach is based on qualitative reasoning techniques [34]. Participants' assessments through linguistic terms are considered qualitative labels in an absolute order-of-magnitude qualitative space [31]. Different levels of precision are used to reflect the distinctions required by evaluators' reasoning processes. Techniques based on order-of-magnitude qualitative reasoning have provided theoretical models that permit operating in conditions of insufficient or non-numerical data [34]. One of the advantages of qualitative reasoning is its ability to tackle problems in such a way that the principle of relevance is preserved, i.e., each variable involved in a real problem is valued with the level of precision required.

The paper comprises six sections. Section 2 introduces the main features of Delphi processes. The theoretical framework for this new approach is then presented in Section 3. In Section 4 the new approach for Delphi processes, based on a group consensus measure with linguistic terms is explained. A real case example in the health-care sector is presented in Section 5 to show the performance of this new approach. Finally, the main conclusions and lines of future research are discussed in Section 6.

2. Delphi Processes: Overview and Key Points

Dalkey and Helmer and the Rand Corporation developed the Delphi technique in 1963 [11]. This technique is usually used for determining the set of possible alternatives, finding implicit assumptions conducting to different judgements, exploring new solutions for a specific problem, or reaching consensus about a specific topic from a panel of experts or stakeholders.

A Delphi process is generally designed through 3 to 4 rounds of questions. In the first round, in order to gather panelists' opinions, open-ended questions are used. The results of this first round are classified into statements which are then valued by the panelists in a second round. In the consecutive rounds the

panelists are showed the values of the total panel and are asked to re-assess their own values in the light of the group's opinion. Frequently, this type of iteration leads to a consensus on the group of significant statements.

The main weaknesses or limitations on the Delphi method are the absence of a definition of a degree of consensus, the difficulty of dealing with the uncertainty involved in panelists' opinions, and the way in which some opinions are suppressed during the consensus process.

Several fuzzy Delphi approaches have been developed in the literature to solve these issues. The application of fuzzy theory to the Delphi method by means of linguistic variables was initially introduced in [27]. A fuzzy Delphi method considering pessimistic, moderate and optimistic assessments of experts via triangular fuzzy numbers was introduced in [23]. Using triangular fuzzy numbers to model the experts' judgments, in [21] consensus is reached in only one round thanks to the implementation of the max-min fuzzy Delphi method and a new Delphi method via fuzzy integration. After reviewing the previous fuzzy Delphi works, a new approach using fuzzy statistics is proposed in [9]. An application of fuzzy Delphi Method to obtain the critical factors of the regenerative technologies by using fuzzy AHP to find the importance degree of each factor is introduced in [20]. A web based consensus support system for group decision making problems and incomplete preferences was introduced in [2]. The method is similar to Delphi technique but it does not rely on the use of questionnaires and the moderator tasks can be replaced. An extension of the recent literature and an implementation of fuzzy Delphi for the adjustment of statistical forecast can be found in [12]. This study presents a fuzzy Delphi adjustment process to improve accuracy and introduced an empirical study to illustrate its performance.

A new approach for Delphi processes is proposed in this paper. It is based on a definition of a degree of consensus that can be used when experts' answers (as from round 2) are given with linguistic terms. Linguistic terms are handled by means of order-of-magnitude qualitative reasoning techniques. [31, 32] offer a detailed application of these methods to group decision-making and consensual processes.

3. Order-of-Magnitude Reasoning Framework

In this section, we briefly review the basic concepts of the qualitative absolute order-of-magnitude model which will be used in the next sections [1, 31, 34]. This paper relies on the use of linguistic terms based on this model. This allows the imprecision involved in panelists' opinions in Delphi processes to be managed.

The qualitative absolute order-of-magnitude model of granularity n considers a finite set of *basic* qualitative labels, $\mathbb{S}_n^* = \{B_1, \dots, B_n\}$, which is totally ordered: $B_1 < \dots < B_n$ [1, 31].

In general, each basic qualitative label corresponds to a linguistic term, for instance for $n = 5$: $B_1 = \text{“Strongly disagree”} < B_2 = \text{“Disagree”} < B_3 = \text{“Neither agree nor disagree”} < B_4 = \text{“Agree”} < B_5 = \text{“Strongly agree”}$.

The *complete universe of description* for the absolute order-of-magnitude space with granularity n , is the set \mathbb{S}_n :

$$\mathbb{S}_n = \mathbb{S}_n^* \cup \{[B_i, B_j] \mid B_i, B_j \in \mathbb{S}_n^*, i < j\},$$

where the *non-basic* label $[B_i, B_j]$ with $i < j$ is defined as the set $\{B_i, B_{i+1}, \dots, B_j\}$, whereas $[B_i, B_i] = B_i$ [1, 31].

Following with the above-mentioned set of $n = 5$ linguistic terms, the non-basic label $[B_1, B_2]$ represents the linguistic term [“Strongly disagree”, “Disagree”]. The linguistic term “Unknown” is represented by [“Strongly disagree”, “Strongly agree”], i.e., $[B_1, B_5]$. This least precise qualitative label is denoted by the symbol $?$, i.e., in \mathbb{S}_n , $[B_1, B_n] \equiv ?$.

This structure permits working with all different levels of precision from the basic labels B_1, \dots, B_n to the $?$ label (see Figure 1).

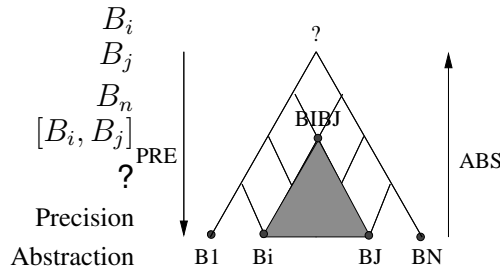


Figure 1: The complete universe of description \mathbb{S}_n [31]

In addition, we also review the concept of extended measure in \mathbb{S}_n and the connex union and intersection operations introduced in [31]:

A normalized measure μ is considered in the set of basic qualitative labels, $\mu : \mathbb{S}_n^* \rightarrow [0, 1]$ such that $\sum_{B_i \in \mathbb{S}_n^*} \mu(B_i) = 1$. This measure is directly extended to \mathbb{S}_n by defining $\mu([B_i, B_j]) = \sum_{k=i}^j \mu(B_k)$.

In order to define the degree of consensus among a set of panelists’ opinions, the connex union and the intersection between qualitative labels are also

considered [31]. Given two qualitative labels $[B_{i_1}, B_{j_1}], [B_{i_2}, B_{j_2}] \in \mathbb{S}_n$, their connex union is the label $[B_{i_1}, B_{j_1}] \sqcup [B_{i_2}, B_{j_2}] = [B_{\min(i_1, i_2)}, B_{\max(j_1, j_2)}]$. When $[B_{i_1}, B_{j_1}] \cap [B_{i_2}, B_{j_2}] \neq \emptyset$, their intersection is the qualitative label $[B_{i_1}, B_{j_1}] \cap [B_{i_2}, B_{j_2}] = [B_{\max(i_1, i_2)}, B_{\min(j_1, j_2)}]$.

Finally, an *iterative relaxation process* is initiated in order to reach a non-empty intersection among the set of qualitative labels when this intersection is initially empty (see a detailed explanation in [31]). The iterative relaxation process is done by means of a dive function ϕ which makes an immersion in a space with a greater granularity (with more levels of precision).

Considering \mathbb{S}_n with basic labels $\mathbb{S}_n^* = \{B_1, \dots, B_n\}$ as the initial space with granularity n , and a space \mathbb{S}_{n+1} with granularity $n + 1$, with basic labels $\mathbb{S}_{n+1}^* = \{B'_1, \dots, B'_{n+1}\}$, the *dive function* ([31]) is the map:

$$\phi : \mathbb{S}_n \rightarrow \mathbb{S}_{n+1},$$

such that, $\phi(B_i) = [B'_i, B'_{i+1}]$ for any basic label $B_i \in \mathbb{S}_n$, and, $\phi([B_i, B_j]) = \bigcup_{k=i}^j \phi_0(B_k) = [B'_i, B'_{j+1}]$ for any non-basic label.

This relaxation process is performed iteratively as many times as necessary in order to get a non-empty intersection. When this process is iterated over time, in each iteration a new measure μ' is computed. For instance, in the first iteration the new measure μ' can be obtained applying the next formulas for the basic labels:

$$\begin{aligned} \mu'(B'_1) &= \frac{1}{2}\mu(B_1), \\ \mu'(B'_i) &= \frac{1}{2}(\mu(B_{i-1}) + \mu(B_i)), \quad i = 2, \dots, n \\ \mu'(B'_{n+1}) &= \frac{1}{2}\mu(B_n), \end{aligned}$$

and the following for non-basic labels:

$$\mu'([B'_i, B'_{j+1}]) = \sum_{k=i}^j \mu'(B'_k), \quad \forall i, j = 1, \dots, n.$$

In the next section, computations in Examples 1 and 3 illustrate this iterative relaxation process and the computation of the new measure μ' .

4. A New Approach for Delphi Processes

The new approach for Delphi processes involves the measure of consensus among panelists with respect to each statement in several rounds. For this reason,

a degree of consensus is defined to order the statements in view of the opinions given by a panel of participants through a Delphi survey. Panelists opinions are expressed using a set of linguistic terms with various levels of precision as introduced in Section 3.

4.1. Measuring Consensus among Panelists

Let us consider a panel of m stakeholders and a set of statements Λ to be assessed from the second round on of the Delphi process. The new approach of Delphi processes proposed in this paper involves the notion of entropy of a qualitative label, defined in \mathbb{S}_n , as a measure of the information provided by the qualitative labels in \mathbb{S}_n , inspired from the Shannon entropy concept in information theory.

Definition 1. *The entropy of a qualitative label $Q \in \mathbb{S}_n$ is defined as:*

$$H(Q) = -\log_2(\mu(Q)),$$

where μ is the considered normalized measure in the set of basic qualitative labels.

Note that Definition 1 is obtained by considering a restriction of the entropy definition introduced in [31]. In this paper, we consider the entropy of each qualitative label in \mathbb{S}_n , instead of considering the entropy of a qualitative description of a set over \mathbb{S}_n . The reason is that, in the proposed approach for Delphi processes, the concept of entropy of each qualitative label in \mathbb{S}_n is needed to define the degree of consensus among panelists with respect to each statement.

Once the notion of entropy of a qualitative label has been defined, the definition of the degree of consensus of the set of panelists with respect to a statement $a \in \Lambda$ is introduced as a quotient of entropies as follows:

Definition 2. *Given m qualitative labels $Q_1, \dots, Q_m \in \mathbb{S}_n$, associated to the assessments of m panelists for a given statement a , such that $\cap_{j=1}^m Q_j \neq \emptyset$, the degree of consensus with respect to a is:*

$$\kappa(Q_1, \dots, Q_m) = \frac{H(\sqcup_{j=1}^m Q_j)}{H(\cap_{j=1}^m Q_j)} = \frac{\log_2(\mu(\sqcup_{j=1}^m Q_j))}{\log_2(\mu(\cap_{j=1}^m Q_j))}$$

If the intersection of all the Q_j is empty then the *iterative relaxation process* mentioned in Section 3 is performed in order to reach a non-empty intersection.

Note that in the case that panelists only use basic qualitative labels, the condition $\bigcap_{j=1}^m Q_j \neq \emptyset$ is only fulfilled when all the panelists' opinions are the same and then the degree of consensus is 1; otherwise, a lower degree of consensus is obtained.

Next example illustrates how the the diving function and the updating measure together with the proposed degree of consensus are computed.

Example 1. *Let us consider the statement $a =$ “Almost continuous pain longer than few hours” and a set of three panelists $\mathbb{E} = \{e_1, e_2, e_3\}$ consisting of a nurse e_1 , a doctor e_2 and a mother e_3 . Let us assume that the assessments of the three panelists with respect to a are represented by three qualitative labels defined as: $Q_1(a) = [B_1, B_2]$, $Q_2(a) = B_3$, $Q_3(a) = B_2$ using the meaning of basic labels B_1, \dots, B_5 given at the beginning of Section 3. Finally, let us define $\mu(B_i) = 1/5, i = 1, \dots, 5$.*

Note that there is not consensus among the panelists' assessments because $Q_1(a) \cap Q_2(a) \cap Q_3(a) = [B_1, B_2] \cap B_3 \cap B_2 = \emptyset$. For this reason, the dive function must be applied once, obtaining a non-empty intersection: $\phi(Q_1(a)) \cap \phi(Q_2(a)) \cap \phi(Q_3(a)) = [B'_1, B'_3] \cap [B'_3, B'_4] \cap [B'_2, B'_3] = B'_3 \neq \emptyset$ and a new measure μ' in S_6^ given by:*

$$\begin{aligned}\mu'(B'_1) &= \frac{1}{2}\mu(B_1) = \frac{1}{10}; \\ \mu'(B'_i) &= \frac{1}{2}(\mu(B_{i-1}) + \mu(B_i)) = \frac{1}{5}, i = 2, \dots, 5; \\ \mu'(B'_6) &= \frac{1}{2}\mu(B_n) = \frac{1}{10}.\end{aligned}$$

Then, since $\sqcup_{k=1}^3 \phi(Q_i(a)) = [B'_1, B'_3] \sqcup [B'_3, B'_4] \sqcup [B'_2, B'_3] = [B'_1, B'_4]$ and $\bigcap_{k=1}^3 \phi(Q_i(a)) = [B'_1, B'_3] \cap [B'_3, B'_4] \cap [B'_2, B'_3] = B'_3$ the degree of consensus is:

$$\begin{aligned}\kappa(Q_1, Q_2, Q_3) &= \frac{H(\sqcup_{k=1}^3 \phi(Q_i(a)))}{H(\bigcap_{k=1}^3 \phi(Q_i(a)))} = \frac{H([B'_1, B'_4])}{H(B'_3)} = \\ &= \frac{\log_2 7/10}{\log_2 1/5} = 0.22\end{aligned}$$

This value of $\kappa(Q_1, Q_2, Q_3)$ suggests a low level of consensus among panelists, consistent with intuition. Next examples present results when different statements and initial panelists' assessments are considered.

Example 2. Analogously to Example 1, let us consider now the statement $b =$ “The pain often cannot be associated with a specific etiology but might well from a combination of things” and assume that the assessments of the three panelists with respect to b are represented by: $Q_1(b) = [B_1, B_2]$, $Q_2(b) = [B_1, B_2]$, $Q_3(b) = B_2$ using the same meaning of basic labels B_1, \dots, B_5 given at the beginning of Section 3. and the same measure μ as in the above example.

In this case there is consensus among the panelists’ assessments because $Q_1(b) \cap Q_2(b) \cap Q_3(b) = B_2$. Then, since $\sqcup_{k=1}^3 (Q_i(b)) = [B_1, B_2]$ and $\cap_{k=1}^3 (Q_i(b)) = B_2$ the degree of consensus is:

$$\begin{aligned} \kappa(Q_1, Q_2, Q_3) &= \frac{H(\sqcup_{k=1}^3 (Q_i(b)))}{H(\cap_{k=1}^3 (Q_i(b)))} = \frac{H([B_1, B_2])}{H(B_2)} = \\ &= \frac{\log_2 2/5}{\log_2 1/5} = 0.57. \end{aligned}$$

This value of $\kappa(Q_1, Q_2, Q_3)$ suggests a greater level of consensus among panelists, consistent with intuition. Finally, let us consider the extreme case in which two panelists’ opinions are “strongly disagree” and “strongly agree”.

Example 3. Let us consider now the statement $c =$ “Daily continuous or intermittent episodes of painful sensations in the newborn” and assume that the assessments of the three panelists with respect to c are represented by: $Q_1(c) = B_1$, $Q_2(c) = B_1$, $Q_3(c) = B_5$ using the same meaning of basic labels and the same measure as above. Obviously, there is not consensus among the panelists’ assessments and the dive function must be applied four times in order to obtain a non-empty intersection: $(\phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1)(Q_1(c)) \cap (\phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1)(Q_2(c)) \cap (\phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1)(Q_3(c)) = [B_1'''' , B_5''''] \cap [B_1'''' , B_5''''] \cap [B_5'''' , B_9''''] = B_5''''$. Then, since $\sqcup_{k=1}^3 \phi^{(4)}(Q_i(c)) = [B_1'''' , B_9''''] = ?$ and $\cap_{k=1}^3 \phi^{(4)}(Q_i(c)) = B_5''''$, computing the their values through the updating measure μ'''' the degree of consensus is:

$$\begin{aligned} \kappa(Q_1, Q_2, Q_3) &= \frac{H([B_1'''' , B_9''''])}{H(B_5'''')} = \\ &= \frac{\log_2 1}{\log_2 16/80} = 0. \end{aligned}$$

In this extreme case in which two panelists' opinions are "strongly disagree" and "strongly agree", the degree of consensus is 0. When the connex union of the initial panelists opinions is the qualitative label Q , the degree of consensus will be 0. For this reason, in real cases applications, outliers will be removed for statements in which panelists extremely disagree.

4.2. The Proposed Approach for Delphi Processes

The new approach for Delphi processes, enables the handling of imprecise information given by evaluators. The proposed approach is based on the degree of consensus introduced in the previous subsection. The degree of consensus allows the ranking and selection of statements. As a result, it has the capacity to obtain consensus automatically without the need for an interaction between participants.

Analyzing the results obtained in Examples 1, 2 and 3 in the previous subsection, it can be seen that qualitative labels with different levels of precision are simultaneously handled to compute the degree of consensus presented. A comparison of the results obtained using the degree of consensus proposed in this paper in Examples 1, 2 and 3 together with the classic statistical parameters that would be used in classic Delphi is presented in Table 1.

Table 1: Results Examples 1 and 2

	Panelists Assessments			Degree of Consensus	Mean	Standard Deviation
	Q_1	Q_2	Q_3	κ	μ	SD
Ex1 (a)	$[B_1, B_2]$	B_3	B_2	0.22	2.17	0.62
Ex2 (b)	$[B_1, B_2]$	$[B_1, B_2]$	B_2	0.57	1.67	0.24
Ex3 (c)	B_1	B_1	B_5	0	2.33	1.89

In addition, the comparison of the outputs produced by the proposed degree of consensus with the classic statistical parameters shows that the new measurement is more consistent with human intuition on consensus. Note that even in the cases where the panelists' opinions have initially no intersection, we are able to compute a measure of consensus. The proposed methodology takes into account the necessary effort that would be needed to reach consensus.

Let us consider the new approach for a Delphi process consisting of three rounds. A scheme for the proposed approach is shown in Figure 2, where the

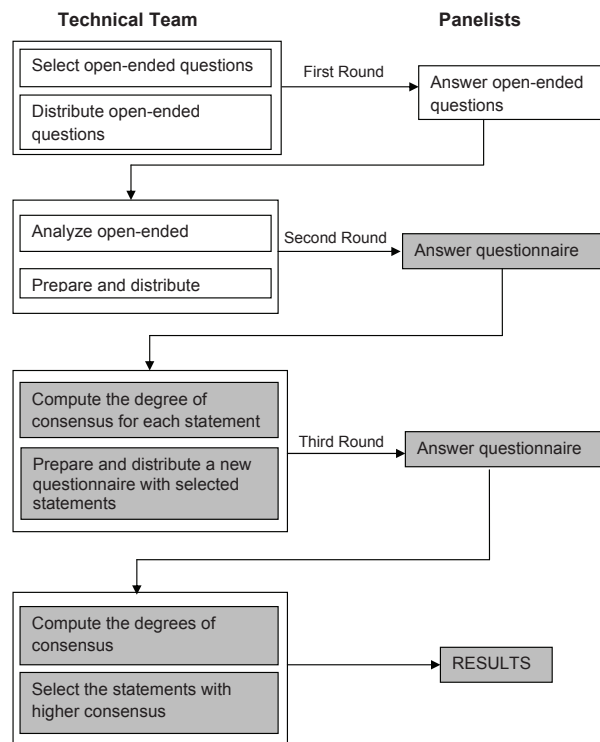


Figure 2: The new scheme of the Delphi process

differences between the proposed approach and the classic Delphi are shadowed.

First, note that the qualitative absolute order-of-magnitude model presented in Section 3 is used in the interaction with the panelists in rounds 2 and 3. When panelists answer the corresponding questionnaires, they can assess statements using linguistic terms with different levels of precision. In addition, the measure of consensus presented in Subsection 4.1 is applied in the way the degree of consensus is computed after rounds 2 and 3, and in the final selection of statements. The proposed consensus scheme allows us to detect statements for which most participants are in consensus, and then, to rank them accordingly to their importance.

5. A Real Case Application to Chronic Pain in Neonates Definition

Nowadays there is a lack of an adequate definition of chronic pain in the newborn period. Research has shown that pain in the newborn period has consequences later in life, such as altered behavior, increased pain sensitivity and

decreased function of the immune system [4, 15, 28]. To date neonatal pain research has mainly focused on acute and procedural pain. Only little is known about other types of pain that are common in the adult, for instance prolonged or chronic pain. Although a small number of studies suggest that chronic pain does exist in the newborn [3, 5, 26], there is no consensus on the definition, etiology is unknown and there are no specific diagnostic determinants. An adequate definition of chronic pain in the newborn period, stating the significant features consensuated by stakeholders of neonate health-care, would enable to design the appropriate treatment.

This section focusses on the definition of chronic pain in neonates. An analysis of the opinions given by doctors, nurses and other stakeholders is conducted by implementing the proposed new approach for Delphi processes in order to select the specific diagnostic determinants for the definition of chronic pain in the newborn period.

A web-based, three round Delphi survey was performed to provide a definition, the etiology and the specific diagnostic determinants of chronic pain in the newborn. The survey, considering the three mentioned aspects, was carried out by the Department of Neonatology of the Máxima Medical Center, Eindhoven Area in The Netherlands. An international panel of experts in the field of neonatology and neonatal pain was invited to participate: health-care providers (doctors and nurses) and parents. The introduced methodology was applied to find a consensus among panelists with respect to the definition of chronic pain in the newborn considering the panelists' responses. A definition of chronic pain in neonates was obtained by using the proposed model. This real case example shows the potential and benefits of the presented methodology.

5.1. Data description

In the first round, participants ($n = 189$) answered the open-ended question *define chronic pain in own words*. The answers were classified and summarized into 114 statements, which were valued by the participants ($n = 189$) on a 5-point Likert scale. In the second round the remaining participants ($n = 72$) were asked to reflect on a selection of 25 statements with a mode or median ≥ 4 or mean ≥ 3.75 . These threshold values were used to provide the opportunity to easily reach consensus in the following round. In the third and last round the remaining participants ($n = 33$) were provided with the values of the total panel responses and their individual response and were asked to re-assess their own values in the light of the group's opinion. Table 2 shows participants' regions of origin and participants' profiles respectively.

Table 2: Distribution of panelists' regions and profiles

Region	<i>n</i>	%	Profile	<i>n</i>	%
Europe	75	39.7	Physician	64	33.9
North America	69	36.5	Clinical Nurse Specialist, Nurse, Practitioner,		
Australia, New Zealand	21	11.1	Physician Assistant	44	23.2
Middle East	8	4.2	Nursing staff	40	21.2
South/Central America	6	3.2	Parent	22	11.6
Asia	6	3.2	Researcher	9	4.8
Africa	4	2.1	Others	10	5.3
Total	189	100.0	Total	189	100.0

5.2. Experimental results

A comparison between the results of the classic Delphi methodology and the new approach presented in this paper has been conducted taking into account the assessments given by participants in the third round about the selected 25 statements. It should be noted that using the classic Delphi methodology those statements with mode, mean and median ≥ 4 simultaneously were selected, resulting in 12 statements. On the other hand, the approach presented in this paper was applied to select the most consensual statements among the obtained 25 statements. The iterative relaxation process explained in Section 4 was applied resulting in 7 statements, in which the participants reached a degree of consensus over 0.20 (see Table 3).

Note that, in Table 3, numbers in bold correspond to those statements selected either by the classic method or the new approach. In addition, shaded rows indicate the 5 statements selected by both methods.

The coincidences and divergences between results of both methods over all the 25 statements are shown in Table 4. The new approach proposed in this paper agreed with classic Delphi in 68% of the statements: 20% of the statements were selected by both methods, whereas 48% were not. Two statements were selected using the new method whereas they were not selected by classic Delphi. These two statements suggest to incorporate a time variable in the definition of chronic pain. However, using classic Delphi method, no time variable was selected. On the other hand, seven statements among those selected by classic Delphi were not selected by the new approach. These seven statements express more than one concept each and, according to health-care providers, this could be quite confusing for the panelists. This confusion is captured by the proposed approach. In the

Table 3: Obtained results

Statements	Classic Delphi				New Approach
	Mode	Median	Mean	<i>SD</i>	Degree of Consensus
Chronic pain in the newborn may lead to a state where any interaction/procedure that is happening to the infant is perceived as painful.	4	4	4.22	0.55	0.37
Chronic pain may likely prolong hospitalization worsening or adding to the existing morbidities.	4	4	4.17	0.65	0.37
Both recurrent and long lasting pain may become chronic.	4	4	4.13	0.78	0.09
Pain that occurs over a period of time, which is ongoing and has no obvious end point in site.	4	4	4.13	0.98	0.09
Chronic pain depletes stress hormones and increases energy consumption therefore interfering with growth.	4	4	4.1	0.92	0.09
Treatment should be based on signs of relief or comfort.	4	4	4.1	0.61	0.37
It can be anywhere in the body depending on the reason for the pain.	4	4	4.07	0.74	0.09
This pain often cannot be associated with a specific etiology but might well be from a combination of things.	4	4	4.07	0.37	0.98
Pain that is ongoing, no longer proximate to a procedure or event.	4	4	4.03	0.69	0.09
A painful event may also alter perception causing events that normally would be tolerate to be perceived as painful, leading to a chronic (longer duration) pain response.	4	4	4.03	0.72	0.37
Poorly controlled acute pain may lead to hyperalgesia, altered pain perception, and possibly a predilection to chronic pain states.	4	4	4.01	0.85	0.09
Daily continuous or intermittent episodes of painful sensations in the newborn.	4	4	4	0.69	0.09
Pain lasting more than 1 week that does not fall under the category of acute pain.	4	3	3.79	0.90	0.24
Pain lasting hours or days.	4	3	3.99	0.94	0.24

group of statements that were selected by both methods, in general, those with high mean values and post hoc calculated small standard deviation show a high level of agreement (degree of consensus) using the new method.

In this example, stakeholders were forced to value statements using a 5-point Likert scale predefined values when they might have wanted to rate them less precisely. This is why, even if we could have dealt with that imprecision with the proposed methodology, we applied it assuming that all the estimations were given by basic labels.

6. Conclusions

The method proposed in this paper, based on a measure of consensus, offers a technique to synthesize a group of stakeholders' opinions through a Delphi sur-

Table 4: Comparison table

	Selected by first method	Not selected by first method
Selected by second method	20%	8%
Not selected by second method	24%	48%

vey. Participants use a set of linguistic labels associated to an order-of-magnitude model to express their evaluations. With this method the group is able to reach consensus automatically without needing neither a moderator nor any interaction between the participants. Moreover, this approach does not need prior normalization to handle imprecise information given by the experts.

There are three main advantages to this approach. First, the different degrees of strictness of the experts' opinions are taken into account. Second, there is no need to compute an average value of ordinal data. And third, this method accommodates "unknown values" by using the label "?" defined in the absolute order-of-magnitude qualitative model.

The proposed method has been used to reach a consensus on the definition of chronic pain in neonates. In addition, a comparison of the results obtained with a statistical study has been performed. The result is a 60% congruence between traditional statistics and this new approach.

Three main lines of future research are being considered. First, from a theoretical perspective, the introduction of machine learning techniques will be explored. This will allow us to update information and landmarks for the selection of statements in each round. Second, a web-based software device for Delphi processes, based on the concepts introduced in this paper is being developed. It will be capable to collect and synthesize opinions expressed with different levels of precision simultaneously. Third, in regard to the real case study presented in this paper, the nature of the 40% difference between both methods will be analyzed and cut-off points will be validated.

To conclude, let us remark that the theory introduced in this paper has a wide domain of potential application in knowledge management, including consumer ratings in marketing research and evaluation or accreditation processes in human resources studies.

Acknowledgement This work is partially supported by the SENSORIAL Research Project (TIN2010-20966-C02-01,02), funded by the Spanish Ministry of Science and Information Technology. The authors would like to thank the Department of Neonatology in Máxima Medical Center, The Netherlands, for providing the research data used in this study.

- [1] Agell, N. et al., 2012. Ranking multi-attribute alternatives on the basis of linguistic labels in group decisions. *Information Sciences*, 209, pp.49-60.
- [2] Alonso, S., Herrera-Viedma, E., Chiclana, F., Herrera, F. 2010. A web based consensus support system for group decision making problems and incomplete preferences. *Information Sciences*, 180 (23), 4477-4495.
- [3] American Academy of Pediatrics, Committee on Fetus and Newborn and Section on Surgery, Section on Anesthesiology and Pain Medicine, and Canadian Paediatric Society and Fetus and Newborn Committee. Prevention and Management of Pain in the Neonate: An Update. *Pediatrics*, 118 (5), 2231-2241. 2006.
- [4] Anand, K. J., Scalzo, F.M. Can Adverse Neonatal Experiences Alter Brain Development and Subsequent Behavior? *BIOLOGY OF THE NEONATE*, 77 (2), 69-82. 2000.
- [5] Anand, K. J., Stevens, B. J., McGrath, P. *Pain in Neonates and Infants*. Pain Research and Clinical Management Series. Elsevier. 2007.
- [6] Belacel, N. 2000. Multicriteria assignment method PROAFTN: Methodology and medical application. *European Journal of Operational Research*, 125, 175-183.
- [7] Ben-Arieh, D., Chen, Z. 2006. Linguistic labels aggregation and consensus measure for autocratic decision-making using group recommendations, *IEEE Trans. Syst., Man Cybernet., Part A: Syst. Hum.* 36 (1) 558-568.
- [8] Canha, L., Ekel, P., Queiroz, J., Schuffner Neto, F. 2007. Models and methods of decision making in fuzzy environment and their applications to power engineering problems. *Numerical Linear Algebra with Applications*, 14, 369-390.
- [9] Chang, P. T., Huang, L. C., Lin, H. J., 2000. The fuzzy Delphi method via fuzzy statistics and membership function fitting and an application to the human resources. *Fuzzy Sets and Systems*, 112, 511-520.

- [10] Choudhury, A.K., Shankar, R., Tiwari, M.K. 2006. Consensus-based intelligent group decision-making model for the selection of advanced technology. *Decision Support Systems*, 42 (3), 1776-1799.
- [11] Dalkey, N. C. and Helmer, O. 1963. An experimental application of the Delphi method to the use of experts. *Management Science*, 9 (3), 458-467.
- [12] Duru, O., Bulut, E., Yoshida, S., 2012. A fuzzy extended DELPHI method for adjustment of statistical time series prediction: An empirical study on dry bulk freight market case. *Expert Systems with Applications* 39, 840-848.
- [13] Eklund, P., Rusinowska, A., De Swart, H. 2007. Consensus Reaching in committees. *European Journal of Operational Research*, 178, 185-193.
- [14] Fu, C., Yang, S-L. 2011. An attribute weight based feedback model for multiple attributive group decision analysis problems with group consensus requirements in evidential reasoning context. *European Journal of Operational Research*, 212, 179-189
- [15] Ganzewinkel, C.J. van, Andriessen, P. Chronic pain in the neonate: A research design connecting ancient delphi to the modern 'dutch polder' *Journal of Research in Nursing*. Published ahead of print, february 8th. 2011.
- [16] Herrera-Viedma, E., Herrera, F., Chiclana, F. 2002. A consensus model for multiperson decision making with different preference structures. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 32, 394-402.
- [17] Herrera-Viedma, E., Martinez, L., Mata, F., Chiclana, F. 2005. A consensus support system model for group decision-making problems with multigranular linguistic preference relations. *IEEE Transactions on Fuzzy Systems* 13(5), 644-658.
- [18] Herrera-Viedma, E., Cabrerizo, F.J., Kacprzyk, J., Pedrycz, W. 2014. A review of soft consensus models in a fuzzy environment. *Information Fusion*, (17), 4-13
- [19] Holsapple, C.W., Joshi, K.D., 2002. Knowledge manipulation activities: results of a Delphi study. *Information & Management* 39, 477-490.

- [20] Hsu, Y-L, Lee, C-H, Kreng, V.B., 2010. The application of Fuzzy Delphi Method and Fuzzy AHP in lubricant regenerative technology selection. *Expert Systems with Applications* 37, 419-425.
- [21] Ishikawa, A., Amagasa, M., Shiga, T., Tomizawa, G., Tatsuta, R., & Mieno, H., 1993. The max-min Delphi method and fuzzy Delphi method via fuzzy integration. *Fuzzy Sets and Systems*, 55, 241-253.
- [22] Juan, Y.K., Castro, D., Roper, K., 2010. Decision support approach based on multiple objectives and resources for assessing the relocation plan of dangerous hillside aggregations *European Journal of Operational Research* 202 (1), 265-272,
- [23] Kaufmann, A., Gupta, M.M., 1988. *Fuzzy Mathematical Models in Engineering and Management Science*, North-Holland, Amsterdam.
- [24] Keeney S., Hasson, F., McKenna, H.P., 2001. A critical review of the delphi technique as a research methodology for nursing *Int J Nurs Stud.* 38, 195-200.
- [25] Linstone, H.A., Turoff, M. 1979. *The Delphi Method: Techniques and Applications*, Addison-Wesley, London, 1979.
- [26] McClain, B. C. and Kain, Z. N. 2005. Procedural Pain in Neonates: The New Millennium. *Pediatrics*, 115 (4), 1073-1075.
- [27] Murray, T. J., Pipino, L. L. and Gigch, J. P., 1985, A Pilot Study of Fuzzy Set Modification of Delphi, *Human Systems Management*, 5, 76-80.
- [28] Page, GG. Are there long-term consequences of pain in newborn or very young infants? *The Journal of Perinatal Education.* 2004;13(3):10-17.
- [29] Polit, D.F., Beck, C.T. 2007. *Nursing research, principles and method* (7th ed). Lippincott Williams & Wilkins, Philadelphia.
- [30] Porcel, C., Herrera-Viedma, E. 2010. Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries. *Knowledge Based Systems*, 23, 32-39.
- [31] Roselló, L., Prats, F., Agell, N., Sánchez, M. 2010. Measuring consensus in group decisions by means of qualitative reasoning. *International Journal of Approximate Reasoning* 51, 441-452,

- [32] Roselló, L., Prats, F., Agell, N., Sánchez, M. 2011. A qualitative reasoning approach to measure consensus. En: E. Herrera-Viedma *et al.* (Eds.), *Consensual Processes*, STUDFUZZ, vol. 267, Springer, Heidelberg, 235-261.
- [33] Roselló, L., Sánchez, M., Agell, N., Prats, F., Mazaira, F.A. 2014. Using consensus and distances between generalized multi-attribute linguistic assessments for group decision-making. *Information Fusion*, 17, 83-92.
- [34] Travé-Massuyès, L., Dague, P. (Eds.), 2003 *Modèles et Raisonnements Qualitatifs*. Hermes Science, Paris.
- [35] Wu, C-R, Lin, C-T, Tsai, P-H. 2010. Evaluating business performance of wealth management banks *European Journal of Operational Research*, 207 (2), 971-979.