KNEE

Management of knee injuries: consensus-based indications from a large community of orthopaedic surgeons

Federico Cabitza · Vincenza Ragone ·
Paolo Arrigoni · Jón Karlsson · Pietro Randelli

Abstract

Purpose To describe preferences and to quantify the amount of agreement among orthopaedic surgeons regarding treatment options for four clinical scenarios of knee pathologies.

Methods A web-based survey was developed to investigate the attitudes of members of an international association of surgeons specialised in sports traumatology and knee surgery European Society of Sports Traumatology, Knee Surgery and Arthroscopy.

Results The response rate was 40 % (412 questionnaires). An inter-rater agreement score (the Normalised Chi-square based Agreement Nomogram, NX2A) was calculated for each question. For scenario 1, 56-year-old male, degenerative medial compartment on both the femoral and tibial side, the surgical approach was preferred to the conservative approach ($p < 0.001$). Biological procedures were not considered appropriate, and in this respect, the respondents achieved a moderate degree of agreement (NX2A = 0.68). For scenario 2, 35-year-old male, early knee medial arthritis, the surgical treatment was preferred to conservative treatment ($p < 0.001$). The traditional closed-wedge tibial osteotomy was not regarded as an appropriate treatment with an excellent degree of agreement among surgeons (NX2A = 0.76). For scenario 3, 46-year-old male, ACL lesion, there was an almost disagreement, as respondents showed no preference between a surgical and conservative approach (NX2A = 0.005). Among surgical treatments, an almost perfect agreement with regard to the appropriateness of arthroscopic single-bundle ACL reconstruction with a semitendinosus/gracilis graft was reached by the surgeons (NX2A = 0.8). For scenario 4, 69-year-old male, diffuse knee arthritis (all the compartments), an almost perfect agreement in favour of a total knee prosthesis was obtained for the management of this scenario (NX2A = 0.85).

Conclusions Web-based survey can help orthopaedic surgeons discuss and propose indications for clinical practice in the management of some of the most common joint diseases.

Level of evidence Cross-sectional survey, Level V.

Keywords Online questionnaires · Web-based surveys · Knee pathology · Treatments · Inter-rater agreement

F. Cabitza
Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano-Bicocca, Viale Sarca 336,
20126 Milan, Italy

V. Ragone (✉) · P. Arrigoni · P. Randelli
Dipartimento di Scienze Medico-Chirurgiche, Università Degli
Studi Di Milano, IRCCS Policlinico San Donato, Piazza
Edmondo Malan, 2, 20097 San Donato Milanese, Milan, Italy
e-mail: ragone.vincenza@inwind.it

J. Karlsson
Department of Orthopaedics, Sahlgrenska University Hospital/
Molndal, 431 80 Molndal, Sweden

Introduction

Surveys are tools using which data on the beliefs, attitudes and behaviour of both patients and doctors can be collected [16] and using which researchers can gather and analyse quantitative data that are essential in clinical epidemiology and health services research [9].

A web-based multi-page survey was developed to investigate the attitudes of a large community of orthopaedic surgeons (members of the European Society of Sports Traumatology, Knee Surgery and Arthroscopy,

ESKA) in terms of the treatments of choice and the management of some of the most common joint diseases.

The survey consists of eight clinical scenarios (4 scenarios of shoulder pathology and 4 scenarios of knee pathology); each scenario was displayed on a different page and was associated with a series of closed-ended questions and qualitative scales using which respondents were requested to assess the appropriateness of a list of alternative options.

The preferences of orthopaedic surgeons for the scenarios relating to the management of clinical shoulder cases have been published in a previous paper [26].

The aim of this paper was to report the preferences relating to clinical scenarios of knee disease management, which were defined in order to investigate the degree of inclusion of evidence-based medicine in daily clinical practice. A variety of diseases can affect the knee joint and their management can be a challenge for orthopaedic surgeons, especially in the borderline cases that we purposely defined to address situations in which choosing the right option is not a trivial task.

In order to enable hidden common preferences to emerge and make surgeons aware of potential differences in practice, we had to assess the degree of agreement reached by the surgeons involved in the survey. Agreement assessment relates to the degree to which surgeons agree about a specific treatment. One important purpose of this work is to report agreement measurements based on the experimentation and consolidation of heuristics proposed elsewhere in the literature and to introduce a novel measurement of agreement that is more sensitive to a large number of respondents and more easy to calculate. In doing this, this paper contributes to unravelling the potential of web-based surveys administered to large communities of practitioners as an effective tool to help surgeons develop consensus-based and truly practice-based guidelines.

Materials and methods

Questionnaire design

The survey under consideration in this paper consisted of 4 clinical knee scenarios (Tables 1, 2, 3, and 4), each displayed on a different page in order to focus attention on its peculiarities.

An international association of surgeons specialising in sports traumatology and knee surgery (ESSKA) was contacted. All the people on the official mailing list of the ESSKA association were regarded as eligible and 1,084 personalised invitations were sent (target population).

For each scenario, the respondents were asked to state whether they would opt for either surgical or conservative treatment on the basis of the available information. Then, according to this first choice, each respondent was asked either to select his/her treatment of choice or to rank alternative treatments; more precisely:

1. Respondents were invited to choose only one of the treatments made available on a list;
2. Respondents were invited to quantify, on an ordinal scale how appropriate they found the treatments for the scenario under consideration.

Two null hypotheses, closely related to the corresponding research questions, were formulated. A categorical variable (nominal) was defined for each item and called 'treatment of choice for the scenario under consideration' and the following null hypothesis was formulated:

$H_0(A)$ all treatments for the scenario under consideration were regarded as equivalent.

Rejecting this hypothesis (chi-square test of the equality of the frequency) is regarded as being the same as claiming that respondents displayed a preference among the treatments made available on a list.

In order to establish whether a proposed treatment was deemed as being significantly relevant to the clinical scenario under consideration, a variable called 'appropriateness of a treatment for the scenario under consideration' was defined. This variable was quantified on an ordinal scale as follows: totally inappropriate, -2; inappropriate; appropriate, 1; and very appropriate, 2. In this way, a negative number would be assigned to a treatment that was not considered appropriate for the clinical scenario under consideration. The following null hypothesis was then formulated:

$H_0(B)$ the proposed treatment is not deemed appropriate for the scenario under consideration.

Rejecting this hypothesis (one-sided sign test on the median being less than or equal to -1) means that the proposed treatment was deemed appropriate by the respondents.

Members of the ESSKA association were contacted by email twice: the first time to present the research initiative and to invite each member to participate in the initiative and complete the online questionnaire; the second time to send a reminder to join the initiative only to those members who had not responded to the survey by that time.

The survey was kept open for 24 days, from 12 May to 4 June 2010; on 26 May (2 weeks after the first invitation), the reminder email was sent and this produced a further 63 % of responses. The date on which the reminder was sent was decided on the basis of considerations aimed at reducing non-response bias [26].

An open-source platform (http://www.limesurvey.org/) was configured to collect the responses anonymously.

Table 1 Survey results for Scenario 1

Scenario 1: 56-year-old male. 175 cm, 78 kg. Pain 7/10 for 7 months. Varus (8°) knee. Lachman−, Posterior drawer test−, Previous medial arthroscopic partial meniscectomy 15 years before. Imaging suggestive of degenerative medial compartment (diffused cartilage damage—on both the femoral and tibial side). Good lateral and femoro-patellar compartments.

	Proportions (%) [95 % CI]	POOA	NX2A
1. What type of treatment would you choose?			
a. Conservative	20 [16.9, 23.1]	0.67	0.35 poor
b. Surgical	80 [76.3, 83.6]		
2. You preferred surgical treatment, how appropriate do you think each of the following techniques is?			
I. High tibial osteotomy (opening wedge)			
a. Very appropriate	54 [49.2, 58.8]	0.84	0.68 moderate
b. Appropriate	37 [32.3, 41.6]		
c. Inappropriate	8 [5.4, 10.6]		
d. Totally inappropriate	1 [0.04, 2]		
II. High tibial osteotomy (closing wedge)			
a. Very appropriate	19 [15, 23]	0.59	0.18 poor
b. Appropriate	52 [47, 57]		
c. Inappropriate	26 [21.6, 30.4]		
d. Totally inappropriate	3 [1.3, 4.7]		
III. Unicompartimental knee prosthesis			
a. Very appropriate	21 [16.9, 25.1]	0.55	0.11 poor
b. Appropriate	46 [41, 51]		
c. Inappropriate	26 [21.6, 30.4]		
d. Totally inappropriate	7 [4.4, 9.6]		
IV. Procedures based on biological response			
a. Very appropriate	2 [0.6, 3.4]	0.84	0.68 moderate
b. Appropriate	7 [4.4, 9.6]		
c. Inappropriate	45 [40, 50]		
d. Totally inappropriate	46 [41, 51]		
V. Microfractures			
a. Very appropriate	5 [2.8, 7.2]	0.63	0.26 poor
b. Appropriate	19 [15, 23]		
c. Inappropriate	44 [38.9, 49.1]		
d. Totally inappropriate	32 [27.2, 36.8]		
3. You preferred conservative treatment, how appropriate do you find each of the following options?			
I. Quadriceps strengthening			
a. Very appropriate	46 [35.4, 56.6]	0.85	0.71 moderate
b. Appropriate	46 [35.4, 56.6]		
c. Inappropriate	7 [1.5, 12.5]		
d. Totally inappropriate	1 [3.1, 1.1]		
II. Cycle of hyaluronic acid			
a. Very appropriate	30 [20, 40]	0.62	0.25 poor
b. Appropriate	45 [34.2, 55.8]		
c. Inappropriate	20 [11.3, 28.7]		
d. Totally inappropriate	5 [0.3, 9.7]		

Table 1 continued

Scenario 1: 56-year-old male. 175 cm, 78 kg. Pain 7/10 for 7 months. Varus (8°) knee. Lachman−, Posterior drawer test−, Previous medial arthroscopic partial meniscectomy 15 years before. Imaging suggestive of degenerative medial compartment (diffused cartilage damage—on both the femoral and tibial side). Good lateral and femoro-patellar compartments.

	Proportions (%) [95 % CI]	POOA	NX2A
III. Cycle of intra-articular cortisone injections			
a. Very appropriate	9 [3, 15]	0.5	0.03 poor
b. Appropriate	31 [20.8, 41.2]		
c. Inappropriate	37 [26.4, 47.6]		
d. Totally inappropriate	23 [13.7, 32.3]		
IV. Swimming and cycling			
a. Very appropriate	49 [38.3, 59.7]	0.97	0.94 excellent
b. Appropriate	50 [39.3, 60.7]		
c. Inappropriate	1 [0, 3.1]		
d. Totally inappropriate	0		
V. Active assisted physiotherapy			
a. Very appropriate	48 [37.1, 58.9]	0.89	0.79 excellent
b. Appropriate	47 [36.1, 57.9]		
c. Inappropriate	4 [0, 8.3]		
d. Totally inappropriate	1 [0, 3.3]		
4. *Irrespective of the conservative treatment you think is the most appropriate, how would you proceed in this case?*			
a. Re-evaluate surgery if cons. treatment fails in 6 months	60 [49.7, 70.3]	0.43	0.24 poor
b. Re-evaluate surgery if cons. treatment fails in 12 months	27 [17.6, 36.4]		
c. Re-evaluate surgery if cons. treatment fails in 24 months	5 [0.4, 9.6]		
d. Surgery has to be excluded for a medium-term period	8 [2.3, 13.7]		

Response rates are reported as proportions with 95 confidence intervals. *NX2A*: normalised Chi-square-based agreement, *Po* percentage of overall agreement. Po and NX2A for ordinal variables were both calculated considering 2 categories ('very appropriate or appropriate' vs. 'inappropriate or totally inappropriate')

In overall terms, at the end of the survey, 374 fully completed questionnaires (36 % of the target population) and 38 partially completed questionnaires were collected.

Agreement assessment

In order to measure the strength of consensus among the raters involved, an inter-rater agreement score was calculated for each question.

Inter-rater agreement relates to the extent to which different evaluating surgeons, each assessing the same clinical scenario, come to the same decision, that is, either select the same treatment or assign the same assessment category (appropriate and not appropriate) to the treatment under consideration.

In the specialist literature, several methods for measuring inter-rater agreement have been proposed [8, 17]. The simplest and most common measurement of inter-rater agreement is the observed proportion of agreement (percentage of overall agreement, Po in what follows POOA). POOA provides an estimation of the probability of two (randomly selected) raters assigning the same appropriateness grade to a given treatment. However, this statistical measurement does not take account of the agreement that would have been expected due solely to chance [13] and it usually overestimates agreement. To assess agreement between multiple respondents, several coefficients of concordance have been proposed [1]. At the same time, common statistical software packages do not provide functionalities to calculate these scores and they can be perceived as difficult by non-specialist users, as surgeons usually are.

Consequently, to make the assessment of agreement easier, we created the nomogram depicted in Fig. 1. The 'Normalised Chi-square based Agreement' (NX2A) makes it possible to achieve a univocal level of agreement (the dependent variable) as a simple linear function of an independent variable that is obtained by normalising the chi-square score associated with the proportions of responses (namely, k) with respect to the number of respondents involved (namely, n).

The nomogram was obtained by correlating scores from the free-marginal multirater kappa, which is an indicator of

Table 2 Survey results for Scenario 2

Scenario 2: 35-year-old male. 185 cm, 100 kg. Pain 8/10. Early knee medial arthritis. Stable. Arthroscopic medial meniscectomy on a buckled handle lesion at 21 years of age.

	Rates and 95 % CI (%)	POOA	NX2A
1. *What type of treatment would you choose for the case mentioned above?*			
a. Conservative	13 [10.4, 15.7]	0.77	0.54 moderate
b. Surgical	87 [84.4, 89.7]		
2. *You preferred conservative treatment. How appropriate do you think each of the following options is?*			
I. Quadriceps strengthening			
a. Very appropriate	56 [42.3, 69.7]	0.88	0.76 excellent
b. Appropriate	38 [24.6, 51.4]		
c. Inappropriate	6 [0, 12.6]		
d. Totally inappropriate	0		
II. Cycle of high-weight hyaluronic acid			
a. Very appropriate	24 [11.8, 36.2]	0.56	0.14 poor
b. Appropriate	45 [30.8, 59.2]		
c. Inappropriate	22 [10.1, 33.9]		
d. Totally inappropriate	9 [0.8, 17.2]		
III. Cycle of low-weight hyaluronic acid			
a. Very appropriate	15 [4.2, 20.8]	0.57	0.17 poor
b. Appropriate	56 [41, 71]		
c. Inappropriate	19 [7.2, 30.8]		
d. Totally inappropriate	10 [1, 19]		
IV. Swimming and cycling			
a. Very appropriate	53 [39.6, 66.4]	0.96	0.92 excellent
b. Appropriate	45 [31.7, 58]		
c. Inappropriate	2 [0, 5.7]		
d. Totally inappropriate	0		
V. Active assisted physiotherapy			
a. Very appropriate	52 [38.5, 65.5]	0.88	0.77 excellent
b. Appropriate	42 [28.6, 55.4]		
c. Inappropriate	4 [0, 9.3]		
d. Totally inappropriate	2 [0, 5.8]		
3. *Irrespective of which of the conservative treatment mentioned above is the most appropriate, how would you proceed in this case?*			
a. Re-evaluate surgery if cons. treatment fails in 6 months	58 [44.6, 71.4]	0.39	0.20 poor
b. Re-evaluate surgery if cons. treatment fails in 12 months	20 [9.1, 30.9]		
c. Re-evaluate surgery if cons. treatment fails in 24 months	14 [4.6, 23.4]		
d. Surgery has to be excluded in a middle-term period	8 [0.7, 15.3]		
4. *You chose surgical treatment. How appropriate do you consider each of the following methods?*			
I. High tibial osteotomy (opening wedge)			
a. Very appropriate	42 [37.3, 46.7]	0.63	0.27 poor
b. Appropriate	35 [30.4, 39.6]		
c. Inappropriate	16 [12.5, 19.5]		
d. Totally inappropriate	7 [4.6, 9.4]		
II. High tibial osteotomy (closing wedge)			
a. Very appropriate	1 [0, 2]	0.88	0.76 excellent
b. Appropriate	5 [2.8, 7.2]		
c. Inappropriate	39 [34.1, 43.9]		
d. Totally inappropriate	55 [50, 60]		

Table 2 continued

Scenario 2: 35-year-old male. 185 cm, 100 kg. Pain 8/10. Early knee medial arthritis. Stable. Arthroscopic medial meniscectomy on a buckled handle lesion at 21 years of age.

	Rates and 95 % CI (%)	POOA	NX2A
III. Unicompartimental knee prosthesis			
a. Very appropriate	10 [6.9, 13.1]	0.50	0.02 poor
b. Appropriate	33 [28.2, 37.8]		
c. Inappropriate	41 [36, 46]		
d. Totally inappropriate	16 [12.2, 19.8]		
IV. Meniscal allograft			
e. Very appropriate	16 [12.2, 19.8]	0.51	0.03 poor
f. Appropriate	43 [37.8, 48.2]		
g. Inappropriate	31 [26.2, 35.8]		
h. Totally inappropriate	10 [6.9, 13.1]		
V. ACI			
a. Very appropriate	20 [16.2, 23.8]	0.56	0.13 poor
b. Appropriate	49 [44.2, 53.8]		
c. Inappropriate	24 [19.9, 28.1]		
d. Totally inappropriate	7 [4.6, 9.4]		
VI. Microfractures			
a. Very appropriate	13 [9.6, 16.4]	0.51	0.03 poor
b. Appropriate	46 [40.9, 51.1]		
c. Inappropriate	30 [25.3, 34.7]		
d. Totally inappropriate	11 [7.8, 14.2]		

Response rates are reported as proportions with 95 confidence intervals. *NX2A* normalised Chi-square based agreement, *Po* percentage of overall agreement. Po and NX2A for ordinal variables were both calculated considering 2 categories ('very appropriate or appropriate' vs. 'inappropriate or totally inappropriate')

ACI autologous chondrocyte implantation

collective agreement that has recently been developed and validated [30], with the ratio between the Chi-square scores associated with the proportions of responses (k) and the number of respondents (n). The functional relationship has been verified for number of respondents (n) larger than 30. The reason why we based our nomogram on the chi-square is that this is a common score in the medical community. All the main statistical software packages calculate a score of this kind (even common spreadsheets like MS Excel) and a number of online calculators are also available (e.g. http://www.jspearson.com/Science/chiSquare.html), as well as a number of look-up, precomposed tables.

Considering the question 2 of scenario 4, the NX2A is obtained following this procedure:

- $k = 3$ (no. of categories from which respondents could make their choice); $n = 348$ (no. of respondents).
- Chi-square score $= 315.7$
- Chi-square score/$n = 0.9$
- Looking to the line associated with $k = 3$ (Fig. 1), NX2A $= [1/(3-1)] \times 0.9$ and it is equal to 0.45 (moderate).

Statistical analysis

The responses collected during the survey have been summarised in terms of the proportions of respondents for each alternative option with respect to the total number of actual respondents; the associated 95 % confidence intervals are also reported. A finite population correction factor was used when the sample size was large enough in comparison with the population size. Statistical analyses were carried out using the SPSS package. Equality of proportions has been tested using a chi-square test. For these and all the other tests mentioned below, a conventional confidence level of 95 % (p value < 0.05) has been adopted in order to regard the results as being statistically significant.

Results

Inferential statistics

Scenario 1 The surgical approach was preferred to the conservative approach to a significant degree ($p < 0.001$).

Table 3 Survey results for Scenario 3

Scenario 3: 46-year-old male. Soccer player—recreational level. Employee. Acute knee sprain 1 month ago. Lachman +, pivot shift +++, MRI consistent with ACL lesion. No meniscal-associated symptoms. No swelling.

	Rates and 95 % CI (%)	POOA	NX2A
1. *What type of treatment would you choose for the case mentioned above?*			
a. Conservative	54 [50, 58]	0.50	0.005 poor
b. Surgical	46 [42, 50]		
2. *You preferred conservative treatment. How appropriate do you think each of the following methods is?*			
I. Quadriceps strengthening			
a. Very appropriate	67 [61, 73]	0.9	0.8 excellent
b. Appropriate	28 [22.3, 33.7]		
c. Inappropriate	4 [1.5, 6.5]		
d. Totally inappropriate	1 [0, 2.3]		
II. Swimming and cycling			
a. Very appropriate	72 [66.3, 77.7]	0.95	0.9 excellent
b. Appropriate	26 [20.5, 31.5]		
c. Inappropriate	2 [0.2, 3.8]		
d. Totally inappropriate	0		
III. Active-assisted physiotherapy			
a. Very appropriate	70 [64.2, 75.8]	0.93	0.86 excellent
b. Appropriate	26 [20.5, 31.5]		
c. Inappropriate	4 [1.5, 6.5]		
d. Totally inappropriate	0		
3. *You preferred surgical treatment. How appropriate do you think each of the following techniques is?*			
I. Arthroscopic ACL reconstruction with an ST-G autologous graft			
a. Very appropriate	60 [53, 67]	0.9	0.8 excellent
b. Appropriate	35 [28.1, 41.9]		
c. Inappropriate	4 [1.2, 6.8]		
d. Totally inappropriate	1 [0, 2.4]		
II. Arthroscopic ACL reconstruction with a BTB autologous graft			
a. Very appropriate	29 [22, 36]	0.75	0.5 moderate
b. Appropriate	57 [49.3, 64.7]		
c. Inappropriate	11 [6.1, 15.9]		
d. Totally inappropriate	3 [0.4, 5.6]		
III. Arthroscopic ACL reconstruction with an autologous quadriceps patellar graft			
a. Very appropriate	16 [10.2, 21.8]	0.57	0.14 poor
b. Appropriate	53 [45, 61]		
c. Inappropriate	26 [19, 33]		
d. Totally inappropriate	5 [1.5, 8.5]		
IV. Arthroscopic ACL reconstruction with an allograft			
a. Very appropriate	9 [4.4, 13.6]	0.49	0.0002 poor
b. Appropriate	42 [34.1, 49.9]		
c. Inappropriate	31 [23.6, 38.4]		
d. Totally inappropriate	18 [11.9, 24.1]		
V. Arthroscopic ACL reconstruction—single bundle			
a. Very appropriate	59 [52, 66]	0.87	0.75 excellent
b. Appropriate	34 [27, 41]		
c. Inappropriate	6 [2.5, 9.5]		
d. Totally inappropriate	1 [0, 2.5]		

Table 3 continued

Scenario 3: 46-year-old male. Soccer player—recreational level. Employee. Acute knee sprain 1 month ago. Lachman +, pivot shift +++, MRI consistent with ACL lesion. No meniscal-associated symptoms. No swelling.

	Rates and 95 % CI (%)	POOA	NX2A
VI. Arthroscopic ACL reconstruction—double bundle			
a. Very appropriate	20 [13.8, 26.2]	0.5	0.06 poor
b. Appropriate	43 [35.3, 50.7]		
c. Inappropriate	29 [21.9, 36.1]		
d. Totally inappropriate	8 [3.8, 12.2]		

Response rates are reported as proportions with 95% confidence intervals. *NX2A* normalised chi-square-based agreement, *Po* percentage of overall agreement. Po and NX2A for ordinal variables were both calculated considering 2 categories ('very appropriate or appropriate' vs. 'inappropriate or totally inappropriate')

ST-G semitendinosus/gracilis tendons, *BTP* bone-patellar tendon-bone

Table 4 Survey results for Scenario 4

Scenario 4: 69-year-old male. Diffuse knee arthritis (all the compartments). Relative improvement from conservative treatments performed elsewhere, VAS 8 to 5. He has returned complaining of pain that has recurred.

	Rates and 95 % CI (%)	POOA	NX2A
1. *Would you consider the option of proposing a total knee prosthesis?*			
a. Yes	96 [94.4, 97.6]	0.91	0.84 excellent
b. No	4 [2.4, 5.6]		
2. *Which kind of implant would you perform?*			
a. Cemented total knee	78 [74.4, 81.6]	0.62	0.45 moderate
b. Uncemented total knee	9 [6.5, 11.5]		
c. Hybrid implant	13 [10.1, 15.9]		
a. Posterior cruciate-retaining implant	48 [43.6, 52.4]	0.49	0.0008 poor
b. Posterior-stabilized total knee prosthesis	52 [47.6, 56.4]		

Response rates are reported as proportions with 95% confidence intervals. *NX2A* normalised chi-square-based agreement, *Po* percentage of overall agreement. Po and NX2A for ordinal variables were both calculated considering 2 categories ('very appropriate or appropriate' vs. 'inappropriate or totally inappropriate')

Fig. 1 This figure illustrates the relationship between the agreement measurement and the normalised Chi-square score, for a set of significant numbers of categories ($k = 2, 3, 4, 5$ and 6). Values of agreement range from 0 to 1 and increase linearly with the Chi-square score. The coefficient of proportionality (slope of the lines) is $1/(k-1)$. Values of agreement less than 0.4 are associated with a 'poor agreement' label, values between 0.4 and 0.75 with 'moderate agreement' and values above 0.75 with 'excellent agreement' [13]. In the figure is shown the inter-rater agreement for the scenario 4 (question 2)
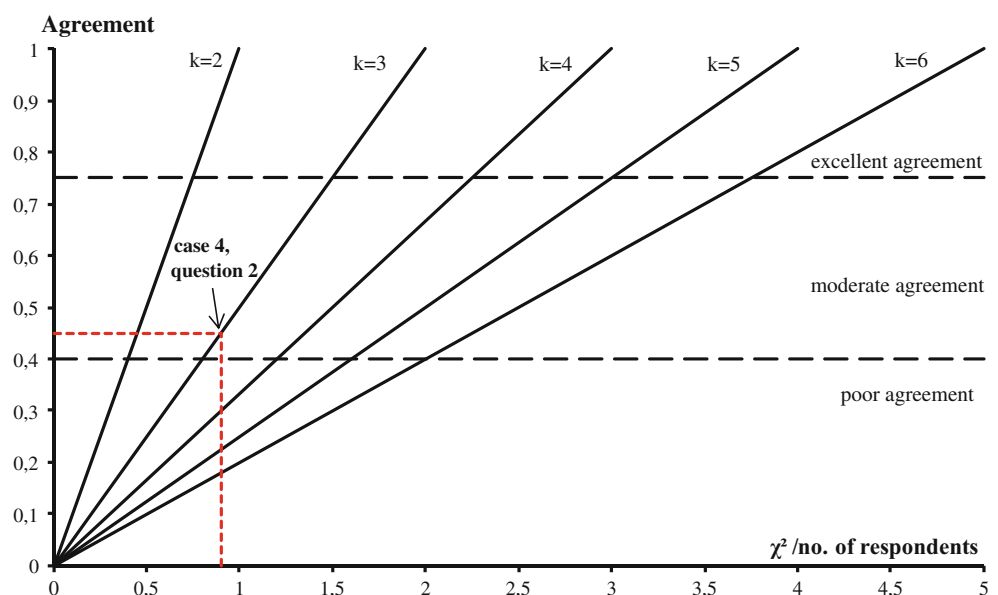
Procedures based on biological response and microfractures were not considered to be appropriate (n.s.). A cycle of intra-articular cortisone injections was not an appropriate conservative option (n.s.).

Among the surgeons those chose a conservative approach, a significantly higher rate of responders preferred to re-evaluate surgery within 6 months ($p = 0.001$).

Scenario 2   The surgical treatment was preferred to conservative treatment ($p < 0.001$). Tibial osteotomy (closing wedge) and a unicompartimental knee prosthesis were not regarded as appropriate surgical options for the management of this clinical scenario (n.s.). The surgeons that preferred a conservative approach chose to re-evaluate surgery in 6 months to a significant degree ($p = 0.001$).

Scenario 3   No significant preference emerged between the surgical and conservative approaches (n.s.). All the proposed conservative treatments were regarded as adequate for the management of this scenario. Among the surgical treatments, arthroscopic ACL reconstruction with an allograft was not regarded as an appropriate option (n.s.).

Scenario 4   Respondents preferred a surgical approach to a significant degree and chose to consider a total knee prosthesis for the management of this clinical scenario ($p < 0.001$).

They preferred a cemented prosthesis ($p < 0.001$), but no preference emerged between a posterior-stabilized implant and a posterior cruciate-retaining design (n.s.).

Agreement assessment

Proportions of responses are summarised in (Tables 1, 2, 3 and 4). Values of agreement based on the POOA and NX2A nomogram are also shown.

Scenario 1   Although there was a high observed proportion in favour of surgical management, the level of agreement among surgeons was poor. It is interesting to observe that almost perfect agreement was obtained for the appropriateness of aerobic exercise.

Scenario 2   A high tibial osteotomy (closing wedge) was not regarded as an appropriate treatment, with an excellent degree of agreement among surgeons.

Scenario 3   There was an almost perfect lack of agreement (poor agreement) for the management of this scenario, as respondents showed no preference at all. Interestingly, almost perfect agreement with regard to the appropriateness of arthroscopic single-bundle ACL reconstruction with a semitendinosus/gracilis graft was reached by the surgeons.

Scenario 4   Almost perfect agreement in favour of a total knee prosthesis was obtained for the management of this scenario.

Discussion

The main aim of this study was to analyse the preferences of a homogeneous community of surgeons with respect to clinical scenarios by means of a specifically designed web-based questionnaire.

Scenario 1

Cartilage repair is only indicated for focal cartilage defects, which can been seen as a precursor of osteoarthritis [27]. In accordance with the literature, biological procedures were not considered appropriate for the management of this clinical scenario, and in this respect, the respondents achieved a moderate degree of agreement.

A cycle of intra-articular cortisone injections was not regarded as appropriate conservative treatment. This finding is supported by the literature that shows that corticosteroid injections are effective in the short term, whereas the benefits in relation to pain and function have not been confirmed in the long term [2, 3].

According to the literature, while strengthening appears to be better treatment in comparison with aerobic exercise in the short term for specific impairment-related outcomes (e.g. pain), aerobic exercise appears to be more effective for functional outcomes in the longer term [4]; an excellent degree of agreement was obtained among surgeons for the appropriateness of the latter conservative treatment.

Scenario 2

The use of autologous chondrocyte implantation (ACI) and other chondral resurfacing techniques is becoming increasingly widespread. However, at the present time, there is insufficient evidence to draw conclusions about the use of ACI for treating articular cartilage defects in the knee [29, 31].

According to these findings, the degree of agreement among the respondents that regarded this treatment as appropriate was low (i.e. 'poor agreement'); this suggests that further randomised controlled trials should be conducted on this topic.

Osteotomy is one of the treatment options for unicompartmental osteoarthritis of the knee. With regard to the kind of techniques, we detected excellent agreement among the surgeons who did not consider the traditional

closed-wedge tibial osteotomy appropriate for the management of this scenario.

This is hardly surprising, as the classic lateral closing-wedge procedure requires a fibular osteotomy that is associated with a risk of damage to the peroneal nerve that is reported to be as high as 11 % [27]. Additional drawbacks include limb shortening, extensive surgical dissection, additional morbidity of a fibular osteotomy and complications in relation to the subsequent placement of the tibial component of a total knee replacement.

On the other hand, although poor agreement was obtained among the respondents, surgeons considered the opening-wedge treatment appropriate, in agreement with several authors [20, 28] who suggest this technique for the valgisation of varus osteoarthritis.

Scenario 3

The evidence from randomised trials to determine whether surgery or conservative management is preferable for ACL injury is insufficient [19]. According to the literature, the preferences of respondents to the survey were distributed fairly evenly between the conservative and surgical options (54 vs. 46 %) and, accordingly, the agreement score was very low.

Among the surgical treatments, arthroscopic ACL reconstruction with an allograft was not regarded as an appropriate option (NS). One possible explanation of this finding may be related to the concern regarding potential complications from the allograft tissue. However, a recent literature review has not found a statistically significant difference in terms of failure rate between autografts and allograft tissue (4.7 vs 8.2 %) [11].

Although a recent meta-analysis has concluded that the current evidence is insufficient to recommend whether a bone-patellar tendon-bone graft or a semitendinosus/gracilis graft is better for ACL reconstruction [23], an excellent degree of agreement on the appropriateness of treatment has been obtained for the latter graft. The lower donor-site morbidity seen in the literature in the case of hamstring autografts could be a factor driving the preferences of the surgeons [7].

Almost perfect agreement on the appropriateness of single-bundle treatment was observed among the surgeons. On the other hand, agreement was poor when it came to the double-bundle reconstruction. Two factors can be considered as playing a role in favour of this finding: on the one hand, there is a lack of evidence relating to the clinical advantages of a double-bundle ACL reconstruction compared with a single-bundle ACL reconstruction [22]; on the other hand, a double-bundle reconstruction can add unnecessary complexity to the surgical procedure [5, 6].

Scenario 4

This is one of the scenarios in which the respondents reached a clear consensus in favour of a surgical approach, although they were obviously unaware of one another's views. The high polarisation of preferences is associated with an almost perfect agreement score.

With regard to the kind of implant, the respondents expressed a clear preference for the cemented total knee implant. There is evidence in the literature of an improved survival rate for the cemented implants compared with the uncemented ones (odds ratio: 4.2, $p < 0.001$) [12], and this could be the main reason for the collective preferences of the surgeons with regard to the management of this clinical scenario.

With regard to the choice of whether to use a posterior cruciate-retaining design or a posterior-stabilized design for total knee arthroplasty, no explicit recommendation can be found in the literature [15]. According to this lack of evidence, the respondents were split fairly evenly in favour of the two techniques and a poor level of agreement was reached; this could suggest that further clinical studies need to be conducted on this specific topic in the near future.

Agreement assessment

Our study has confirmed other work [16] relating to the fact that online survey systems are a flexible tool for collecting the preferences and attitudes of doctors towards appropriate treatments in medical practice.

The ideally 'best treatment' or 'treatment of choice' for the varying levels of severity of joint injuries is a challenging topic, as clear treatment recommendations are either frequently not available, not agreed on or just spread among the orthopaedic community. This condition can be found in almost any medical setting [24, 32].

In the literature, there are some reports of investigations in which web-based questionnaires have aimed to investigate consensus on complex and borderline cases and to determine whether surgeons agree on the kind of patient who needs surgery and the type of surgery that should be recommended for treatment in clinical scenarios of this kind [18, 21].

Our work is intended to act as an advance with respect to those contributions where agreement is identified on the basis of a merely conventional threshold (or cut-off value) of the proportion of respondents in favour of a given option, where agreement is only limited to two alternative options, where no measurement of agreement among the respondents is given or where measurements of agreement that have been acknowledged to be prone to overestimation or other systematic biases are adopted [10, 14, 21].

The proposed NX2A nomogram makes it possible to overcome naïve considerations that are based on the mere proportion of preferences and to obtain access to indications with the same strength as those based on the free-marginal multirater kappa [30]. For instance, in Table 1, it would be wrong to consider the agreement obtained by a community, where 80 % of its respondents prefer one option (i.e. surgical), as 'good agreement', as in this case, other considerations (relating to the number of possible alternatives, the number of respondents and the influence of chance in the preferences, for example, to mention only the main factors) that actually make a percentage of this kind indicative of 'poor agreement' would be neglected.

The NX2A nomogram provides doctors with both an ordinal and an interval evaluation of the detected inter-rater agreement (see the vertical axis in Fig. 1). The former kind of indication is useful in terms of making sense, the appropriateness of the indicator and communication with patients; the latter is useful in order to compare the agreement reached in different groups (independent samples) or to look for any within-group variation in a pretest and post-test fashion (paired samples).

Furthermore, this kind of reliable information on collective agreement could be used by surgeons when counselling patients on the treatments that are available for their joint pathologies. For variables on which agreement is poor or moderate, surgeons may want to advise patients about the objective variability (or uncertainty) of choice among a large number of orthopaedic surgeons. Conversely, for variables on which excellent agreement is achieved, surgeons may confidently advise patients that there is consensus among their peers about the treatment that should be chosen in cases similar to the one under discussion. In this respect, it should, however, be made clear that excellent agreement does not necessarily imply a good outcome per se; it can instead be presented as a simple, pragmatic (and understandable for the patient as well, we believe) indication of the technique almost any referred surgeon would prefer to use. Obviously, areas of significant clinical uncertainty (or disagreement) should be the focus of future research, or more intensive medical education and training for orthopaedic surgeons who treat the kind of injuries that were the subject of the survey.

Our point is that, when backed up by an objective and quantitative assessment of their strength within a large community of practitioners (as in the case of the NX2A measurement), collective indications could be related to a kind of 'four-and-a-half' level of evidence, that is, a level of 'evidence' that is based on the 'consensus of many experts', rather than on the opinion of a few, albeit respected, experts (level of evidence 5); as such, we believe this evidence can correctly summarise the indications of choice that are actually applied within even wide communities of

practitioners or, as in our case, within a whole medical association/society.

Conclusion

The assessment of agreement by the proposed NX2A nomogram has a twofold aim. On the one hand, it can foster discussion among the surgeons in those areas in which adequate evidence exists in the literature to support clinical decision-making. In this case, a certain degree of disagreement among orthopaedic surgeons could be attributed to a lack of adequate peer-reviewed literature on the topic; to an existing controversy among available scientific publications; to the inadequate dissemination and adoption of this evidence; to a conservative attitude and preference to rely on personal experience and tradition; or to a combination of these factors. On the other hand, a tool of this kind, when it is employed in the interpretation of even complex (i.e. multi-page, multi-item and multi-option) online surveys, could help scientific committees discuss and propose indications for practice. These considerations provide the rationale for a possible extension of the classical taxonomy of levels of evidence that is adopted in orthopaedic research [25].

References

1. Banerjee M, Capozzoli M, McSweeney L, Sinha D (1999) Beyond kappa: a review of interrater agreement measures. Can J Stat 27(1):3–23
2. Bannuru RR, Natov NS, Obadan IE, Price LL, Schmid CH, McAlindon TE (2009) Therapeutic trajectory of hyaluronic acid versus corticosteroids in the treatment of knee osteoarthritis: a systematic review and meta-analysis. Arthritis Rheum 61(12):1704–1711
3. Bellamy N, Campbell J, Robinson V, Gee T, Bourne R, Wells G (2006) Intraarticular corticosteroid for treatment of osteoarthritis of the knee. Cochrane Database Syst Rev (2):CD005328
4. Bennell K, Hinman R (2005) Exercise as a treatment for osteoarthritis. Curr Opin Rheumatol 17(5):634–640
5. Beynnon BD, Johnson RJ, Abate JA, Fleming BC, Nichols CE (2005) Treatment of anterior cruciate ligament injuries, part 2. Am J Sports Med 33(11):1751–1767
6. Beynnon BD, Johnson RJ, Abate JA, Fleming BC, Nichols CE (2005) Treatment of anterior cruciate ligament injuries, part I. Am J Sports Med 33(10):1579–1602
7. Biau DJ, Tournoux C, Katsahian S, Schranz PJ, Nizard RS (2006) Bone-patellar tendon-bone autografts versus hamstring autografts for reconstruction of anterior cruciate ligament: meta-analysis. BMJ 332(7548):995–1001
8. Cabitza F (2012) Harvesting collective agreement in community oriented surveys: the medical case. In: Dugdale J, Masclet C, Grasso MA, Boujut JF, Hassanaly P (eds) From research to practice in the design of cooperative systems: results and open challenges. Springer, London

9. Eysenbach G, Wyatt J (2002) Using the internet for surveys and health research. J Med Internet Res 4(2):E13

10. Feinstein AR, Cicchetti DV (1990) High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 43(6):543–549

11. Foster TE, Wolfe BL, Ryan S, Silvestri L, Kaye EK (2010) Does the graft source really matter in the outcome of patients undergoing anterior cruciate ligament reconstruction? An evaluation of autograft versus allograft reconstruction results: a systematic review. Am J Sports Med 38(1):189–199

12. Gandhi R, Tsvetkov D, Davey JR, Mahomed NN (2009) Survival and clinical function of cemented and uncemented prostheses in total knee replacement: a meta-analysis. J Bone Joint Surg Br 91(7):889–895

13. Gwet K (2001) Handbook of inter-rater reliability. STATAXIS Publishing Company, Gaithersburg

14. Hunt RJ (1986) Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. J Dent Res 65(2):128–130

15. Jacobs WC, Clement DJ, Wymenga AB (2005) Retention versus sacrifice of the posterior cruciate ligament in total knee replacement for treatment of osteoarthritis and rheumatoid arthritis. Cochrane Database Syst Rev (4):CD004803

16. Kellerman SE, Herold J (2001) Physician response to surveys. A review of the literature. Am J Prev Med 20(1):61–67

17. Krippendorff K (2004) Content analysis: an introduction to its methodology. Sage Publications Inc., Thousand Oaks

18. Lee JY, Hohl JB, Fedorka CJ, Devin C, Brodke DS, Branch CL Jr, Vaccaro AR (2011) Surgeons agree to disagree on surgical options for degenerative conditions of the cervical and lumbar spine. Spine (Phila Pa 1976) 36(3):E203–E212

19. Linko E, Harilainen A, Malmivaara A, Seitsalo S (2005) Surgical versus conservative interventions for anterior cruciate ligament ruptures in adults. Cochrane Database Syst Rev (2):CD001356

20. Luites JW, Brinkman JM, Wymenga AB, van Heerwaarden RJ (2009) Fixation stability of opening-versus closing-wedge high tibial osteotomy: a randomised clinical trial using radiostereometry. J Bone Joint Surg Br 91(11):1459–1465

21. Marx RG, Jones EC, Angel M, Wickiewicz TL, Warren RF (2003) Beliefs and attitudes of members of the American Academy of Orthopaedic Surgeons regarding the treatment of anterior cruciate ligament injury. Arthroscopy 19(7):762–770

22. Meredick RB, Vance KJ, Appleby D, Lubowitz JH (2008) Outcome of single-bundle versus double-bundle reconstruction of the anterior cruciate ligament: a meta-analysis. Am J Sports Med 36(7):1414–1421

23. Mohtadi NG, Chan DS, Dainty KN, Whelan DB (2011) Patellar tendon versus hamstring tendon autograft for anterior cruciate ligament rupture in adults. Cochrane Database Syst Rev 9:CD005960

24. Mosca L, Linfante AH, Benjamin EJ, Berra K, Hayes SN, Walsh BW, Fabunmi RP, Kwan J, Mills T, Simpson SL (2005) National study of physician awareness and adherence to cardiovascular disease prevention guidelines. Circulation 111(4):499–510

25. Obremskey WT, Pappas N, Attallah-Wasif E, Tornetta P 3rd, Bhandari M (2005) Level of evidence in orthopaedic journals. J Bone Joint Surg Am 87(12):2632–2638

26. Randelli P, Arrigoni P, Cabitza F, Ragone V, Cabitza P (2011) Current practice in shoulder pathology: results of a web-based survey among a community of 1,084 orthopedic surgeons. Knee Surg Sports Traumatol Arthrosc. doi:10.1007/s00167-011-1673-z

27. Ronn K, Reischl N, Gautier E, Jacobi M (2011) Current surgical treatment of knee osteoarthritis. Arthritis 2011:454873

28. Song EK, Seon JK, Park SJ, Jeong MS (2010) The complications of high tibial osteotomy: closing-versus opening-wedge methods. J Bone Joint Surg Br 92(9):1245–1252

29. Vasiliadis HS, Wasiak J (2010) Autologous chondrocyte implantation for full thickness articular cartilage defects of the knee. Cochrane Database Syst Rev (10):CD003323

30. Warrens MJ (2010) Inequalities between multi-rater kappas. Adv Data Anal Classif 4:271–286

31. Wasiak J, Clar C, Villanueva E (2006) Autologous cartilage implantation for full thickness articular cartilage defects of the knee. Cochrane Database Syst Rev (3):CD003323

32. Webster BS, Courtney TK, Huang YH, Matz S, Christiani DC (2005) Physicians' initial management of acute low back pain versus evidence-based guidelines. Influence of sciatica. J Gen Intern Med 20(12):1132–1135