



## Original article

## French translation and cultural adaptation of a questionnaire for patients with hip or knee prosthesis



Roger Erivan<sup>a,\*</sup>, Guillaume Villatte<sup>a</sup>, Thibault Chaput<sup>b</sup>, Aurélien Mulliez<sup>c</sup>,  
Matthieu Ollivier<sup>d</sup>, Stéphane Descamps<sup>a</sup>, Stéphane Boisgard<sup>a</sup>

<sup>a</sup> Université Clermont Auvergne, CHU Clermont-Ferrand, CNRS, SIGMA Clermont, ICCF, 63000 Clermont-Ferrand, France

<sup>b</sup> Université Clermont Auvergne, CHU Clermont-Ferrand, 63000 Clermont-Ferrand, France

<sup>c</sup> Délégation à la Recherche Clinique et aux Innovations (DRCI) - CHU Clermont-Ferrand, 63000 Clermont-Ferrand, France

<sup>d</sup> Aix-Marseille Université, CNRS, ISM UMR 7287, 13288, Marseille cedex 09, France Institute for Locomotion, Department of Orthopedics and Traumatology, St. Marguerite Hospital, 270, boulevard Sainte-Marguerite, BP 29, 13274 Marseille, France

## ARTICLE INFO

## Article history:

Received 8 November 2018

Accepted 11 January 2019

## Keywords:

Orthopedics

Patient follow-up

Questionnaire

Translation

Total hip arthroplasty

Total knee arthroplasty

## ABSTRACT

**Introduction:** The Kingsbury questionnaire offers the possibility of follow-up by means of an X-ray and a simple questionnaire sent to the home address of the patient, who will not need to come in consultation if there are no problems. The questionnaire detects any anomaly in follow-up. In case of anomaly in the questionnaire or radiograph, the patient is contacted and/or seen again, as appropriate. The Kingsbury questionnaire has had no transcultural validation, and we therefore conducted a prospective study in order to 1) translate into French the questionnaire, previously validated in English; 2) adapt it for good understanding according to cultural habits; and 3) assess the translated version on a test-retest procedure.

**Hypothesis:** The study hypothesis was that the translated questionnaire would show good test-retest reproducibility.

**Material and methods:** The exact English version of the questionnaire was obtained directly from the authors of the index publication. A methodology of translation, back-translation and test-retest enabled assessment of the translation and of the reproducibility of the French version. The reference method of cultural adaptation of self-administered questionnaires and patient information documents was used. The questionnaire was tested prospectively.

**Results:** One hundred patients were contacted, providing 73 clinical tests with radiographic validation and 48 complete test-retests in a representative population of total hip and total knee arthroplasty (THA, TKA). Internal coherence showed a KR-20 coefficient of 0.71 and Cronbach alpha of 0.76: e.g., good internal coherence. Item difficulty, requiring renewed contact, was low for all questions. Mean variance was low on the first 7 questions: 0.08 (range, 0.02–0.16). Correlation was close to 0.5 for each question. Analysis of reproducibility found excellent agreement (>90%) for the first 7 questions, which were binary; for question 8, agreement was good (83.3%) considering that there were 5 possible responses. For 19 of the 73 respondents, the questionnaire results indicated a need for further contact. After analysis of their radiographs, 4 needed to be seen in consultation again. The other 15 had unfavorable responses but without deterioration since the last classical consultation or any radiologic abnormality consultation, and were not called back for consultation.

**Conclusion:** The French version of the Kingsbury questionnaire provided reproducible assessment, avoiding the need to call the patient back for consultation unnecessarily. The questionnaire needs validating in a larger sample before being widely used: the present study was just a first step.

**Level of evidence:** IV, Prospective without control group.

© 2019 Elsevier Masson SAS. All rights reserved.

\* Corresponding author. Orthopedic and Trauma Surgery Department, Hôpital Gabriel Montpied, CHU de Clermont Ferrand BP 69, 63003 Clermont Ferrand, France.  
E-mail address: [rerivan@chu-clermontferrand.fr](mailto:rerivan@chu-clermontferrand.fr) (R. Erivan).

## 1. Introduction

The number of total hip and knee arthroplasties (THA, TKA) is constantly increasing and is set to continue doing so in coming years [1–3]. Ever more patients are undergoing arthroplasty, and population aging means that they are less and less able to easily come for follow-up consultations. Follow-up becomes voluntary: it is proposed, but in practice many patients are lost to follow-up or do not comply with the proposal [4]. In the case of prostheses, long-term follow-up is necessary, to detect loosening but also to assess the patient's behavior and monitor behavioral change and new implants.

A study in 2016 showed that follow-up by means of an X-ray and a simple questionnaire sent to the patient's home address was feasible [4], avoiding having to call the patient in if there are actually no problems. The questionnaire detects any anomaly in follow-up. If an anomaly is detected on the questionnaire or on the X-ray, the patient is contacted and/or seen again, depending on the findings. The aim is thus simply to judge whether it is worth calling the patient back for consultation [4]. If all questionnaire responses are favorable and there is no radiographic abnormality, there is no reason to importune the patient: the economic and time savings are thus potentially considerable.

The questionnaire has not been subjected to transcultural validation, and we therefore conducted a prospective study in order to.

- translate into French the questionnaire, previously validated in English;
- adapt it for good understanding according to cultural habits;
- assess the translated version on a test-retest procedure. The hypothesis was that the French version would show good reproducibility.

## 2. Materials and methods

### 2.1. Patients

The study had review board approval (CpP n° 2018-A01022-53). The exact English questionnaire was procured directly from the authors of the index publication [4] (Appendix 1). Requisite sample size was determined to be 50 subjects, following Terwee et al. [5].

The French version was tested prospectively in 100 patients, to ensure test-retest in about 50 patients, considering, in the light of recent similar studies [6], that 25% of patients were liable not to respond to the first questionnaire and 25% to the second. The questionnaires and radiography prescription were sent to the home address of patients meeting the inclusion criteria.

Selection aimed at a representative sample of the target THA/TKA population in terms of gender, postoperative follow-up and urban/rural place of residence. The inclusion criteria were those of our usual 1, 3, 5 and 10 year follow-up, selecting the first 100 patients seen at the beginning of 2018. Questionnaires were then analyzed, and patients were called back in consultation when necessary.

### 2.2. Method

The translation, back-translation, test-retest methodology was used [7,8], with a simple and reproducible method of translation. The translation/back-translation method [9] is the reference for cultural adaptation of self-administered questionnaires and patient information documents.

Translation: two independent English-to-French translations were made, by two native French-speaking translators (RE and TC),

and compared to reach a consensus approved by the head of project (RE).

Back-translation: back-translation from French back into English was performed by PC, a native English-speaker.

Translation meeting: a meeting with the head of project analyzed discrepancies between the source document and the translated/back-translated versions, making necessary changes and reaching a consensus between the members of the translation team.

Cultural adaptations, both linguistic and related to the specificities of the French health and social system, were made, as in previous studies of this type [10–12]. The COSMIN guidelines were also applied in adapting the questionnaire [13].

Test/retest was performed with a 2-week interval, such that patients would not remember their initial responses but would not have changed in clinical status. At first self-administration, patients were not told that there would be a retest 2 weeks later, so as not to bias the initial responses; the need for a retest was, however, explained when the second, identical, questionnaire was sent out. Responses were considered “unfavorable” when they required the patient's file to be re-examined: i.e., a “yes” to questions 1 to 4 or a “no” to questions 5 to 7. Responses to question 8 helped give an overall impression, but did not enter into the interpretation of the results. An unfavorable response to questions 1 to 3 automatically entailed the patient being called back in, while an unfavorable response to questions 4 to 8 entailed closer analysis of the file, with the patient being called back in only if need be. Questionnaires and radiographs were analyzed for respondents to the first questionnaire/radiograph.

### 2.3. Statistics

Statistical analysis used Stata software, v12 (Stata-Corp, Texas, USA) following the validity criteria for studies based on questionnaires [5]. Descriptive statistics were reported as numbers and percentages for categorical variables and mean  $\pm$  standard deviation for continuous variables. Analysis concerned respondents to both successive questionnaires; there were thus no missing questionnaires. In case of missing response on an item, the patient was recontacted to complete the questionnaire; there were thus no missing data. Comparison of means used the Student test (population  $> 30$ ).

Internal coherence was analyzed on the Kuder-Richardson Formula 20 (KR-20) for binary questions (i.e., 1-7). Question 8 had 5 possible responses, and Cronbach's alpha (applicable to non-binary variables) was calculated. In both cases, values ranged from 0 to 1, with  $> 0.60$  considered acceptable,  $> 0.70$  good and  $> 0.90$  excellent [14]. Item difficulty, item variance and item-rest correlation were calculated. Item difficulty corresponded to cases in which the patient's file needed re-examining. Item variance corresponded to criterion variability, with less variability the closer the value to 1 (or 0, as appropriate). Item-rest correlation corresponded to the item's correlation with all other items (except itself), low values meaning the item correlated poorly with the others (i.e., was not measuring the same thing). Variability is greatest for correlations around 0.5. The various items were assessed successively [15].

Reproducibility was first assessed in terms of percentage agreement and Cohen kappa coefficient [16]. As distributions were asymmetric, with many “yes” or many “no” responses, liable to bias the kappa coefficient, Gwet's AC1 coefficient was also calculated [17]. Values  $< 0.20$  indicated slight concordance, 0.21-0.40 acceptable concordance, 0.41-0.60 moderate concordance, 0.61-0.80 good concordance, and  $> 0.81$  near-perfect concordance [18]. The Chi<sup>2</sup> test was used to assess difference in type of implant, gender and

## SUIVI DE PATIENTS DANS LE SUIVI DES ARTHROPLASTIES TOTALES

Nom du patient : \_\_\_\_\_ Date du jour : \_\_\_\_\_

Quelle intervention avez-vous subie ? (entourez la bonne réponse) Prothèse de genou droite / gauche Prothèse de hanche droite / gauche Date de l'intervention : \_\_\_\_\_

Etiquette : \_\_\_\_\_

**1. Avez-vous eu des problèmes avec la prothèse depuis votre dernière visite de contrôle ?** Cochez la bonne case.  
Si oui, expliquez  
Oui  Non

**2. Avez-vous des douleurs en rapport avec votre prothèse de hanche / genou ?**  
Si oui, expliquez et notez sur 10 ex : 2/10  
Oui  Non

**3. Prenez-vous régulièrement des antalgiques à cause de votre prothèse ?**  
Si oui, lesquels ?  
Oui  Non

**4. Utilisez-vous une canne ou des béquilles ?**  
Si oui, décrivez votre mobilité  
Oui  Non

**5. Est-ce que votre opération a amélioré vos activités quotidiennes ?**  
Si non, dites pourquoi  
Oui  Non

**6. Est-ce que votre opération de hanche ou genou a amélioré vos activités personnelles ? par exemple se laver, prendre une douche ...**  
Si non, dites pourquoi  
Oui  Non

**7. L'intervention a-t-elle été bénéfique ?**  
Si non, dites pourquoi Date de l'intervention : \_\_\_\_\_  
Oui  Non

**8. Notez votre satisfaction globale après votre prothèse**



Fig. 1. French version of total arthroplasty patient follow-up questionnaire.

area of residence. All tests were 2-tailed, with the significance threshold set at  $p < 0.05$ .

### 3. Results

Adhering to the translation procedure described in the Methods section resulted in the validated French version of the questionnaire, shown in Fig. 1.

The questionnaire and radiography prescription were first sent to 100 patients, 73 of whom responded by mail and had the X-ray taken. The questionnaire was then sent out to these responders

(without radiography prescription) 2 weeks later, and received 48 responses. The flowchart is shown in Fig. 2.

The 100 patients had 50 THAs and 50 TKAs, and a mean age at arthroplasty of  $70.7 \pm 10.6$  years [range, 44–92 years]. Mean age at questionnaire response was  $75.5 \pm 9.9$  years [range, 54–101 years], for a mean interval of  $3.9 \pm 3.3$  years [range, 1–10 years]. Sixty patients (60%) were female. 56 were urban residents (56%).

The 27 non-respondents were contacted by telephone, to which 11 did not respond. Twelve telephone respondents said they had not replied to the questionnaire as they did not wish to take part in the study; 4 were willing to fill out the questionnaire but not to get and send back the X-ray, for which they would have had to pay

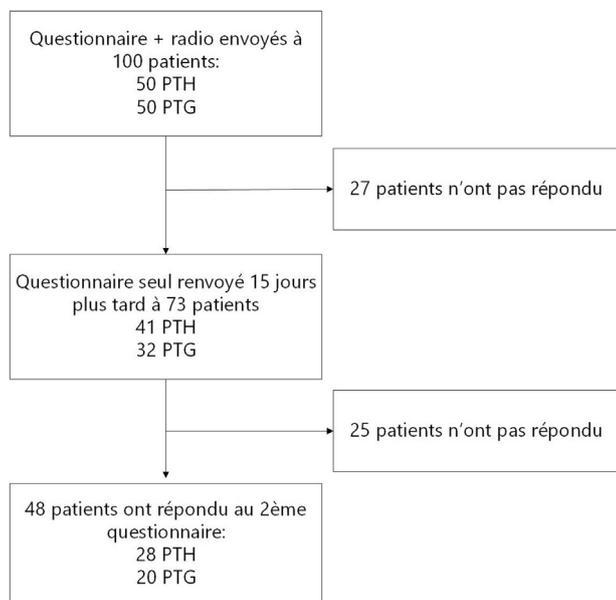


Fig. 2. Flowchart.

postage. The 27 non-respondents comprised 9 THAs and 18 TKAs ( $p=0.12$ ), with mean age at arthroplasty of  $73.7 \pm 9.8$  years [range, 61–92 years] ( $p=0.67$ ) and mean age at the time of the questionnaire of  $78.8 \pm 10.1$  years [range, 63–101 years] ( $p=0.13$ ) for a mean interval of  $3.8 \pm 3.4$  years [range, 1–10 years] ( $p=0.95$ ). Seventeen were female (63.0%) ( $p=0.78$ ), and 13 urban (48%) ( $p=0.47$ ).

The 73 initial respondents (questionnaire + radiograph) had 41 THAs and 32 TKAs ( $p=0.42$ ), mean age at arthroplasty of  $69.5 \pm 10.7$  years [range, 44–91 years] ( $p=0.49$ ) and mean age at questionnaire response of  $74.3 \pm 9.6$  years [range, 54–96 years] ( $p=0.42$ ), for a mean interval of  $4.0 \pm 3.4$  years [range, 1–10 years] ( $p=0.99$ ). Forty-three were female (59%) ( $p=0.89$ ), and 30 urban (41%) ( $p=0.053$ ).

Nineteen of the 73 initial respondents (26.0%) had an unfavorable response among questions 1 to 3, and were therefore automatically recontacted. Thirteen (17.8%) had an unfavorable response among questions 4 to 8, with re-examination of their file and no cases in which they were called back for consultation. Radiographic analysis of the 73 patients found 8 with abnormalities: 3 cases of periprosthetic radiolucency, 2 of wear and 3 of heterotopic ossification. In all, after renewed contact and X-ray analysis, 4 of the 73 patients were called back for consultation. The other 15 of the 19 patients with unfavorable questionnaire responses showed no progression since the last classic follow-up and no radiographic abnormality, and were therefore not called back in. The 4 patients called back for consultation presented pain and periprosthetic radiolucency in 3 cases: 1 around the THA stem and 2 at the TKA tibial plateau, requiring close follow-up; the other patient presented Brooker grade 2 ossification [19], also requiring follow-up.

The 48 patients studied on test-retest had a mean age at arthroplasty of  $69.4 \pm 9.6$  years [range, 44–90 years] ( $p=0.46$ ) and mean age at questionnaire of  $74.1 \pm 8.7$  years [range, 55–92 years] ( $p=0.18$ ), for a mean interval of  $3.8 \pm 3.2$  years [range, 1–10 years] ( $p=0.96$ ). Twenty-five were female (52.1%) ( $p=0.36$ ), and 30 urban (62.5%). There were 28 THAs and 20 TKAs ( $p=0.34$ ).

Test-retest results are shown in Table 1.

Ten of the 48 test-retest respondents (20.8%) had an unfavorable response among questions 1 to 3, with automatic renewed contact at first test, and 9 (18.8%) at retest; 8 (16.7%) had an unfavorable response among questions 4 to 8 with re-examination of their file at first test and 4 (8.3%) on retest, none of which led to renewed consultation. Overall, at both stages, 2 of these 48 patients were

Table 1

Responses on questionnaire test-retest (favorable = not requiring new consultation).

	Question	1st test	2nd test
% favorable response	Q1	93.8%	93.8%
	Q2	79.2%	81.3%
	Q3	93.8%	87.5%
	Q4	85.4%	89.6%
	Q5	89.6%	95.8%
	Q6	91.7%	95.8%
	Q7	97.9%	100.0%
	Q8	n° 1	0.0%
% response		n° 2	2.1%
		n° 3	6.3%
		n° 4	31.3%
		n° 5	60.4%
			54.2%

Table 2

Internal coherence of test-retest difficulty, variance and item-rest correlation.

	Item difficulty	Item variance	Item-rest correlation	Alpha
q1	0.0625	0.0586	0.657	0.617
q2	0.2083	0.1649	0.3028	0.721
q3	0.0625	0.0586	0.311	0.694
q4	0.1458	0.1246	0.4702	0.655
q5	0.1042	0.0933	0.3812	0.679
q6	0.0833	0.0764	0.4567	0.660
q7	0.0208	0.0204	0.5474	0.670

Table 3

Test-retest reproducibility, with percentage agreement, and kappa and Gwet Ac coefficients.

Item	% agreement	Kappa	Gwet AC
q1	100	1	1
q2	97.9	0.934	0.97
q3	93.8	0.634	0.925
q4	95.8	0.81	0.947
q5	93.8	0.543	0.928
q6	91.7	0.294	0.906
q7	97.9	0	0.979
q8	83.3	0.698	0.796

seen again; the other 8 had unfavorable questionnaire responses but showed no progression since the last classic follow-up and no radiographic abnormality, and were therefore not called back in. The 2 patients called back in presented pain and periprosthetic radiolucency: 1 on a THA stem and 1 on the tibial plateau of a TKA, requiring close follow-up.

Internal coherence analysis found KR-20 = 0.71 on the 7 binary questions. Including question 8, Cronbach alpha was 0.76. The questionnaire thus showed good internal coherence (see Table 2). Item difficulty (i.e., percentage “poor response” requiring renewed contact) was low for all questions. Variance was also low, at a mean 0.08 (range, 0.02–0.16) for the 7 binary questions.

Reproducibility analysis results (percentage agreement, kappa and Gwet Ac coefficients) are shown in Table 3. Percentage agreement on the first 7 (binary) questions was excellent, exceeding 90%; for question 8, it was still good, considering that 5 responses were possible.

#### 4. Discussion

The study hypothesis was confirmed. The translation and cultural adaptation of the self-administered questionnaire for follow-up of THA or TKA recipients bearers was feasible, and checked in a representative sample of the target population. We kept all 8 questions. Correlation coefficients and internal coherence on test-retest were excellent.

The optimal follow-up rhythm after joint arthroplasty is not well defined. The literature on Kingsbury-type questionnaires is still sparse, as they are recent [20]. The methodology of the present translation, on the other hand, is well known; we used Coudeyre et al.'s method [8], ensuring a valid reproducible translation. According to international guidelines, such a translation may be used; however, although we validated the translation and test in a preliminary study on our small sample, application and validation in a larger population will be necessary before widespread use [7,21]. The Pearson coefficient was not used, being unsuited to test-retest analysis. The index Kingsbury study distinguished THA and TKA, which we put together as the method had already been validated and sample size precluded subgroup analysis.

Lovelock et al. [22] performed a “virtual clinic” study on a topic similar to the questionnaire translated here, but did not use a specific follow-up questionnaire but only the Oxford Hip Score or Knee Score, which were not designed for follow-up and need interpreting in order to decide whether a patient requires physical consultation or not; they found that 23% of patients needed to be seen in consultation again. This is why we consider it preferable to use a questionnaire validated specifically for follow-up purposes and identification of patients requiring further consultation. Likewise, a study of telemedicine in Australia was conducted by Caffery et al. [23], but using videoconference after standard questionnaires not specifically designed to identify patients requiring real clinical follow-up.

In the present study, differences emerged between the first and second questionnaire administrations, with more patients contacted after the first. However, only 4 patients actually needed to be seen again clinically, and all were well detected on both questionnaire steps; they will need close follow-up.

One study limitation was the small number of respondents, as the questionnaire was voluntary; there was, however, no significant corresponding difference in patient characteristics. It may be that patients who are not doing well are more likely to respond to a questionnaire or consult with the surgeon or another physician. The 100 selected patients were not contacted in advance. Some may have failed to respond due to an erroneous address or to death not known to our software. This questionnaire, moreover, requires the patient to be free of cognitive impairment, which is difficult to assess except in presence of the patient, and this could be another reason for loss to follow-up. Printing and mailing the questionnaire involved costs that were not assessed; nor were the costs of mailing the responses, which contributed to lowering the response rate. These economic factors deserve specific analysis. An important point with this kind of instrument is that a consultation costs money: the questionnaire uses up some medical time but is not charged as a consultation and some means of payment needs devising. Radiograph + questionnaire follow-up would probably be more effective and reach more patients, as classical follow-up is more onerous; however, this remains to be proven. A final point is that the present response rate was better in males, THA-bearers and rural dwellers among the group of 73, although this was not found in the group of 48, despite there being no significant differences between the two, notably as regards age.

## 5. Conclusion

The French version of the questionnaire enabled reproducible assessment without calling the patient in for consultation unnecessarily. A larger-scale study is required before routine implementation, to analyze usefulness and applicability more closely.

## Disclosure of interest

The authors declare that they have no competing interest. SB is a consultant for Zimmer outside of the present study. MO is a consultant for Arthrex, Stryker and New-clip.

## Funding

None.

## Author contributions

RE was involved in design, investigation, writing and surgery. GV was involved in design, writing and surgery. TC was involved in design, investigation and writing. AM was involved in design and statistical analysis. MO was involved in the statistics, writing and revision. SD was involved in design, revision, surgery and supervision. SB was involved in design, revision, surgery and supervision.

## Acknowledgment

Thanks to Philip Chennell for back-translation into English.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.otsr.2019.01.011>.

## References

- [1] Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. *Lancet* 2007;370:1508–19.
- [2] Maradit Kremers H, Larson DR, Crowson CS, Kremers WK, Washington RE, et al. Prevalence of total hip and knee replacement in the United States. *J Bone Joint Surg Am* 2015;97:1386–97.
- [3] Nemes S, Rolfson O, W-Dahl A, Garellick G, Sundberg M, et al. Historical view and future demand for knee arthroplasty in Sweden. *Acta Orthop* 2015;86:426–31.
- [4] Kingsbury SR, Dube B, Thomas CM, Conaghan PG, Stone MH. Is a questionnaire and radiograph-based follow-up model for patients with primary hip and knee arthroplasty a viable alternative to traditional regular outpatient follow-up clinic? *Bone Joint J* 2016;98:201–8.
- [5] Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- [6] Olsson A, Sandblom G, Fränneby U, Sondén A, Gunnarsson U, Dahlstrand U. The Short-Form Inguinal Pain Questionnaire (sf-IPQ): An Instrument for rating groin pain after inguinal hernia surgery in daily clinical practice. *World J Surg* 2019;43:806–11.
- [7] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186–91.
- [8] Coudeyre E, Descamps S, Intyre JM, Boisgard S, Poiraudou S, Lefevre-Colau MM. Translation and French cultural adaptation of a decision making tool for patients' orientation after total hip or knee arthroplasty. *Ann Phys Rehabil Med* 2009;52:694–703.
- [9] Klouche S, Giesinger JM, Soriali EH. Translation, cross-cultural adaptation and validation of the French version of the Forgotten Joint Score in total hip arthroplasty. *Orthop Traumatol Surg Res* 2018;104:657–61.
- [10] Delaunay C, Epinette J-A, Dawson J, Murray D, Jolles BM. Cross-cultural adaptations of the Oxford-12 HIP score to the French speaking population. *Orthop Traumatol Surg Res* 2009;95:89–99.
- [11] Jenny JY, Diesinger Y. The Oxford Knee Score: compared performance before and after knee replacement. *Orthop Traumatol Surg Res* 2012;98:409–12.
- [12] Jenny JY, Diesinger Y. Validation of a French version of the Oxford knee questionnaire. *Orthop Traumatol Surg Res* 2011;97:267–71.
- [13] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
- [14] Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. *Med Decis Making* 2015;35:114–31.
- [15] Putman S, Rémy F, Pasquier G, Gougeon F, Migaud H, Duhamel A. Validation of a French patient-reported outcome measure for patello-femoral disorders: The Lille Patello-Femoral Score. *Orthop Traumatol Surg Res* 2016;102:1055–9.
- [16] Reed JF. Homogeneity of kappa statistics in multiple samples. *Comput Methods Programs Biomed* 2000;63:43–6.

- [17] Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol* 2008;61:29–48.
- [18] Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977;33:363–74.
- [19] Brooker AF, Bowerman JW, Robinson RA, Riley LH. Ectopic ossification following total hip replacement. Incidence and a method of classification. *J Bone Joint Surg Am* 1973;55:1629–32.
- [20] Hamilton DF, Giesinger JM, Giesinger K. It is merely subjective opinion that patient-reported outcome measures are not objective tools. *Bone Joint Res* 2017;6:665–6.
- [21] Wild D, Grove A, Martin M, Eremenco S, McElroy S, et al. Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005;8:94–104.
- [22] Lovelock T, O'Brien M, Young I, Broughton N. Two and a half years on: data and experiences establishing a "Virtual Clinic" for joint replacement follow up. *ANZ J Surg* 2018, <http://dx.doi.org/10.1111/ans.14752>.
- [23] Caffery LJ, Taylor M, North JB, Smith AC. *J Telemed Telecare. Tele-orthopaedics: A snapshot of services in Australia* 2017;23:835–41.