

Resident self-assessment of operative experience in primary total knee and total hip arthroplasty: Is it accurate?

Tomce Trajkovski, MD
 Christian Veillette, MSc, MD
 David Backstein, MD, MEd
 Veronica M.R. Wadey, MD, MA
 Bill Kraemer, MD

From the Department of Surgery,
 University of Toronto, Toronto, Ont.

Presented at the 2010 Orthopaedic
 Surgery Research Day, University of
 Toronto, Toronto, Ont., Oct. 20, 2010,
 winning the Robin Sullivan Award for
 best research project.

Accepted for publication
 Oct. 3, 2011

Correspondence to:
 T. Trajkovski
 736 Bay St., Ste. 810
 Toronto ON M5G 2M4
 tom.trajkovski@utoronto.ca

DOI: 10.1503/cjs.035510

Background: A prerequisite for a valuable surgical case log is the ability to perform an accurate self-assessment. Studies have shown mixed results when examining residents' ability to self-assess on varying tasks. We sought to examine the correlation between residents' self-assessment and staff surgeons' evaluation of surgical involvement and competence in performing primary total knee (TKA) and hip arthroplasty (THA).

Methods: We used the intraclass correlation coefficient (ICC) to evaluate interobserver agreement between residents' self-perception and staff surgeons' assessment of involvement. An assessment of competency was performed using a categorical global scale and evaluated with the κ statistic. We piloted a structured surgical skills assessment form as an additional objective appraisal of resident involvement.

Results: We analyzed assessment data from 65 primary TKA and THA cases involving 17 residents and 17 staff surgeons (93% response rate). The ICC for resident involvement between residents and staff surgeons was 0.80 (95% confidence interval [CI] 0.69–0.88), which represents substantial agreement. The agreement between residents and staff surgeons about residents' competency to perform the case had a κ value of 0.67 (95% CI 0.50–0.84). The ICC for resident, staff surgeon and third-party observer using the piloted skills assessment form was 0.82 (95% CI 0.75–0.88), which represents substantial agreement.

Conclusion: This study supports the ability of orthopedic residents to perform self-assessments of their degree of involvement and competency in primary TKA and THA. Staff surgeons' assessment of resident involvement correlated highly with the surgical skills assessment form. Self-assessment is a valuable addition to the surgical case log.

Contexte: Un préalable à la tenue d'un registre de cas chirurgicaux valable est la capacité de s'auto-évaluer correctement. Des études ont fait état de résultats mitigés quant à la capacité des résidents de s'auto-évaluer face à diverses tâches. Nous avons analysé la corrélation entre l'auto-évaluation des résidents et leur évaluation par des chirurgiens attachés à l'établissement pour ce qui est de la participation aux chirurgies et des compétences lors d'interventions pour prothèse totale primaire du genou (PTG) et de la hanche (PTH).

Méthodes: Nous avons utilisé un coefficient de corrélation intraclass (CCI) pour évaluer la concordance inter-examineurs entre la perception des résidents et l'évaluation de leur participation par les chirurgiens. Nous avons évalué les compétences à l'aide d'une échelle nominale globale et à l'aide du test κ . Nous avons aussi administré un formulaire structuré d'évaluation des habiletés chirurgicales comme mesure objective additionnelle de la participation des résidents.

Résultats: Nous avons étudié les données d'évaluation relatives à 65 cas de PTG et de PTH primaires, auxquels 17 résidents et 17 chirurgiens ont participé (taux de réponse 93 %). Le CCI afférent à la participation des résidents entre ces derniers et les chirurgiens a été de 0,80 (intervalle de confiance [IC] à 95 % 0,69–0,88), ce qui représente une concordance substantielle. La concordance entre la compétence des résidents évaluée par ces derniers et par les chirurgiens pour ce qui est de la réalisation des interventions s'accompagnait d'une valeur κ de 0,67 (IC à 95 % 0,50–0,84). Le CCI pour les résidents, les chirurgiens et les tiers examineurs à partir du formulaire d'évaluation des habiletés a été de 0,82 (IC à 95 % 0,75–0,88), ce qui représente une concordance substantielle.

Conclusion: Cette étude confirme la capacité des résidents en orthopédie d'évaluer leur degré de participation et leurs compétences lors d'interventions primaires pour PTG et PTH. L'évaluation de la participation des résidents par les chirurgiens a été en forte corrélation avec les résultats du formulaire d'évaluation des habiletés chirurgicales. L'auto-évaluation est un ajout valable au registre des cas chirurgicaux.

An orthopedic surgical trainee is required to balance the acquisition of factual knowledge and the development of psychomotor skills. Although these 2 aspects of an orthopedic training program are interdependent, surgical experience is the essential step in training a surgeon.¹ However, of all the components of surgical training, operative experience is frequently the most poorly documented and least assessed by both trainees and accrediting bodies.

Case logs have been used as a means for assessing surgical exposure and involvement in operative procedures. The opportunity for personal auditing of surgical experience is of definite educational benefit to the trainee and further verifies the adage that "audit seeks to extend the knowledge of practitioners about their own practice."^{2,3} Orthopedic training programs may also benefit from the implementation of a surgical logbook system. The ability to track the progress and identify gaps in experience for individual trainees allows the optimal customization of the educational experience.⁴ Although there are examples of developed electronic logbooks, there is debate about what would constitute a perfect logbook.

Because case logs require a degree of self-assessment and reflection, the objectivity and accuracy of such records have been questioned. Numerous studies have shown mixed results when examining residents' ability to perform self-assessment on varying tasks.⁵ A meta-analysis of 44 self-assessment studies in higher education reported a mean correlation between self- and expert assessments of 0.39.⁶ The performance evaluations used in these studies encompassed multiple domains of clinical competence and

did not focus on surgical residents alone. The purpose of the present study was to examine orthopedic residents' self-assessment of surgical involvement and competence in performing primary total knee (TKA) and total hip arthroplasty (THA). This information is crucial if case logs are to be adopted as a necessary component of surgical training.

METHODS

Participants

Orthopedic residents from the University of Toronto who performed primary TKA and THA over a 9-month span (Oct. 2009 to June 2010) were asked to participate. Data were collected from 5 hospital sites.

Senior residents were defined as those who had previously completed a dedicated arthroplasty rotation during 1 of their first 2 years of training.

Task

Cases of primary TKA and THA were observed by an independent third-party observer (T.T.). A maximum of 4 cases were observed for each resident and staff surgeon to try to minimize fatigue factor and increase accuracy of evaluations. Furthermore, only 1 case was observed in a particular operating room per day. When eligible cases were identified and taking place at different sites on the same day, the case that was chosen to be observed and included in data collection was selected at random by computer to minimize selection bias. Immediately after

Part 1: For each segment of the case, write the percentage that you/resident were involved as a primary surgeon (scalpel/needle driver/component in hand [does not include retracting], includes surgical decision-making steps)

	Percentage (0%–100%)
Exposure	
Implant insertion	
Closure	
Total case performed	

Part 2: In regards to the case performed, do you feel you/resident:

- 1) Require further training
- 2) Are capable of performing the procedure under supervision
- 3) Are capable of performing the procedure independently

Fig. 1. Excerpt from the questionnaire. Part 1: Assessment of involvement; Part 2: Assessment of competency.

completion of the cases, the orthopedic staff surgeon and resident independently completed a paper questionnaire on resident involvement and competency. The amount of resident involvement as the primary surgeon was strictly defined as the percentage of the case the resident operated with the scalpel, needle driver, drill or component in hand (not including retracting), including surgical decision-making steps (Fig. 1). Strict operational definitions were used for compartmentalizing each segment of the case into exposure, implant insertion and closure, and explained to all participants. An assessment of competency was performed using a nonvalidated categorical global scale (Fig. 1). Secondary data collection included self-assessment of complexity of case performed on a 10-point numerical rating scale.

Structured surgical skills assessment form

We piloted a structured surgical skills assessment form as an objective appraisal of resident involvement and compared the perceptions of residents and staff. The skills assessment form was developed to include all major objective and subjective surgical decision-making steps performed during primary TKA and THA (Fig. 2). It was developed in concert with several fellowship-trained arthroplasty surgeons, and surgeons involved in surgical education and the evaluative process of residents. Several drafts were circulated for content validation until consen-

sus was reached for all components. Construct validity was evaluated by comparing the mean involvement of junior versus senior residents using the Student *t* test and found to be significant ($t = 7.3, p < 0.001$), as was done previously for similar studies.⁷ Beside each component was a rating scale from 0 to 2. The observer assigned a “0” if the resident observed the step being performed, a “1” if the resident was actively involved in the component being performed and a “2” if the resident performed the component independently. A rating of N/A was reserved for components not performed owing to staff surgeon preference. An overall percentage was calculated for each subsection and the overall case.

Statistical analysis

Before data collection was started, a sample size calculation was performed. A sample size of 51 cases with 3 observations per participant achieves 80% power to detect an intraclass correlation of 0.2 under the alternative hypothesis when the intraclass correlation under the null hypothesis is 0.4 using an *F* test with a significance level of 0.05. Data were analyzed using SAS version 9.2 (SAS Institute Inc.) We used the intraclass correlation coefficient (ICC) to evaluate interobserver agreement between residents’ and staff surgeons’ assessment of involvement. Intraclass correlation coefficient values can be interpreted as weighted κ values where the weights are quadratic. Values range from 0, reflecting no

	0	1	2	N/A
Approach				
Marks skin incision				
Makes skin incision				
Subcutaneous dissection to capsule				
Capsulotomy and deep dissection				
Releases, soft tissue/osteophytes				
Identifies need for selective release				
Femur				
Identification of femoral start point				
Drilling femoral canal				
Assessment of femoral alignment, degrees of valgus cut				
Distal femoral cut				
Identifies and releases ACL only (CR), ACL/PCL retaining knee (PS)				
Size femur in AP plain				
Assessment of rotation, degrees of external rotation				
Anterior femoral cut				
Posterior femoral cut				
Chamfer cuts				
Box cut/hotch cut				
Removal of osteophytes				

Fig. 2. Excerpt from the total knee arthroplasty assessment form. 0 = the resident observed the step being performed; 1 = the resident was actively involved in the component being performed; 2 = the resident performed the component independently; N/A = component not performed owing to staff surgeon preference.

ACL = anterior cruciate ligament; AP = anteroposterior; CR = cruciate retaining; PCL = posterior cruciate ligament; PS = posterior stabilized.

interobserver agreement, to 1, reflecting perfect interobserver agreement. Interobserver agreement for competency was evaluated with the κ statistic, as categorical variables were used. We used the standard and well-established criteria of Landis and Koch for the interpretation of agreement studies.⁸ They define poor agreement as 0–0.2, fair agreement as 0.21–0.40, moderate agreement as 0.41–0.60, substantial agreement as 0.61–0.80 and nearly perfect agreement as 0.81 or greater. Moreover, other investigators who have validated scales have used values of 0.65 or greater to represent substantial agreement.⁹

RESULTS

Seventeen residents and 17 staff surgeons participated in the study (93% response rate). The breakdown was 7 junior residents (5 from postgraduate year [PGY]-1, 2 from PGY-2) and 10 senior residents (5 from PGY-3, 3 from PGY-4 and 2 from PGY-5).

The ICC between staff surgeons' and residents' assessment of surgical involvement varied from 0.64 to 0.84, depending on the segment of the case being analyzed (Table 1). The ICC for the involvement of the total case was 0.80 (95% confidence interval [CI] 0.69–0.88). When the structured surgical skills assessment form was introduced and compared with staff surgeons' and residents' assessments, the correlation was 0.82 (95% CI 0.75–0.88). This illustrated substantial agreement of staff surgeons' and residents' assessment of resident involvement with results of the assessment form. Residents were as good as staff physicians at assessing involvement in comparison with the assessment form. The correlation between staff surgeons' and residents' assessment of competency had a κ value of 0.67 (95% CI 0.50–0.84), which was also in the substantial agreement category. Interestingly, when there was disagreement between resident and staff surgeon on assessment of competency, it was the staff surgeon who usually recorded a higher level of competency.

Using the surgical skills assessment form, we found a significant difference in resident involvement based on level of training. Junior residents were acting as primary surgeons for only 28% of the cases on average, whereas senior residents were acting as primary surgeons 66% of the time ($p < 0.001$). This attests to the construct validity of

the assessment tool piloted. This also provides valuable information suggesting, as one would expect, that as residents progress through their training they are more active participants in surgical cases.

In regard to the level of complexity of the cases performed, there was a statistically significant difference between resident and staff assessment ($p < 0.001$). Residents on average tended to underestimate the level of complexity by an average of 1.64 (95% CI 1.37–1.94) points on a 10-point scale ($p < 0.001$).

DISCUSSION

This study aimed to verify the accuracy of self-assessment for surgical involvement and competency in a group of orthopedic surgery residents performing primary TKA and THA. Participants were accurate judges of their surgical involvement with substantial agreement with staff surgeon assessment and a structured surgical skills assessment form. They were also able to accurately self-assess their level of competency in comparison to staff surgeon opinion ($\kappa = 0.67$, 95% CI 0.50–0.84). However, residents on average tended to underestimate the level of complexity of cases in comparison to staff surgeon opinion.

The finding that residents underestimate the complexity of primary arthroplasty cases in comparison to staff surgeons is of utmost importance. This may represent a gap in the understanding of principles of arthroplasty or simply a lack of experience. This finding illustrates the importance of acquisition of knowledge, surgical experience and technical skills as a true reflection of competency.

Obtaining an accurate appraisal of resident involvement requires using a measurement tool that represents a true picture of the operative experience. In surgical education studies, staff surgeons have often been used as the expert opinion and gold standard when assessing residents. In our study, staff surgeons were able to assess resident involvement with a high degree of reliability in comparison to an objective surgical assessment form. Moreover, assessment forms, such as the one used in the present study, have previously been shown to have a high interrater reliability.⁷ These findings indicate that staff surgeons are able to provide a “true” account of the operative experience.

There are several possible explanations that may

Table 1. Intraclass correlation coefficient values for percentage of surgical involvement

Surgical involvement	Comparison; intraclass correlation coefficient			
	Resident and staff	Staff and assessment form	Resident and assessment form	Resident, staff and assessment form
Exposure	0.64 (0.45–0.78)	0.64 (0.45–0.78)	0.63 (0.43–0.77)	0.64 (0.50–0.76)
Implants	0.84 (0.75–0.90)	0.88 (0.81–0.93)	0.86 (0.77–0.91)	0.86 (0.80–0.91)
Closure	0.80 (0.69–0.87)	0.58 (0.40–0.72)	0.68 (0.52–0.79)	0.69 (0.58–0.79)
Total	0.80 (0.69–0.88)	0.85 (0.76–0.90)	0.82 (0.71–0.88)	0.82 (0.75–0.88)
Agreement	Substantial	Substantial	Substantial	Substantial

account for the better-than-expected self-assessment scores. These findings may reflect self-assessment abilities for the performance of a compartmentalized technical task, which may allow for a more objective self-evaluation. Additionally, continuous observation and informal feedback may occur more often for surgical skills than for other clinical skills. Furthermore, exposure to a large volume of operative arthroplasty cases may enable surgical residents to readily identify their level of involvement and competency. These results are not likely applicable to all orthopedic surgeries owing to varying degrees of complexity and residents' limited exposure to uncommon procedures.

An accurate assessment of involvement in operative cases is valuable information not only to the trainee, but also to the training program. It serves as an important mode of feedback for residents and allows them to chart their involvement in cases over time. The information could also allow educators to identify residents who are not obtaining adequate operative experience and may require further training or remediation before the end of the surgical rotation. Furthermore, it may help to identify hospitals where residents are actively involved in operative cases, which, in turn, may be more ideal settings for training purposes.

Limitations

There are limitations to this study. Only a subset of residents were eligible for participation in the study, as only 5 hospital sites were used as data collection sites. Furthermore, some residents were not performing arthroplasty during the data collection period. The accuracy of resident self-assessment may have been biased by the knowledge that staff surgeons were rating the residents. Residents may also have altered their responses because they were being evaluated by an objective observer. However, we do not feel these biases undermine the importance of our results. Being evaluated likely stimulated focused self-reflection throughout the cases, which may have improved residents' ability to self-assess. Our results attest to the ability of residents to perform accurate self-assessments immediately after a focused technical task while being evaluated by a staff surgeon.

In the current study, a single categorical global scale was used to assess competency. This by no means is a comprehensive assessment tool. A single question is not enough to assess overall competency to perform primary TKA and THA, as competency encompasses multiple facets including knowledge acquisition, surgical acumen and technical skills. However, we feel it provides a quick, feasible and valuable assessment of residents' performance on the cases performed. Simply because a resident feels competent to perform 1 case does not mean they have reached a level of

competency at which they can perform all cases of primary arthroplasty, as no 2 cases are alike. If data on a multitude of cases were collected as part of the surgical case log, one could better monitor a global measure of competency over time.

CONCLUSION

Our results demonstrate that orthopedic trainees have the potential to provide accurate assessments of their degree of involvement and competency in primary TKA and THA. Future studies are needed to address whether our results can be expanded to a multitude of orthopedic procedures. However, it is our opinion that inclusion of a self-assessment is a valuable addition to the surgical case log for primary TKA and THA.

Competing interests: None declared for T. Trajkovski, C. Veillette, V.M.R. Wadey and B. Kraemer. D. Backstein declares having been a paid educational consultant to Zimmer Canada and Stryker Canada.

Contributors: All authors designed the study, reviewed the article and approved its publication. T. Trajkovski, C. Veillette and V.M.R. Wadey acquired the data, which T. Trajkovski and B. Kraemer analyzed.

References

1. Sambrook P, Smith A, Hewitt G, et al. Development of a national clinical logbook system for oral and maxillofacial surgery trainees — the RACDS, OMS system. *Int J Oral Maxillofac Surg* 1996;25:242-4.
2. Cool P, Morgan-Jones RL, Oliver CW, et al. The audit of orthopaedic trainee teaching improves supervision rates. *Ann R Coll Surg Engl* 1999;81:195-7.
3. Beard JD, Brennan JA, Budd JS. A computerised personal audit for surgical trainees. *Ann R Coll Surg Engl* 1990;72(4 Suppl):46-8.
4. Chou S, Lockyer J, Cole G, et al. Assessing postgraduate trainees in Canada: Are we achieving diversity in methods? *Med Teach* 2009;31:e58-63.
5. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 1991;66:762-9.
6. Falchikov N, Boud D. Student self-assessment in higher education: a meta-analysis. *Rev Educ Res* 1989;59:395-430.
7. Winckel CP, Reznick RK, Cohen R, et al. Reliability and construct validity of a structured technical skills assessment form. *Am J Surg* 1994;167:423-27.
8. Sackett DL, Haynes RB, Guyatt GH, et al. *Clinical epidemiology: a basic science for clinical medicine*. 2nd ed. Boston (MA): Little, Brown; 1991. p. 30.
9. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991;44:1271-8.