



## Original research

## Is there more than one approach to evaluating the variability of surgeons' performance?

Mikhail Saltychev <sup>a,\*</sup>, Katri Laimi <sup>a</sup>, Petri Virolainen <sup>b</sup>, Merja Eskola <sup>c</sup>, Ville Äärimala <sup>b</sup><sup>a</sup> Department of Physical and Rehabilitation Medicine, Turku University Hospital and University of Turku, Turku, Finland<sup>b</sup> Department of Orthopedics and Traumatology, Turku University Hospital and University of Turku, Turku, Finland<sup>c</sup> Department of Rehabilitation and Brain Trauma and Department of Expert Services, Turku University Hospital and University of Turku, Turku, Finland

## H I G H L I G H T S

- There were statistically significant differences between seven surgeons' performances of shoulder surgery.
- However, only a non-significant part of this variability could be interpolated across the entire shoulder surgery field.
- Variability in performance within a group of surgeons cannot be generalized without additional statistical analyses.

## A R T I C L E I N F O

## Article history:

Received 30 September 2014

Received in revised form

9 February 2015

Accepted 12 February 2015

Available online 18 February 2015

## Keywords:

Performance

Intraclass correlation coefficient

Variability

Variance

## A B S T R A C T

**Introduction:** To demonstrate that the variability found to be significant between surgeons' performances within a small group does not necessarily mean that this significance applies to the entire field of that specific type of surgery. It is common for inferences and recommendations for an entire field to be based on the variability within a small group of surgeons. The variability between groups usually remains unknown.

**Methods:** An analysis of variance was used to assess the statistical significance of the variability among surgeons' performances of a specific type of surgery within the studied sample. The intraclass correlation coefficient was used to investigate how large a segment of this variability can be explained by a surgeon-related factor for the entire surgeon population of a specific field. The topic was illustrated using data obtained from a group of seven surgeons who operated on the penetrating rotator cuff tears of 742 patients.

**Results:** There were statistically significant differences between seven surgeons in the improvement of pain and the range of shoulder joint motion. However, only a small ( $\leq 2\%$ ) and statistically non-significant part of this variability could be explained by a difference between surgeons when the results were interpolated across the entire population of shoulder surgeons.

**Discussion and conclusion:** Variability in performance within a group of surgeons performing a specific type of surgery cannot be generalized to include the performance of all surgeons doing the same type of surgery without additional statistical analyses.

© 2015 Surgical Associates Ltd. Published by Elsevier Ltd. All rights reserved.

## 1. Introduction

Even when a specific type of treatment is implemented in essentially similar settings, its effects vary [1]. Numerous factors may be responsible. Differences in the severity of the condition between patients, co-morbidity, age, gender, general health,

occupation, health behaviour, and motivation for post-treatment rehabilitation are among such factors. How much of this variability can be explained by the influence of the actions of a single expert?

A comparison of the performance of healthcare actors is especially challenging, as the lasting and multistage treatment process usually has outcomes that are not easily measurable and that manifest themselves years after the treatment. In addition, the process, which is often multidisciplinary, does not make it possible to easily distinguish the impact inflicted by a particular expert. In this respect, surgical procedures create a special situation, as the

\* Corresponding author. Department of Physical and Rehabilitation Medicine, Turku University Hospital, PO Box 28, 20701 Turku, Finland.

E-mail address: [mikhail.saltychev@gmail.com](mailto:mikhail.saltychev@gmail.com) (M. Saltychev).

success or failure of surgical treatment is commonly perceived by patients and health professionals as being tightly linked to the surgeon who performed the procedure. Therefore, the performance of operating physicians has probably been assessed more often than that of any other medical specialties. It has been suggested that different surgeons obtain different results and that this so-called 'surgeon-related' factor causes substantial outcome variability [1–4]. Differences in surgeons' performances have been reported for different disorders and subspecialties and explained, for example, by the surgeon's experience, training level, and annual frequency of performed operations [1–12].

Previous reports do not contain evidence that any of the surgeon-related factors found to affect the results of procedures within each studied sample are responsible for any substantial portion of the overall fluctuation in the results across the field. The purpose of our study was to demonstrate that the variability in surgeons' performances, observed within a small study sample, does not permit inferences regarding the entire field. In this study, point was first illustrated by assessing the statistical significance of variance for a certain group of surgeons and then determining the portion of the overall variability of performance that can be explained by a found surgeon-related factor. We used our data on surgical procedures performed by a group of seven orthopaedists who had operated on penetrating tears of rotator cuff tendons. The purpose was not to evaluate the effectiveness of a specific procedure but, instead, to demonstrate the appropriate use of some statistical techniques.

## 2. Materials and methods

### 2.1. Settings and study population

The data for this demonstration were obtained from an on-going prospective study on shoulder surgery described in more detail elsewhere [4,13,14]. Seven consultants in orthopaedics who had operated on consecutive adult patients with a tear of a rotator cuff between June 2007 and December 2013 in the Department of Orthopaedics and Traumatology at a university hospital. To minimize possible outliers, the data on surgeons who operated on fewer than 20 patients during the period of enrolment were excluded from the analysis. The office nurse electronically booked an appointment for an orthopaedist based on the next available outpatient appointment time. The orthopaedic surgeon who had examined the patient in the outpatient clinic was the one who also performed the surgery. Data were collected from electronic patient records, from findings obtained during the clinical examinations by the orthopaedic surgeon and a physiotherapist, and from a structured patient survey. The surgeon also recorded data during the operation using a structured electronic form. The surgical techniques and operating equipment were similar for all of the orthopaedists. The outcome was assessed 12 months after the surgery. The Ethics Committee of the Turku University Hospital approved the study.

### 2.2. Definitions of the descriptive variables

The main *diagnosis* was made by an orthopaedist using the code of the International Classification of Diseases, 10th revision. *Age* was calculated in full years at the time of the surgery. The *time between the clinical examination and the surgery* was recorded in days. The patient's *body mass index (BMI)* was calculated during the clinical examination. The *cause of the tear* was dichotomized as traumatic or non-traumatic. The *duration of symptoms* was dichotomized as  $\leq 3$  months (acute) or  $> 3$  months (chronic). *Smoking* was dichotomized as yes or no. The *work situation* was dichotomized as working or studying versus retired, on sick leave, or unemployed. *Previous*

*surgery* on a rotator cuff was dichotomized as yes or no. The *presence of osteoarthritis* was confirmed by a surgeon during the operation. Observed osteoarthritis of any grade on either one of joint surfaces was dichotomized as yes or no. The *sagittal size of the penetrating tear* in the rotator cuff tendon was measured in millimetres by the surgeon during the operation.

### 2.3. Definitions of outcome variables

*Pain severity* was assessed using the numeric rating scale 0 to 10 (0 representing 'no pain' and 10 being 'the worst possible pain'). The range of arm motion was recorded with a precision of  $10^\circ$ . The results were also measured by the standardized *Constant Shoulder Score* [15,16]. *Satisfaction with the surgery* was dichotomized as yes or no at 12 months after the procedure.

### 2.4. Theory

It is common in previous reports that, when a difference in performance within a certain group of surgeons is found to be statistically significant, inferences and recommendations are made regarding the entire field of that particular type of surgery. Are such inferences justified? After all, what does the 'statistical significance' of the variability in the performance reported by previous studies mean? In this case, statistical significance may denote two diverse points of view. Firstly, patients, the supervisor, the hospital administration, or the surgeons themselves may be interested to know how uniform the results of a *certain* group of surgeons (e.g. who work in the same department) are, and what the 'ranking' of surgeons regarding outcome is. Secondly, policy makers and researchers may be concerned with the extent to which a surgeon-specific factor may explain the overall variability in performance of *any* possible surgeon (or any group of surgeons) across entire field. These two very different questions require different statistical approaches. The tests for significance used in previous studies have been limited to the first approach. Statistically significant p-values or confidence intervals signal that differences in outcome endure even though the *same surgeons* perform a procedure an infinite number of times on an infinite sample of patients. The observed significance is, however, only true for the studied, usually small, group of surgeons operating in specific settings. It says nothing about the variability in performance outside the studied group. Previous reports do not contain evidence that any of the surgeon-related factors found to affect the results of procedures within each studied sample may be responsible for any substantial portion of the overall fluctuation in the results across the field. This point is important since imprecise statistical inferences supported by misunderstood or misused methods may lead to incorrect guidelines and recommendations.

### 2.5. Calculation

When appropriate, the descriptive and outcome variables were assessed by calculating frequencies, means, and standard deviations (SD). Changes in the outcome variables during follow-up were calculated as means and SDs. We used a one-way analysis of variance (ANOVA) to test the significance of the differences between the surgeons' results with respect to changes in each continuous outcome variable. To ensure ANOVA adjustment for the presence of osteoarthritis [14] and the size of rotator cuff tear [13], we carried the analysis out twice using both GLM (general linear modelling) and MIXED (mixed linear models) procedures. A chi-square test was used for the dichotomized variable (satisfaction with surgery). The one-way, random intraclass correlation coefficient (ICC) method with class set to surgeon identification number

Download English Version:

<https://daneshyari.com/en/article/4285991>

Download Persian Version:

<https://daneshyari.com/article/4285991>

[Daneshyari.com](https://daneshyari.com)