The Journal of Arthroplasty 31 (2016) 81–86

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

Analysis of Longitudinal Outcome Data with Missing Values in Total Knee Arthroplasty

Yeon Gwi Kang, MS [a], Jang Taek Lee, PhD [b], Jong Yeal Kang, MD [a], Ga Hye Kim, BA [c], Tae Kyun Kim, MD, PhD [a]

[a] Joint Reconstruction Center, Seoul National University Bundang Hospital, Seongnam-si, South Korea
[b] Department of Statistics, Dan Kook University, Yongin-si, South Korea
[c] Department of Psychology, Princeton University, Princeton, New Jersey

ARTICLE INFO

Article history:
Received 10 August 2014
Accepted 29 June 2015

Keywords:
missing data
total knee arthroplasty
repeated measures ANOVA
generalized estimating equation
mixed model repeated measures

ABSTRACT

We sought to determine the influence of missing data on the statistical results, and to determine which statistical method is most appropriate for the analysis of longitudinal outcome data of TKA with missing values among repeated measures ANOVA, generalized estimating equation (GEE) and mixed effects model repeated measures (MMRM). Data sets with missing values were generated with different proportion of missing data, sample size and missing-data generation mechanism. Each data set was analyzed with three statistical methods. The influence of missing data was greater with higher proportion of missing data and smaller sample size. MMRM tended to show least changes in the statistics. When missing values were generated by 'missing not at random' mechanism, no statistical methods could fully avoid deviations in the results.

In many clinical studies, the data used for analysis are collected with repeated measurements within same subjects, followed over a period of time. This kind of data is called longitudinal data. Since functional recovery after total knee arthroplasty (TKA) occurs over a long period, the data collected for evaluation of post-TKA outcome are longitudinal in their characteristics. However, it is inevitable that these data come with missing values because not all patients can visit the clinic at every time point of evaluation. Therefore, proper statistical methods should be applied to deal with the influence of missing values.

Several statistical methods have been introduced for the analysis of longitudinal data with missing values. To draw a valid conclusion in clinical studies, it is imperative to understand the strengths and shortcomings of each method and to apply relevant ones for the given data. Traditional method for handling longitudinal data with missing values is repeated measures analysis of variance (RM-ANOVA). With this method, only the subjects with complete data are included, and those with any missing values are excluded from data analysis. This may cause loss of information and deviation in the results. Recently, generalized estimating equation (GEE) and mixed effects model repeated measures (MMRM) have gained popularity. GEE is a method relying on statistical models that take into account estimated correlations between successive data points, based on a pre-determined structure and the data itself [1]. MMRM relies on statistical models that contain both (or mixed — that is why its name includes 'mixed') fixed and random effects to account for missing values [2]. These methods have been increasingly employed because they can utilize all the observed data even in the subjects with missing values, and therefore minimize the influence of missing data.

The influence of missing data on the study results would differ according to many factors like the proportion of missing data, sample size, type of data and mechanism of missing data generation. The influence would be greater with increasing proportion of missing data and smaller sample size. Certain type of data may be more vulnerable to the effect of missing values. If the missing values are generated not at random but in particular pattern, the results may be substantially deviated. For example, if dissatisfied patients do not return for routine follow-ups, the observed outcomes would look better than the reality. Therefore, to minimize the deviations in study results, it is very important to recognize the impact of the missing data and choose appropriate statistical methods for the analysis. However, in the field of orthopedic surgery, there are few studies which focus on how the missing data affect the results and which statistical method is appropriate for the analysis of those data.

We conducted this study to determine how the missing values in outcome data of TKA affect the results in various conditions with different proportions of missing data, sample sizes, types of data and missing-

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to http://dx.doi.org/10.1016/j.arth.2015.06.067.

Each author certifies that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

This study was performed at Seoul National University Bundang Hospital, Bundang-gu, Gyeonggi-do, Korea.

Reprint requests: Tae Kyun Kim, MD, PhD, Joint Reconstruction Center, Seoul National University Bundang Hospital, 300 Gumi-dong, Bundang-gu, Seongnam-si, Gyeonggi-do 463-707, South Korea.

http://dx.doi.org/10.1016/j.arth.2015.06.067
0883-5403/

data generation mechanisms. We applied three different statistical methods, RM-ANOVA, GEE and MMRM, for the analysis and tried to determine which one is the most appropriate for the given data. We hypothesized that the deviations in the result would be more prominent with higher proportion of missing data, smaller sample size and when the missing values are not generated randomly. We also thought that the influence of missing values would differ between different types of data. Finally, we hypothesized that GEE or MMRM is more appropriate than RM-ANOVA for the analysis of outcome data of TKA with missing values.

Materials and Methods

We retrospectively reviewed the medical records of the patients who were diagnosed as primary osteoarthritis and received primary TKA in our institution from June 2004 to November 2005. This study was approved by the institutional review board of our hospital. Of the 291 patients who received TKA, 100 patients with complete longitudinal outcome data were included for the analysis. Of the 100 patients, 97 were female, and three were male. Mean (standard deviation) age, body weight, height and BMI were 68.4 (5.6) years, 61.3 (9.3) kg, 152.2 (5.8) cm, and 26.5 (3.5) kg/m^2.

Range of motion (ROM) of the knee and WOMAC score, which are commonly used outcome measures after TKA, were measured in the patients who underwent TKA in our institution. The measurements were taken at four different time points: before surgery, 6 months, 1 year and 2 years after surgery. ROM of all patients was measured by single investigator (YGK) to the nearest 5° using a standard clinical goniometer with the patients supine. WOMAC score, a clinical index designed to evaluate the status or treatment outcomes of the osteoarthritic patients, is measured by the patients using the questionnaire which consists of 24 items [3]. It consists of three subscales to measure pain, stiffness and function, which has 5, 2 and 17 items, respectively. The score of each item ranges from 0 (no symptom) to 4 (worst symptom) so that the highest possible score is 96 points overall.

Complete data set from 100 patients without missing values was analyzed by three statistical methods (i.e. RM-ANOVA, GEE and MMRM), and the results were set as reference values for subsequent analysis. Means and standard errors of means were calculated at each time point, and *P*-values for difference of means between two successive measurements were obtained. Subsequently, mimicking actual clinical circumstances, missing values were generated at different postoperative evaluation time points (6 months, 1 year, and 2 years after surgery). Missing-data generation mechanism can be divided into 'missing completely at random (MCAR),' 'missing at random (MAR)' and 'missing not at random (MNAR)' [4,5]. All missing values in our study were generated in MCAR mechanism using a randomization table except for the data sets used for the analysis of missing-data generation mechanisms, in which case, the missing values were created with MNAR mechanism. These new data sets with missing values were then analyzed using the three statistical methods. Among these three statistical methods, the one which gave the statistics whose values were closest to the reference values was considered most useful.

To address each question, we generated data sets under four different situations. First, to determine whether proportion of missing data affects the result, missing values were generated to account for 10%, 20% and 30% of the ROM data at each postoperative evaluation time point. Each data set was analyzed using the three statistical methods, and the results were compared with the reference values. Second, to determine the effect of sample size on the result, we randomly selected 50 out of the 100 cases for which missing values were generated as 10%, 20% and 30% of the ROM data at each postoperative evaluation time point. Each data set was analyzed using the three statistical methods, and the results were compared with those obtained from 100 cases. Third, we generated missing values in WOMAC score data with the same method as in ROM data. Then the statistics obtained from the ROM data and WOMAC score data were compared to determine whether the results differ according to type of data. Lastly, we generated data sets with missing values under MNAR mechanism. We made two different assumptions. First, assuming that the patients with poor outcomes did not return to clinic for the aforementioned reason, missing values were generated in the patients with poor outcome (worst scenario). Data were erased serially from the worst outcome (i.e., lowest ROM) to make the proportion of missing data 10%, 20% and 30%. Second, assuming the opposite situation, missing values were generated in the patients with favorable outcome (best scenario). Data were erased serially from the best outcome. The missing values were generated at the proportion of 10%, 20% and 30% in the data of 6 months, 1 year and 2 years after surgery. The results of these data sets were then compared with the reference values. All statistical analyses were performed with SPSS for Windows (version 17.0, SPSS, Chicago).

Results

In the analysis of the 100 cases with randomly generated missing values in ROM data, mean values did not change substantially even though the proportion of missing data increased. However, standard errors and *P*-values markedly increased with increasing proportion of

Table 1
The Statistics and Their Changes From Reference Values According to the Proportion of Missing Data Obtained by Using Three Statistical Methods for Analyzing the Range of Motion Data From 100 Patients.

Method	Proportion of Missing Data[a]	Preoperative		6 Months		1 Year		2 Years		*P*-Value (Pre-6 Months)	*P*-Value (6 Months–1 Year)	*P*-Value (1 Year–2 Years)
		Mean	SE	Mean	SE	Mean	SE	Mean	SE			
RM-ANOVA	0% (100)	142.8	1.3	132.7	1.0	134.4	0.9	132.3	1.1	0.000	0.046	0.001
	10% (72)	143.5 (+1)	1.4 (+8)	132.8 (+0)	1.1 (+10)	134.7 (+0)	1.1 (+22)	132.3 (0)	1.2 (+9)	0.000	0.146	0.002
	20% (50)	144.4 (+1)	1.5 (+15)	133.1 (+0)	1.4 (+40)	135.3 (+1)	1.3 (+44)	133.8 (+1)	1.4 (+27)	0.000	0.164	0.264
	30% (34)	143.4 (+0)	2.3 (+77)	132.9 (+0)	1.8 (+80)	134.7 (+0)	1.9 (+111)	132.2 (−0)	1.9 (+73)	0.000	0.601	0.086
GEE	0% (100)	142.8	1.3	132.7	1.0	134.4	0.9	132.3	1.1	0.000	0.038	0.001
	10% (90)	142.8	1.3	132.2 (−1)	1.0 (0)	134.7 (+0)	1.0 (+11)	132.3 (0)	1.1 (0)	0.000	0.023	0.010
	20% (80)	142.8	1.3	132.2 (−1)	1.1 (+10)	134.6 (+0)	1.1 (+22)	133.0 (+1)	1.1 (0)	0.000	0.039	0.379
	30% (70)	142.8	1.3	133.6 (+1)	1.1 (+10)	135.0 (+0)	1.1 (+22)	131.6 (−1)	1.3 (+18)	0.000	0.605	0.007
MMRM	0% (100)	142.8	1.3	132.7	1.0	134.4	0.9	132.3	1.1	0.000	0.046	0.001
	10% (90)	142.8	1.3	132.4 (−0)	1.0 (0)	134.6 (+0)	0.9 (0)	132.2 (−0)	1.1 (0)	0.000	0.025	0.002
	20% (80)	142.8	1.3	132.4 (−0)	1.1 (+10)	134.4 (+0)	1.0 (+11)	132.5 (+0)	1.1 (0)	0.000	0.072	0.025
	30% (70)	142.8	1.3	133.4 (+1)	1.1 (+10)	134.9 (+0)	1.0 (+11)	132.2 (−0)	1.1 (0)	0.000	0.392	0.004

The changed statistics were presented with the amount of change (%) in the parenthesis, with references to the values of the original data without missing values (i.e. 0% missing data): increment, +; decrement, −.

[a] Proportion of data which were actually included in the analyses is presented in the parenthesis.