Original Article Clinics in Orthopedic Surgery 2021;13:549-557 • https://doi.org/10.4055/cios19152

Reliability of Goldberg Scoring System in the Radiographic Evaluation of Bony Union after Bone Grafting

Young Choi, MD, Young Hoon Yang, MD, Young-Ho Kwon, MD

Department of Orthopaedic Surgery, Kosin University College of Medicine, Busan, Korea

Background: Evaluation of bony union after bone grafting is very important in orthopedic surgery. The aim of this study was to verify inter- and intraobserver reliability of the Goldberg scoring system for radiographic evaluation of bony union after bone grafting in various situations of animal models.

Methods: Twenty-seven male C57/BL6 mice, which lack the ability to synthesize galactose-alpha-1,3-galactose (GalT KO mice), and 9 C57/BL6 mice carrying a wild-type gene were used as animal models. We divided the mice into four groups. In group 1, syngenic bone grafting and intramedullary fixation were performed (9 wild type C57BL/6 mice). In group 2, allogenic bone grafting was performed (9 GalT KO mice). In group 3, an alpha-galactosidase-treated porcine xenograft was transplanted into the femur to reduce the antigenicity (9 GalT KO mice). In group 4, a non-treated porcine xenobone grafting was performed (9 GalT KO mice). The level of radiographic bony union (Goldberg method) was assessed by three orthopedic surgeons. Intra- and interobserver reliability for radiographic evaluation was assessed.

Results: In the Goldberg scoring system, most of the radiographic measurements showed substantial to almost perfect intraobserver reliability. The total score showed substantial intraobserver reliability. The kappa coefficient (κ) of the first examiner was 0.603, the κ of the second examiner was 0.790, and the κ of the third examiner was 0.758. The scoring system showed substantial interobserver reliability. The κ of the first session was 0.641 and the κ of the second session was 0.649.

Conclusions: The Goldberg scoring system is a reliable tool for radiographic evaluation of bony union after bone grafting.

Keywords: *Goldberg scoring system, Bone grafting, Bony union*

Bony union after bone grafting is a very important criterion in outcome assessment to determine the healing status, effectiveness of treatment, and follow-up care plan.

Bone grafting is used in patients with congenital abnormalities such as pseudoarthrosis, delayed union/nonunion of fractures, osseous defects caused by tumors, trauma, and infection.[1)] Autologous graft is the best bone graft option, causing less problem of transplant rejection.[1,2)] However, autologous grafts are limited in their amount and occasionally additional surgery is required for postoperative complications, such as pain, donor site fracture, or meralgia paresthetica.[2)] Therefore, allogenic bone grafts have emerged as a substitute source of bone grafts. Although allografts solve the problem of donor site morbidity, they have inherent disadvantages because of limited graft availability, cost, the potential for disease transmission, and in many clinical settings, slower and less complete graft incorporation than autograft tissue.[3)] Despite these limitations, however, the interest has increased in recent years for the use of allografts in reconstructive orthopedic surgery. For these reasons, substitutes including biosynthetic materials or demineralized bone matrix have

Received November 5, 2019; Revised June 22, 2020;
Accepted January 12, 2021

Correspondence to: Young-Ho Kwon, MD
Department of Orthopaedic Surgery, Kosin University College of Medicine,
262 Gamcheon-ro, Seo-gu, Busan 49267, Korea
Tel: +82-51-990-6467, Fax: +82-51-243-0181
E-mail: ks880087@kosinmed.or.kr

Clinics in Orthopedic Surgery • pISSN 2005-291X eISSN 2005-4408

been studied for clinical purposes. Moreover, many studies on xenografts have been performed.[4-7] The problems of allografts, such as the transmission of disease, antigenicity, accessibility, and psychological distance, have led to the interest in xenografts as a graft material.

Materials from inorganic bovine can be used for onlay graft because of superior integration with newly formed bone. As bone grafting increasingly occupies an important and broad place in orthopedic surgery, further investigations and studies on the evaluation methods of bony union are necessary.

Simple radiographs are commonly used to evaluate bony union after bone grafting. The Goldberg scoring system[8] is widely used for radiographic assessment of bony union after bone grafting, which is considered the radiographic evaluation, as well as the histologic evaluation. Although the Goldberg scoring system is advantageous for evaluation of bony union based on radiographic findings, it has disadvantages such as lack of reproducibility. Therefore, it is important to confirm reliability of the Goldberg scoring system for radiographic assessment.

The Kappa coefficient (κ) is used to measure the reliability of ratings or measurements in statistics. It represents how strongly units in the same group are similar to each other. Although it is thought as a type of correlation coefficient, it is different from most other correlation coefficients in that it is based on data structured as groups, instead of data structured as coupled observations.[9]

The κ is typically used to quantify the extent to which individuals with a fixed degree of relevance resemble each other in terms of categorical characteristics. Another common usage is the evaluation of the consecutiveness or reproducibility of qualitative measurements observed by different individuals estimating the same quality.[9]

There are two types of changeability, interobserver variability and intraobserver variability. The former (interobserver variability) represents systematic differences between observers—one observer could persistently estimate patients at a risky level compared to other observers. The latter (intraobserver variability) represents the divergence of a certain observer's grade on a certain patient, which is not part of a systematic gap.

The aim of this study was to verify inter- and intraobserver reliability of the Goldberg scoring system for radiographic evaluation of bony union after bone grafting in various situations in animal models.

METHODS

This study was approved by the Institutional Animal Care and Use Committee and conducted from May 2010 to December 2010 (No. BA1108-088/047-01).

Materials

Animals

This study included 27 male alpha-gal knockout C57/BL6 mice (9–10 weeks old, 23–27 g in weight) and 9 wild type C57/BL6 mice. The animals were nurtured in cages with a temperature of 23°C ± 1°C and humidity of 60% ± 5%. A day and night cycle of 12 hours was provided to maintain the biological rhythm, and formula feeding with distilled water was provided ad libitum. The animals were observed closely every day for any pain and distress.

This study was set and performed in various conditions of bony union after bone grafting. The mice were divided into four groups: (1) group 1 included C57/BL6 wild type mice that received a syngenic bone graft from wild type mice; (2) group 2 included C57/BL6 alpha-gal knockout mice that received a bone graft from C57/BL6 wild type mice; (3) group 3 contained C57/BL6 GalT KO mice that underwent untreated porcine xenogenic bone grafting; and (4) group 4 had C57/BL6 GalT KO mice that underwent alpha-galactosidase-treated porcine xenogenic bone grafting.

Bone grafts and enzyme preparation

The synergic bone grafts of mice were set for autografts and the allogenic bone grafts of mice were set for allograft. And porcine bone grafts were set for xenograft. A cortical bone graft was harvested from the tibia of 26-month-old female pathogen-free pigs. The bones were frozen at low temperatures, thawed at room temperature, and washed with sterile saline to get rid of the cell debris from the porcine bone. The bones were chopped into 3 × 3 × 5 mm pieces and a longitudinal central hole was made to enhance intramedullary fixation during operation.

The alpha-gal epitope is the primary antigen that can cause an immune reaction to the porcine organs in the human body. Humans and primates, having one percent of anti-gal antibody in blood, have an immune response to alpha-gal epitope. Acute immune rejection could be caused by even a small amount of exposure to the alpha-gal epitope that is situated on the surface of porcine bone osteocytes. We used purified recombinant alpha-galactosidase to remove the alpha-gal epitope from the bone grafts for 12 hours at 26°C.

Methods

Operative procedure

1) Anesthesia

Anesthesia was performed by injecting 25 mg/kg xylazine (Xylazine hydrochloride; Bayer Korea, Seoul, Korea) and 75 mg/kg ketamine (Ketamine hydrochloride, Daihan Pharm, Seoul, Korea) into the peritoneum of the animals. The right thigh was prepared and sterilized. For prophylactic antibiotics, intraperitoneal gentamicin sulfate (Kuhnil Pharm, Seoul, Korea) was administered before skin incision.

2) Operation

By using the lateral approach, the periosteum was lifted, a 5-mm bone block was removed from the midpoint of the femoral shaft under the third trochanter (Fig. 1), and a bone graft was inserted into the bone defect. Intramedullary fixation was performed using a 21-G needle by two orthopedic surgeons (YHK and YC) with five and nine years of orthopedic experience, respectively (Fig. 2).

In group 1, a syngenic bone graft obtained from the femoral shaft of C57/BL6 wild type mice was inserted to C57/BL6 wild type mice. In group 2, a bone graft obtained from the femoral shaft of C57/BL6 wild type mice was inserted to C57/BL6 alpha-gal knockout mice. In group 3, an alpha-galactosidase-untreated porcine xenobone graft was inserted to C57/BL6 alpha-gal knockout mice. In group 4, alpha-galactosidase-treated porcine xenobone graft was inserted to C57/BL6 alpha-gal knockout mice.

3) Postoperative care

In all mice, 0.1 mg/kg buprenorphine (SK Chemicals Life Science, Seoul, Korea) was administered subcutaneously every 10–14 hours to relieve pain after the operation. Weight-bearing was not limited, and the environment of the cage was maintained as that of the preoperative condition.

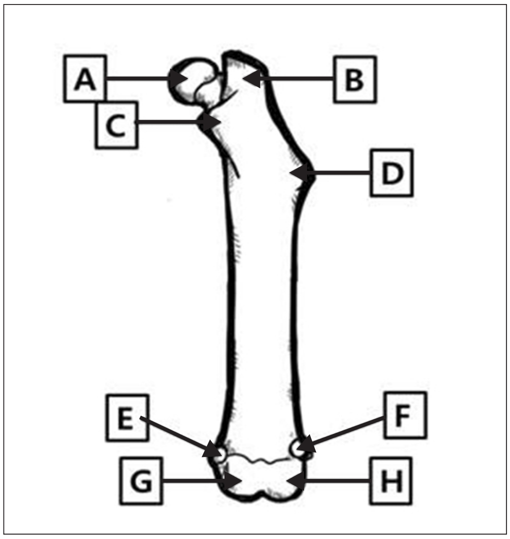

Fig. 1. Anatomy of the femur of the rat. A: head of the femur, B: greater trochanter, C: lesser trochanter, D: third trochanter, E: medial fabella, F: lateral fabella, G: medial condyle, H: lateral condyle.

4) Sample (or tissue) harvest

At 5 weeks after operation, cervical dislocation was used to sacrifice the animal under anesthesia. After obtaining the blood sample from the inferior vena cava, the right femur was taken out.[10]

5) Radiographic evaluation of bony union

The extracted femur was prepared by removing the surrounding soft tissue and anteroposterior radiographs of the femur were taken 5 weeks after operation. UT 2000 radiograph machine (Philips Research, Eindhoven, the Netherlands) was used to obtain femoral radiographs including anteroposterior, lateral, and Mortis views. The extracted femur was placed at approximately 100 cm from the source in the anatomical position. The setting of the radiograph depended on the extracted femur size and was 46 to 50 kVp and 4.5 to 5 mAs (Fig. 3).

Radiographic bony union was evaluated using the Goldberg method.[8] First, the appearance of a graft was evaluated and scored as 0 for resorbed, 1 for mostly resorbed, 2 for largely intact, and 3 for reorganized. Second, bony union was evaluated and scored as 0 for nonunion, 1 for possible union, and 2 for radiographic union from the proximal to the distal end of the graft (Table 1).

Consensus building of radiographic measurements

We had a formal meeting for consensus building according to the access procedure.[11-13] The examiners of the study (1) investigated current practices, (2) presented a specific and systematic review of related articles, (3) discussed thoroughly, and (4) voted on preferences in private.

Prior to the study, following the previously proposed process, a consensus-building session was held by all examiners before measurements of radiographs. Three items

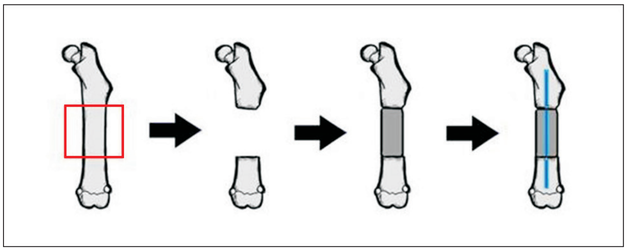

Fig. 2. Bone graft operation. After cutting of the diaphyseal portion of the host femur, the bone graft was placed in the gap and fixed with an intramedullary rod.

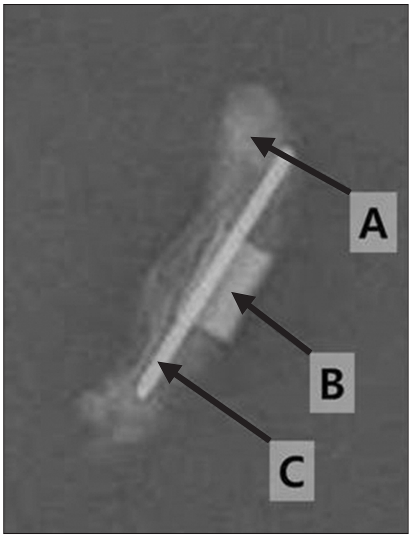

Fig. 3. Radiographic findings of a bone graft fixed with an intramedullary rod. A: femur, B: bone graft, C: intramedullary rod.

of radiographic measurements in the Goldberg scoring system to be assessed were reviewed from literatures and consensus was reached by the orthopedic surgeons. Previous studies were reviewed,[8,14-17] and one of the examiners (YHK) defined the measurements that we believed would be relevant to assessment of bony union after bone grafting. Standards for measurements were selected based on the consensus among the three orthopedic surgeons (YHK, YC, and SWM) who belonged to the consensus development panel.

We tried to cover as many radiographic measurements as we could. Radiographic measurement methods were chosen by eliminating unnecessary and less frequently used methods. The three members of the panel (YHK, YC, and SWM) were orthopedic surgeons with five, eight, and seven years of experience each. One of the examiners (YHK) was an arbitrator and each item was selected by the agreement of the panel. The discussion focused mostly on the universal use and significance of each measurement.

Intra- and interobserver reliability test

Intra- and interobserver reliability of radiographic measurements was assessed among the three examiners (YHK, YC, and SWM). A prior sample size estimation by precision analysis indicated that a minimum of 36 radiographs should be assessed for bony union after bone grafting.[18] Intra- and interobserver reliability of the radiographic measurement of bony union after bone grafting was assessed with use of the measurements obtained by the three examiners. As the interval for intraobserver reliability testing needs to be long enough so that it does not cause a recall bias, we reviewed previous literatures.[19-22] After the review, the three examiners performed assessments in two sessions that were separated by a 3-week interval.

Each examiner was blinded to the other measurements and to all mice data. All measurements were collected by a research assistant (EYL) who did not otherwise participate in the study.

Table 1. Goldberg Radiographic Scoring System

Radiographic assessment	Point
Appearance of graft	
Resorbed	0
Mostly resorbed	1
Largely intact	2
Reorganizing	3
Union (proximal and distal evaluated separately)	
Nonunion	0
Possible union	1
Radiographic union	2
Total points possible per category	
Graft	3
Proximal union	2
Distal union	2
Maximum score	7

Statistical Methods

This study was designed such that the κ could be used to examine reliability.[23] Reliability was assessed with the κ at a target point of 0.6, the minimum sample size of 36 radiographs, and Bonett method of approximation.[18] The data on the femur in each mouse were selected by means of block randomization and were included for statistical analysis.

The κ was calculated in the setting of a two-way random-effect model, assuming a single measurement and absolute agreement.[24] In this study, the κ was characterized as poor for < 0.00, slight for 0.00 to 0.20, fair for 0.21 to 0.40, moderate for 0.41 to 0.60, substantial for 0.61 to 0.80, and almost perfect for 0.81 to 1.00.[25] A κ can be interpreted as follows: κ = 1 indicates perfect reliability, κ = 0 indicates opposite, and a κ of > 0.6 indicates substantial reliability. A *p*-value of < 0.05 was considered significant. The data were analyzed statistically using IBM SPSS ver. 26.0 (IBM Corp., Armonk, NY, USA).

RESULTS

Thirty-six mice with bone grafting were enrolled in the study group. The average score for the appearance of graft, proximal union, and distal union was 2.5 (standard deviation [SD], 0.5; range, 2–3), 1.7 (SD, 0.5; range, 1–2), and 1.6 (SD, 0.5; range, 0–2), respectively. The average total score was 5.8 (SD, 1.3; range, 3–7) (Table 2). Group 1 had the most advanced bony union and group 2 had more advanced bony union than groups 3 and 4 evaluated by the Goldberg scoring system (Table 3).

The intraobserver and interobserver reliability tests for the Goldberg scoring system were performed without the results of one research assistant (EYL), whose interobserver reliability was relatively low. The overall intra- and interobserver reliability showed excellent agreement.

The κ for intraobserver reliability of radiographic measurements ranged from 0.653 to 0.889 for each value and from 0.603 to 0.758 for the total score. The κ for interobserver reliability of radiographic measurements ranged from 0.671 to 0.888 for each value and from 0.641 to 0.649 for the total score (Tables 4-7).

Table 2. Summary of Measured Data

Parameter	Mean (range)	Standard deviation
Appearance of graft	2.5 (2–3)	0.5
Union (proximal)	1.7 (1–2)	0.5
Union (distal)	1.6 (0–2)	0.5
Total Score	5.8 (3–7)	1.3

Intraobserver Reliability Test

Measurements of the appearance of graft and the union (proximal and distal evaluated separately) showed satisfactory overall reliability. Most of the radiographic measurements showed good-to-excellent reliability.

For the first examiner (YHK), the κ was the highest for the appearance of graft (κ, 0.889), and distal union showed the lowest intraobserver reliability (κ, 0.653). The score for proximal union was 0.738. All radiographic measurements showed more than substantial reliability, and the total score showed substantial reliability (κ, 0.603).

For the second examiner (YC), the κ was the high-

Table 3. Summary of Measured Data

Variable	Group 1	Group2	Group3	Group 4
Appearance of graft				
Examiner 1	3.0 ± 0.0	2.9 ± 0.3	2.0 ± 0.0	2.0 ± 0.0
Examiner 2	3.0 ± 0.0	2.7 ± 0.5	2.1 ± 0.3	2.0 ± 0.0
Examiner 3	3.0 ± 0.0	2.8 ± 0.4	2.0 ± 0.0	2.0 ± 0.0
Union: proximal				
Examiner 1	1.9 ± 0.3	2.0 ± 0.0	1.6 ± 0.5	1.3 ± 0.5
Examiner 2	2.0 ± 0.0	2.0 ± 0.0	1.7 ± 0.5	0.9 ± 0.3
Examiner 3	2.0 ± 0.0	2.0 ± 0.0	1.6 ± 0.5	1.2 ± 0.4
Union: distal				
Examiner 1	1.9 ± 0.3	2.0 ± 0.0	1.3 ± 0.7	1.2 ± 0.4
Examiner 2	2.0 ± 0.0	2.0 ± 0.0	1.1 ± 0.6	1.2 ± 0.7
Examiner 3	2.0 ± 0.0	2.0 ± 0.0	1.3 ± 0.8	1.1 ± 0.3
Total score				
Examiner 1	6.8 ± 0.7	6.9 ± 0.3	4.9 ± 0.9	4.6 ± 0.7
Examiner 2	7.0 ± 0.0	6.7 ± 0.5	4.9 ± 1.2	4.1 ± 0.8
Examiner 3	7.0 ± 0.0	6.8 ± 0.4	4.9 ± 0.8	4.3 ± 0.5

Values are presented as mean ± standard deviation.

Table 4. Intra- and Interobserver Reliability of Appearance of Graft as Measured on Radiographs of Femur

Variable	Coefficient	*p*-value
Intraobserver reliability	Cohen's kappa	
First examiner	0.889	
Second examiner	0.888	
Third examiner	0.888	
Interobserver Reliability	Fleiss Kappa	
First session	0.888	< 0.005
Second session	0.837	< 0.005

Table 5. Intra- and Interobserver Reliability of Proximal Union as Measured on Radiographs of Femur

Variable	Coefficient	*p*-value
Intraobserver reliability	Cohen's kappa	
First examiner	0.738	
Second examiner	0.880	
Third examiner	0.884	
Interobserver reliability	Fleiss Kappa	
First session	0.746	< 0.005
Second session	0.767	< 0.005

Table 6. Intra- and Interobserver Reliability of Distal Union as Measured on Radiographs of Femur

Variable	Coefficient	*p*-value
Intraobserver reliability	Cohen's kappa	
First examiner	0.653	
Second examiner	0.776	
Third examiner	0.869	
Interobserver reliability	Fleiss Kappa	
First session	0.671	< 0.005
Second session	0.699	< 0.005

Table 7. Intra- and Interobserver Reliability of Total Score as Measured on Radiographs

Variable	Coefficient	*p*-value
Intraobserver reliability	Cohen's kappa	
First examiner	0.603	
Second examiner	0.790	
Third examiner	0.758	
Interobserver reliability	Fleiss Kappa	
First session	0.641	< 0.005
Second session	0.649	< 0.005

est for the appearance of graft (κ, 0.880) and distal union showed the lowest intraobserver reliability (κ, 0.776). The score for proximal union was 0.880. All radiographic measurements showed more than substantial reliability, and the total score showed substantial reliability (κ, 0.790).

For the third examiner (SWM), the κ was the highest for the appearance of graft (κ, 0.888), and distal union showed the lowest intraobserver reliability (κ, 0.869). The score for proximal union was 0.884. All radiographic measurements also showed almost perfect reliability, and the total score showed substantial reliability (κ, 0.758).

Interobserver Reliability Test

In terms of interobserver reliability, the appearance of graft and the union (proximal and distal evaluated separately) showed substantial reliability. On the first session, the κ was the highest for the appearance of graft (κ, 0.888, $p < 0.005$), and distal union showed the lowest interobserver reliability (κ, 0.671, $p < 0.005$). The score for proximal union was 0.746 ($p < 0.005$). Radiographic measurements showed substantial to almost perfect reliability, and the total score showed substantial reliability (κ, 0.641, $p < 0.005$).

On the second session, the κ was the highest for the appearance of graft (κ, 0.837, $p < 0.005$), and distal union showed the lowest interobserver reliability (κ, 0.699, $p < 0.005$). The score for proximal union was 0.767 ($p < 0.005$). Radiographic measurements showed substantial to almost perfect reliability, and the total score showed substantial reliability (κ, 0.649, $p < 0.005$).

DISCUSSION

We believe that the reliability of many radiographic evaluations on bony union have not been established although radiographic measurements are used frequently.[14,15,17] In this study, the Goldberg scoring system, which included the appearance of graft, proximal union, and distal union as investigating items, was found to be reliable for evaluating bony union after bone grafting.[8]

Some limitations of this study should be addressed before discussing these findings in detail. First, although the reliability of radiographic evaluation was tested, the validity of radiographs was not tested in this study. The validity and reliability of the reference standard method should be documented before addressing the validity of evaluation of bony union in this study. Despite the relatively high reliability, however, it has not been elucidated whether radiography is a valid method for evaluating bony union. Therefore, it is important to further evaluate validity of radiographic measurement for clinical use. However, radiographs have been widely used to evaluate bony union, and the validity of radiographic evaluation of bony union was demonstrated in cadaveric studies. Its use has been advocated in numerous studies. Other methods such as computed tomography and histologic evaluation are valid tools for measuring the level of bony union and can be used as reference standards. Second, the interval between measurements for intraobserver reliability testing needs to be long enough so that it does not cause recall bias. Although the period of 3 weeks between independent measurements was to determine intraobserver reliability, this period was rather short and could have increased intraobserver reliability. Nevertheless, our results showed similar interobserver reliability without recall bias, and there are many studies that used a 3-week interval for intraobserver reliability testing.[20-22] Third, the variation range must be wide enough. In a narrow variation range, one can easily guess that reliability is superior to a wide range. In other words, correlation may be higher when subjects are homogeneous, and correlation may be lower when subjects are heterogeneous.[26] In the Goldberg scoring system, there are three component items and the range of score is from 0 to 3 for appearance of graft and 0 to 2 for proximal and distal union. These scores have narrow ranges, which may affect reliability. Forth, reliability of the Goldberg scoring system tested through animal experiments could not be perfectly applied to human studies. Clinically, the Goldberg scoring system was designed for human body, but we evaluated bone union of mice in this study. The advantage of animal experiment is that we could evaluate bony union of diverse bone graft materials in a short-term study. We also could evaluate bony union under the same location and duration with mice. However, it cannot be perfectly applied clinically since DNA of mice and human is different. So, further prospective studies for human bony union after bone grafting should be conducted.

The κ is used to evaluate the consistency of the measurements performed by many different examiners measuring the same categorical variables.[23] If several examiners are asked to score the results of bony union after bone grafting, we can assess how consistent the scores are to each other.

An important aspect of this problem is that there is both interobserver and intraobserver variability. Interobserver variability refers to systematic differences among the observers—one observer may consistently score patients at a higher risk level than other observers. Intraobserver variability refers to deviations of a particular observer's score on a particular patient that are not part of a systematic difference.[9]

The κ is made to be used for exchangeable measurements; that is, grouped data without a meaningful way to order measurements within a group. In this study, on the Goldberg scoring system, most of the radiographic measurements showed substantial to almost perfect intraobserver reliability and excellent interobserver reliability. However, on intraobserver reliability, the items showed some different values.

The subsection of the appearance of graft showed the highest intraobserver reliability for measurement of bony union after bone grafting. The appearance of graft is both intuitive and easily understood. It is especially obvious between the mostly resorbed (score 1) and largely intact (score 2). We think that between two subsections, the acceptable range is broad, and we speculate that this factor may have influenced the highest intraobserver reliability results.

The subsection of the bony union state of the proximal area showed relatively high intraobserver reliability for measurements compared to the union of distal area. This may be because in the proximal area of the femur, there was a large area to contact bone grafting and bony fixation was convenient. In this state, distinction was facilitative for measurements and in particular, the sort of “nonunion (score 0)” and “possible union (score 1)” was convenient because meaning of nonunion is clear to evaluation. We think that these conditions may have affected the results.

With regard to bony union of the distal area, intraobserver reliability for measurements was relatively low compared to that of the union of the distal area. Distal area of the femur was tiny and small to contact bone graft and fixation was not stable compared to the proximal area. Nevertheless, there was an equal provision of the union of proximal area, which is obvious to measurement.

In this study, we demonstrated that the Goldberg scoring system was a reliable indicator for evaluating bony union. With more subsections and advanced classification, the system may offer more detailed and accurate results for evaluation of bony union.

The Goldberg scoring system was found to be a reliable tool for describing radiographic bony union after bone grafting in an animal model. Further study is required to determine its usefulness in the evaluation of human bony union after bone grafting.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

We would like to thank Eun Young Lee (Bachelor of Business Administration, Seoul National University Bundang Hospital) and Sang-won Moon (Department of Orthopaedic Surgery, Inje University Haeundae Paik Hospital) for their research assistance.

ORCID

Young Choi	https://orcid.org/0000-0003-3929-6315
Young Hoon Yang	https://orcid.org/0000-0003-1922-0057
Young-Ho Kwon	https://orcid.org/0000-0002-0811-3750

REFERENCES

1. Finkemeier CG. Bone-grafting and bone-graft substitutes. J Bone Joint Surg Am. 2002;84(3):454-64.
2. Nandi SK, Roy S, Mukherjee P, Kundu B, De DK, Basu D. Orthopaedic applications of bone graft & graft substitutes: a review. Indian J Med Res. 2010;132:15-30.
3. Bauer TW, Muschler GF. Bone graft materials: an overview of the basic science. Clin Orthop Relat Res. 2000;(371):10-27.
4. de Oliveira E Silva M, Pelegrine AA, Alves Pinheiro da Silva A, et al. Xenograft enriched with autologous bone marrow in inlay reconstructions: a tomographic and histomorphometric study in rabbit calvaria. Int J Biomater. 2012;2012:170520.
5. Voor MJ, Yoder EM, Burden RL Jr. Xenograft bone inclusion improves incorporation of hydroxyapatite cement into cancellous defects. J Orthop Trauma. 2011;25(8):483-7.
6. Bi L, Hu Y, Fan H, et al. Treatment of contaminated bone defects with clindamycin-reconstituted bone xenograft-composites. J Biomed Mater Res B Appl Biomater. 2007;82(2):418-27.
7. Wang ZG, Liu J, Hu YY, et al. Treatment of tibial defect and bone nonunion with limb shortening with external fixator and reconstituted bone xenograft. Chin J Traumatol. 2003;6(2):91-8.
8. Goldberg VM, Powell A, Shaffer JW, Zika J, Bos GD, Heiple KG. Bone grafting: role of histocompatibility in transplantation. J Orthop Res. 1985;3(4):389-404.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74.
10. Garcia P, Histing T, Holstein JH, et al. Rodent animal models of delayed bone healing and non-union formation: a comprehensive review. Eur Cell Mater. 2013;26:1-12.
11. Junger S, Brearley S, Payne S, et al. Consensus building on access to controlled medicines: a four-stage Delphi consensus procedure. J Pain Symptom Manage. 2013;46(6):897-910.
12. Mull HJ, Nebeker JR, Shimada SL, Kaafarani HM, Rivard PE, Rosen AK. Consensus building for development of outpatient adverse drug event triggers. J Patient Saf. 2011;7(2):66-71.
13. Vitale MG, Riedel MD, Glotzbecker MP, et al. Building consensus: development of a Best Practice Guideline (BPG) for surgical site infection (SSI) prevention in high-risk pediatric spine surgery. J Pediatr Orthop. 2013;33(5):471-8.
14. Goldberg VM. The biology of bone grafts. Orthopedics. 2003;26(9):923-4.
15. Goldberg VM, Shaffer JW, Field G, Davy DT. Biology of vascularized bone grafts. Orthop Clin North Am. 1987;18(2):197-205.
16. Goldberg VM, Stevenson S. The biology of bone grafts. Semin Arthroplasty. 1993;4(2):58-63.
17. Goldberg VM, Stevenson S. Bone graft options: fact and fancy. Orthopedics. 1994;17(9):809-10, 821.
18. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med. 2002;21(9):1331-5.
19. McKelvie SJ. Does memory contaminate test-retest reliability? J Gen Psychol. 1992;119(1):59-72.
20. Lee KM, Chung CY, Park MS, Lee SH, Cho JH, Choi IH. Reliability and validity of radiographic measurements in hindfoot varus and valgus. J Bone Joint Surg Am. 2010;92(13):2319-27.

21. Lee YK, Chung CY, Koo KH, Lee KM, Kwon DG, Park MS. Measuring acetabular dysplasia in plain radiographs. Arch Orthop Trauma Surg. 2011;131(9):1219-26.

22. Park MS, Chung CY, Lee KM, Kim TW, Sung KH. Reliability and stability of three common classifications for Legg-Calve-Perthes disease. Clin Orthop Relat Res. 2012;470(9): 2376-82.

23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.

24. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420-8.

25. Campos S, Zhang L, Sinclair E, et al. The palliative performance scale: examining its inter-rater reliability in an outpatient palliative radiation oncology clinic. Support Care Cancer. 2009;17(6):685-90.

26. Kurande VH, Waagepetersen R, Toft E, Prasad R. Reliability studies of diagnostic methods in Indian traditional Ayurveda medicine: an overview. J Ayurveda Integr Med. 2013; 4(2):67-76.