

Clinical Research

Translation and Validation of the German New Knee Society Scoring System

Mahmut Enes Kayaalp MD, Thomas Keller PhD, Wolfgang Fitz MD, Giles R. Scuderi MD, Roland Becker MD, PhD

Received: 29 May 2018 / Accepted: 10 October 2018 / Published online: 26 December 2018
Copyright © 2018 by the Association of Bone and Joint Surgeons

Abstract

Background In 2011 the Knee Society Score (KSS) was revised to include patient expectations, satisfaction, and physical activities as patient-reported outcomes. Since the new KSS has become a widely used method to evaluate patient status after TKA, we sought to translate and validate it for German-speaking populations.

Questions/purposes After translation of the new KSS into German using established guidelines, we sought to test the new German version for (1) validity; (2) responsiveness; and (3) reliability.

Methods The new KSS form was translated and adapted according to the available guidelines. The final version was used to validate the German version of the new KSS

One of the authors certifies that he (TK) has received benefits, during the study period, in an amount of less than USD 10,000 from Medizinischen Hochschule (Brandenburg, Germany). One of the authors certifies that he (WF) has received royalties and stock options, during the study period, in an amount of USD 100,001 to USD 1,000,000 from Conformis (Billerica, MA, USA), outside the submitted work. One of the authors certifies that he (GRS) has received royalties and consulting fees, during the study period, in an amount of USD 100,001 to USD 1,000,000, from Zimmer Biomet (Warsaw, IN, USA); research support and consulting fees, during the study period, in an amount of USD 10,000 to USD 100,000 from Pacira (Parsippany, NJ, USA); consulting fees in an amount of less than USD 10,000 from Medtronic (Minneapolis, MN, USA); and an honorarium during the study period in an amount of less than USD 10,000 from Acelyty (San Antonio, TX, USA), outside the submitted work.

Clinical Orthopaedics and Related Research® neither advocates nor endorses the use of any treatment, drug, or device. Readers are encouraged to always seek additional information, including FDA approval status, of any drug or device before clinical use. Each author certifies that his institution approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research.

This work was performed in the Department of Orthopaedics and Traumatology, Medical School Theodor Fontane University Hospital Brandenburg an der Havel, Germany.

M. E. Kayaalp, Department of Orthopaedics and Traumatology, Cerrahpasa Faculty of Medicine, Istanbul University, Istanbul, Turkey

T. Keller, StatConsult GmbH, Magdeburg, Germany

W. Fitz, Brigham and Women's Hospital, Boston, MA, USA

G. R. Scuderi, Northwell Orthopedic Institute, New York, NY, USA

M. E. Kayaalp, R. Becker, Department of Orthopaedics and Traumatology, University Hospital Brandenburg, Medical School Brandenburg Theodor Fontane, Brandenburg an der Havel, Germany

M. E. Kayaalp (✉), Department of Orthopaedics and Traumatology, Istanbul University-Cerrahpasa Faculty of Medicine, Kocamustafapasa Cd. 53, 34098 Istanbul, Turkey, email: mek@mek.md

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

(GNKSS) in 133 patients undergoing TKA, of which 100 patients were included in the study as per inclusion criteria. Patients completed the GNKSS form along with the German WOMAC and the German SF-36 scores preoperatively and at the 2-year postoperative followup. Construct validity was tested by comparing domain scores of the GNKSS with domain scores of the German WOMAC and the SF-36. Responsiveness was evaluated by comparing pre- and postoperative scores in all questionnaires in all patients using standardized response means. To evaluate reliability, every second patient ($n = 50$) in the whole group was asked to complete the GNKSS form a second time 1 week after their 2-year followup; 39 patients responded. This sample group was considered representative after testing the difference among age, sex, body mass index, operation side, preoperative or postoperative GNKSS, and WOMAC scores with the original group. Intraclass correlation coefficients (ICCs) were used to assess reliability and Cronbach's α was an indicator of internal consistency of each domain score.

Results Construct validity was excellent pre- and postoperatively between the GNKSS and the WOMAC for domains including symptoms, satisfaction, total functional score, and total score and activity subdomains, except the expectation domain and advanced and discretionary subdomains of the GNKSS and the stiffness domain of WOMAC. The expectation domain showed either no significant correlation or only weak correlations with the domains of WOMAC pre- as well as postoperatively (r ranging between -0.19 and -0.34). Correlation of the function section of the GNKSS as well as the physical function and role-physical domains of the SF-36 pre- and postoperatively were moderate to strong, respectively, with statistically significant ($p < 0.001$) r values of 0.49 and 0.48 preoperatively and 0.73 and 0.65 postoperatively. Correlation of the symptom section of the GNKSS and bodily pain domain of the SF-36 was also strong pre- and postoperatively. Regarding responsiveness, all domains of the GNKSS showed large changes except the expectation domain. The symptom and functional sections of the GNKSS showed higher responsiveness than the corresponding pain and function domains of the WOMAC and bodily pain and physical function domains of the SF-36. Also, the total score changes were larger for the GNKSS compared with the WOMAC. No floor or ceiling effect was observed. Reliability was excellent with ICCs of 0.83 to 0.97 as an indicator of test-retest reliability and Cronbach's α values of 0.78 to 0.85 preoperatively and 0.92 to 0.94 postoperatively as an indicator of internal consistency for all domains and subdomains.

Conclusions The GNKSS is a valid, responsive, reliable, and consistent outcome measurement tool that may be used to evaluate the outcome of TKA.

Level of Evidence Level II, diagnostic study.

Introduction

The Knee Society Score (KSS) has been widely accepted and used worldwide after its initial introduction in 1989 [11]. There have been concerns regarding its reliability and responsiveness, however, and contemporary patients may have expectations that differ from those of nearly 30 years ago [22]. The increased number of younger patients undergoing TKA and the projected growth of TKAs have also contributed to the need for an updated scoring system that reflects enhanced functional and recreational activities after TKA [20]. Consequently, a new KSS was developed in 2011 that included a new patient-reported outcome section that also measures satisfaction, functional activities, and expectations. The new KSS has been validated in terms of reliability and consistency [20].

The increase in TKAs in Germany mirrors that of the United States [33]. To compare international outcomes after TKA, accepted outcome scoring instruments locally adapted to each language and culture are necessary. For this purpose, the new KSS has been translated and validated in various languages including Dutch, Japan, Chinese, and Korean in previous studies [10, 13, 16, 31]. However, a validated German-language version of the new KSS does not exist.

The purpose of this study therefore was to establish a validated German version of the new KSS for German-speaking individuals undergoing TKA by translating the new KSS into German and testing it in terms of (1) construct validity; (2) responsiveness; and (3) reliability.

Patients and Methods

Translation

The translation and adaptation of the score were completed according to previously published guidelines [2]. Two independent native German-speaking translators (RB, VM) translated the new KSS from English to German. One translator was aware of the process and the other unaware. Both versions were evaluated and merged into a single translation draft. This draft was then backtranslated by two other independent translators (WF, MEK). Finally, a review committee evaluated the translations and established a prefinal version of the scoring tool. This version was tested on 30 patients with osteoarthritis to identify comprehension issues or problems associated with completing the questionnaire. We made some minor modifications, and the final version was established (Appendices, [Supplemental Digital Content 1](#) and [Supplemental Digital Content 2](#)).

Study Design

We obtained ethical approval from the ethical committee of the state of Brandenburg, Germany, before starting the study. A total of 133 patients were identified who had undergone primary TKA from the first quarter of 2014 to the third quarter of 2015. Inclusion criteria were patients undergoing primary TKA who provided their consent, could complete the questionnaire without assistance, and were willing to complete the questionnaire at followup visits. Patients were excluded if they underwent revision TKA during the followup period (n = 1), had surgery on the contralateral side (n = 6), had accompanying hip or lumbar pain (n = 8), or did not complete all three questionnaires sufficiently as required for score calculations (n = 18) as per user manuals for each questionnaire as explained subsequently. The resulting 100 patients, 60 of whom were women, were included in the study (Table 1). We determined the number of patients and methods to be used in this validation study based on prior reports [2, 29]. The well-defined adaptation guidelines from Beaton et al. [2] suggest a sample size of 30 to 40 patients for pretesting a new scoring tool, but does not suggest a minimum group size for validation. However, Terwee et al. [29] recommend at least 100 patients for internal consistency, and 50 patients are needed to effectively evaluate floor and ceiling effects and to evaluate construct validity in validation studies. Still, a prior meta-analysis indicated a lack of clear guidance and consensus about sample size determination [1]. Finally, because the new KSS is designed to assess patients before and after TKA, it has two separate versions for each application with totally different questions in the expectation section. Therefore, separate assessments of both forms for construct validity were necessary. However, some prior translation and validation studies of the new KSS included either postoperative patients or mixed groups of patients from preoperative and postoperative

periods by stating the sum of both groups as the sample size group of their study [10, 16, 31]. In the current study, the same sample group of patients was included pre- and postoperatively to prevent any related responder issues and to validate both versions to make them available at the same time.

All patients were evaluated preoperatively and underwent primary TKAs at the University Hospital of Brandenburg Medical School Department of Orthopaedics and Traumatology and were then reevaluated at the 2-year followup point. The mean followup time interval was 24 months (range, 22-26 months). The mean age and body mass index (BMI) at the time of preoperative evaluation were 72 (± 9) kg/m² and 31 (± 5) kg/m², respectively (Table 1).

All patients were asked to complete the three questionnaires (German New KSS [GNKSS], WOMAC, SF-36) pre- and postoperatively at the 2-year followup. The patients completed all the forms except the objective portion of the new KSS without any assistance. A senior orthopaedic resident (MEK) under the supervision of the senior attending (RB) completed the clinical and radiologic examinations on all patients. Radiographs were used to document implant position and to exclude component loosening. The WOMAC scores were converted into a 100-point scale before statistical analyses were made, as mentioned in a previous study [26]. Scores from the German SF-36 were included as eight individual single-domain scores with the mental and physical summary scores on a 100% scale as previously recommended [15, 28].

Every second one of the randomly listed patients (n = 50) by the data recording program were asked to repeat the questionnaire 1 week after the 2-year followup. Patients did not receive any treatment of any kind in this timeframe to test reliability of the GNKSS as recommended in the literature [31]. Patients were asked to send these forms per mail to the hospital. The symptom section of these forms was specifically designed to let patients fill out only the 10-point scales but not the scoring boxes, because this part of the questionnaire should be filled out or calculated by the physician as the developers intended [30]. Thirty-nine patients out of 50 sent back the completed forms by mail to the hospital. For a Type I error rate, Van der Straeten et al. [31] recommended a sample size of 38 for a power of 0.90 and an α of 0.05, but in their mixed population, groups from pre- and postoperative periods were used without a clear definition of the sample configurations. However, the sample group in this study included patients only from the postoperative period, which creates a more uniform sample group with likely more powerful results.

The sample group was compared with the original group in terms of age, gender, BMI, and preoperative and postoperative GNKSS total and functional scores. The chi-square test, Student's t-test, and Mann-Whitney U tests

Table 1. Demographic data of the sample group population (n = 100)

Variable	Value
Age (years; mean ± SD)	72 ± 9
Sex, number (%)	
Women	60 (60%)
Men	40 (40%)
Side, number (%)	
Right	47 (47%)
Left	53 (53%)
BMI (kg/m ² ; mean ± SD)	31 ± 5
Women	32 ± 6
Men	30 ± 3

BMI = body mass index.

were performed as required for evaluating the differences between the groups. No significant differences for gender ($p = 0.912$), age ($p = 0.592$), BMI ($p = 0.421$), preoperative GNKSS total ($p = 0.454$) and functional ($p = 0.590$) scores, or postoperative GNKSS total ($p = 0.150$) and functional ($p = 0.153$) scores were observed. The sampling group was considered representative of the original group according to these findings.

All collected clinical data were then entered into a computerized database (Microsoft Access®, Microsoft Office 2013; Microsoft, Redmond, WA, USA).

Statistical Analysis

Construct Validity

Construct validity shows to what degree a test measures what it is intended for. Because the GNKSS was intended to reflect patient status before and after primary TKA, a comparison with already existing and validated scoring tools was deemed necessary. Therefore, we analyzed the construct validity using the German WOMAC and the German SF-36, because there is no accepted reference method to reflect the status of patients before and after TKA and because these tests have been validated in previous studies [8, 27]. A comparison among domain scores of all three tools was made pre- and postoperatively in all 100 patients. Using Spearman's coefficient, we computed correlations. These correlations were hypothesized to be either less converging or divergent for mental domains, including expectations, and strongly converging for physical domains as seen in previous studies [20, 31]. The strength of converging correlation was considered weak, moderate, or strong for coefficient values of 0.35, 0.35 to 0.5, and > 0.5 , respectively.

Responsiveness

Responsiveness is the ability of a test to show differences after a defined treatment method. This means differences in pre- and postoperative scores of the GNKSS should reflect the improvements after primary TKA. Greater differences between scores before and after a treatment method would show greater ability of the tool to reflect changes when compared with other scoring tools. To assess responsiveness, correlations of the results of related domains among all tests were evaluated preoperatively and postoperatively using standardized response means (SRM). SRM values were calculated as the mean difference between preoperative and 24-month postoperative scores divided by the SD of the score, whereby the 95% confidence intervals (CIs) were calculated with a jackknife

procedure. The aim was to prove the ability of the outcome measure to reflect the effect of TKA. SRM values were graded in means of change as small (values of 0.2-0.5), moderate (0.5-0.8), and large (> 0.8). We thought that SRM values for each domain of the GNKSS except expectations would be > 0.8 because TKA would be expected to provide better functional results and decrease pain associated with osteoarthritis.

A scoring tool is expected to reflect patients' status in outcome with a normal distribution of scores. In case of an accumulation of scores toward the maximal or minimal zone, it is called a ceiling or floor effect, which shows the limitations of the scoring tool to differentiate among patients' status in outcome. These effects were examined by analysis of distribution of scores to show the ability of the GNKSS to differentiate improvements and further prove its responsiveness.

Reliability

Test-retest reliability tests the stability and consistency of a scoring tool over a time period. Patients complete the forms a second time after a certain timeframe without receiving any additional interim treatment. A timeframe of 1 week was chosen for the evaluation of test-retest reliability because other investigators [18] have suggested that a timeframe ranging from 2 days to 2 weeks seems to be a reasonable compromise between avoiding changes in a patient's condition and preventing recall bias. Test-retest reliability intraclass correlation coefficients (ICCs) were assessed with a 95% CI and evaluated internal consistency was evaluated with Cronbach's α coefficients, which show the ability to maintain coherence of the different components of the scale. Variance components calculated by a random-effects analysis of variance were used to calculate ICCs. Reproducibility was accepted as excellent for an ICC value > 0.8 . Internal consistency was evaluated as fair (0.7 α values), good (0.8 α values), or excellent (0.9 α values).

Other Considerations

When a missing answer was detected, the following process was observed: For the GNKSS, dummy values equal to the average of all of the other items in the same domain were entered. If the patient indicated fewer activities than required in the discretionary activities section, a mean score was inserted for the missing item. Patient responses of "I never do this" were rated as 0 point. All of these steps were in compliance with suggestions made in the user manual from the developers [30]. For the WOMAC, we followed a similar method that was also recommended in the WOMAC user guide [3]. For the SF-36, the missing item

percentage of the related domain was relevant, so when the patient answered more than half of the questions in the respective domain, the missing item was replaced with an average score derived from answers to other questions of the same domain, as recommended [32]. As an overview, 10.6% missing items were detected in the German SF-36, 3% in the German WOMAC, and 1% in the GNKSS, very similar to percentages previously published [31]. If more than the allowed missing items for each test according to relevant guidelines were detected, or two or more domains or whole tests were missing, those patients were excluded.

The results were analyzed using SAS, Version 9.4, software (SAS Institute, Cary, NC, USA).

Statistical analyses were performed by a certified statistician (TK). All the scores included mean and SD values and p values of < 0.05 were considered statistically significant.

Results

Validity

Convergent validity was used to assess the construct validity of the GNKSS, which shows the theoretical correlations between two scoring tools. Correlation of corresponding scores from the GNKSS, the WOMAC, and the SF-36 was strong, suggesting that the GNKSS is able to reflect patient status similar to these already validated tests. WOMAC scores correlated negatively with the GNKSS in reciprocal nature, that is, higher scores on the WOMAC reflected worse clinical outcome, whereas they reflected better clinical outcomes in the new KSS and vice versa. Construct validity, as indicated by Spearman coefficients, was strong among all subdomains of the GNKSS and WOMAC pre- and postoperatively with values between -0.51 ($p < 0.001$) and -0.82 ($p < 0.001$); the only exception was the expectation domain of the new KSS and stiffness domain of the WOMAC, where the correlations showed weak converging results. This was the result of the fact that these domains of the scoring tools did not correspond with any other domains directly, that is, the WOMAC does not include questions regarding patient expectations and the KSS does not evaluate knee stiffness like in the WOMAC, so a high correlation ratio would not be expected. Moreover, all the domains and the four subdomains of the function section of the GNKSS, except the expectation section, as a result of same reasons, correlated moderately or strongly as expected with the German WOMAC total score (Table 2).

Correlations between the physical domains of the SF-36 and the GNKSS such as bodily pain with symptoms and physical function and physical role with activity domains of the GNKSS were moderate to strong in a converging manner preoperatively and postoperatively, as

anticipated a priori, suggesting that the GNKSS also performed similarly with the formerly validated SF-36 in corresponding domains. However, a strong correlation among all domains was not expected, because the SF-36 is a general health-related quality-of-life assessment tool and the GNKSS differs from it because it aims primarily to reflect patient status regarding primary TKA. Coefficient values ranged from 0.48 to 0.73, proving the moderate-to-strong correlations. Furthermore, the correlation of the symptom section of the GNKSS and bodily pain domain of the SF-36 was strong pre- and postoperatively. Mental domains such as vitality and emotional role functioning showed diverging or weakly converging results with all the domains and subdomains of the GNKSS, because these domains do not directly correspond to any domain of the GNKSS as explained previously. These results were in line with the a priori hypothesis that mental components would not show strong correlations with the domains of the GNKSS (Table 3).

Responsiveness

Responsiveness evaluation, as demonstrated by SRMs, showed that all GNKSS domains had large changes, proving the superior ability of the GNKSS to reflect improvements in outcome after primary TKA with SRM values ranging between 1.65 and 2.36, except for the expectation domain, as hypothesized a priori. The symptom section along with the functional subdomains also showed large changes, which were greater than the corresponding pain and function domains of the WOMAC and the physical function and bodily pain domains of the SF-36, which indicates the GNKSS is more responsive in corresponding domains, which are the pain and functional domains. Total score changes were reflected with SRM values of 1.87 for the GNKSS and -1.49 for the WOMAC, which demonstrated that overall responsiveness of the GNKSS was also larger than that of the WOMAC (Table 4), which means that the total GNKSS score is also a more sensitive parameter of showing improvements after primary TKA when compared with the WOMAC total score.

Analysis of distribution of scores at the second year followup showed that there were no floor or ceiling effects.

Reliability

Regarding reliability, all domains of the GNKSS showed excellent results for all domains both in terms of test-retest reliability represented by ICC values between 0.82 and 0.97 and internal consistency represented by Cronbach's α values between 0.78 and 0.85 preoperatively and 0.92 and 0.94 postoperatively (Table 5).

Table 2. Construct validity between the German New Knee Society Score (GNKSS) and the German WOMAC

			German New KSS										
Mean Score (SD)		Pre-op	9.13 (3.84)	13.33 (7.55)	13.76 (2.26)	37.71 (15.53)	13.47 (7.45)	12.09 (5.2)	6.86 (4.68)	5.94 (3.40)	72.69 (22.94)		
		Post-op	20.25 (4.11)	28.39 (8.27)	9.08 (2.39)	65.57 (19.86)	20.56 (8.66)	21.2 (5.46)	12.93 (5.95)	10.88 (3.53)	122.63 (31.46)		
Pre-op	Post-op	Correlation chart	Symptom	Satisfaction	Expectation	Total functional score	Activities				Total		
							Functional	Standard	Advanced	Discretionary			
German WOMAC	56.29 (20.32)	19.84 (19.60)	Pain	t1	-0.63 (<0.001)	-0.52 (<0.001)	-0.17 (0.116)	-0.51 (<0.001)	-0.43 (<0.001)	-0.57 (<0.001)	-0.31 (0.002)	-0.30 (0.035)	-0.62 (<0.001)
				t2	-0.74 (<0.001)	-0.81 (<0.001)	-0.34 (<0.001)	-0.69 (<0.001)	-0.58 (<0.001)	-0.65 (<0.001)	-0.54 (<0.001)	-0.52 (<0.001)	-0.78 (<0.001)
	52.05 (27.71)	29.2 (24.67)	Stiffness	t1	-0.26 (0.010)	-0.32 (0.002)	-0.04 (0.742)	-0.30 (0.003)	-0.27 (0.008)	-0.24 (0.021)	-0.10 (0.354)	-0.17 (0.100)	-0.35 (<0.001)
				t2	-0.61 (<0.001)	-0.68 (<0.001)	-0.20 (0.048)	-0.51 (<0.001)	-0.43 (<0.001)	-0.56 (<0.001)	-0.37 (<0.001)	-0.37 (0.002)	-0.60 (<0.001)
	52.46 (22.08)	22.96 (20.52)	Function	t1	-0.52 (<0.001)	-0.52 (<0.001)	-0.20 (0.072)	-0.52 (<0.001)	-0.42 (<0.001)	-0.54 (<0.001)	-0.36 (<0.001)	-0.36 (0.004)	-0.63 (<0.001)
				t2	-0.70 (<0.001)	-0.77 (<0.001)	-0.30 (0.002)	-0.78 (<0.001)	-0.65 (<0.001)	-0.74 (<0.001)	-0.62 (<0.001)	-0.60 (<0.001)	-0.81 (<0.001)
	53.23 (20.57)	23.54 (19.64)	Total	t1	-0.56 (<0.001)	-0.54 (<0.001)	-0.19 (0.078)	-0.54 (<0.001)	-0.44 (<0.001)	-0.57 (<0.001)	-0.36 (<0.001)	-0.36 (0.004)	-0.65 (<0.001)
				t2	-0.72 (<0.001)	-0.80 (<0.001)	-0.30 (0.003)	-0.77 (<0.001)	-0.65 (<0.001)	-0.76 (<0.001)	-0.61 (<0.001)	-0.59 (<0.001)	-0.82 (<0.001)

KSS = Knee Society Score; SD = Standard deviation; Pre-op = Preoperative; Post-op = Postoperative mean scores respectively; Spearman’s correlation coefficients (r) when comparing GNKSS domains and subdomains with the domains of German WOMAC; t1 and t2 indicates pre-operative and post-operative respective r values. WOMAC scores correlate negatively with the GNKSS due to reciprocal nature of these two scoring tools.

Discussion

The new KSS has been widely adopted worldwide, and several translation and validation studies have been published in other languages. Our purpose was to produce a German version of the new KSS using well-defined guidelines to confirm that it has valid measurement properties [2, 13, 29]. Our results indicate that the GNKSS proposed in this

study is a valid, responsive, reliable, and consistent outcome tool to be used in German-speaking populations to evaluate the pre- and post-TKA status of patients.

The current study has several limitations. First, like in all adaptation and validation studies, patients had to complete three separate scoring tools simultaneously, which may have resulted in missing or invalid responses as a result of an increase in responders’ burden. The developer of

Table 3. Construct validity between the German New Knee Society Score (GNKSS) and the German WOMAC

		German New KSS											
Mean Score (SD)		Pre-op		9.13 (3.84)	13.33 (7.55)	13.76 (2.26)	37.71 (15.53)	13.47 (7.45)	12.09 (5.2)	6.86 (4.68)	5.94 (3.40)	72.69 (22.94)	
		Post-op		20.25 (4.11)	28.39 (8.27)	9.08 (2.39)	65.57 (19.86)	20.56 (8.66)	21.2 (5.46)	12.93 (5.95)	10.88 (3.53)	122.63 (31.46)	
		Pre-op	Post-op	Correlation chart		Symptom	Satisfaction	Expectation	Total functional score	Activities			
								Functional	Standard	Advanced	Discretionary		
German SF-36	28.02 (20.79)	57.63 (25.37)	Physical function	t1	0.18 (0.086)	0.24 (0.021)	0.09 (0.405)	0.49 (<0.001)	0.49 (<0.001)	0.53 (<0.001)	0.34 (0.001)	0.31 (0.003)	0.50 (<0.001)
				t2	0.53 (<0.001)	0.51 (<0.001)	0.33 (0.001)	0.73 (<0.001)	0.62 (<0.001)	0.63 (<0.001)	0.59 (<0.001)	0.65 (<0.001)	0.70 (<0.001)
	26.90 (39.99)	61.89 (42.35)	Role-physical	t1	0.33 (0.003)	0.17 (0.130)	0.02 (0.860)	0.48 (<0.001)	0.50 (<0.001)	0.41 (<0.001)	0.25 (0.029)	0.34 (0.002)	0.43 (<0.001)
				t2	0.45 (<0.001)	0.58 (<0.001)	0.30 (0.006)	0.65 (<0.001)	0.58 (<0.001)	0.58 (<0.001)	0.55 (<0.001)	0.58 (<0.001)	0.65 (<0.001)
	25.81 (16.08)	66.69 (25.09)	Bodily pain	t1	0.52 (<0.001)	0.45 (<0.001)	0.16 (0.153)	0.52 (<0.001)	0.49 (<0.001)	0.52 (<0.001)	0.30 (0.003)	0.41 (<0.001)	0.63 (<0.001)
				t2	0.71 (<0.001)	0.78 (<0.001)	0.44 (<0.001)	0.71 (<0.001)	0.69 (<0.001)	0.62 (<0.001)	0.54 (<0.001)	0.53 (<0.001)	0.79 (<0.001)
	54.65 (18.89)	61 (15.07)	General health	t1	0.31 (0.003)	0.26 (0.014)	0.15 (0.175)	0.29 (0.005)	0.37 (<0.001)	0.18 (0.082)	0.10 (0.343)	0.22 (0.035)	0.37 (<0.001)
				t2	0.18 (0.085)	0.09 (0.412)	0.20 (0.064)	0.21 (0.043)	0.24 (0.025)	0.24 (0.025)	0.15 (0.171)	0.06 (0.586)	0.17 (0.106)
	49.03 (8.79)	50.82 (7.42)	Vitality	t1	-0.07 (0.511)	0.01 (0.892)	0.08 (0.468)	0.06 (0.573)	0.09 (0.415)	0.01 (0.911)	0.08 (0.430)	0.06 (0.560)	0.01 (0.911)
				t2	-0.02 (0.875)	-0.09 (0.390)	-0.11 (0.281)	-0.06 (0.563)	-0.05 (0.618)	0.02 (0.876)	-0.08 (0.429)	-0.08 (0.428)	-0.09 (0.379)
	70.26 (28.95)	85.89 (21.67)	Social function	t1	0.25 (0.016)	0.30 (0.003)	0.20 (0.062)	0.45 (<0.001)	0.56 (<0.001)	0.36 (<0.001)	0.29 (0.005)	0.15 (0.156)	0.48 (<0.001)
				t2	0.36 (<0.001)	0.48 (<0.001)	0.22 (0.033)	0.51 (<0.001)	0.44 (<0.001)	0.40 (<0.001)	0.43 (<0.001)	0.46 (<0.001)	0.52 (<0.001)
	60.26 (47.46)	78.97 (39.67)	Role-emotion	t1	0.26 (0.024)	0.29 (0.01)	0.16 (0.199)	0.34 (0.003)	0.40 (<0.001)	0.19 (0.097)	0.12 (0.315)	0.28 (0.014)	0.46 (<0.001)
				t2	0.37 (<0.001)	0.46 (<0.001)	0.19 (0.084)	0.43 (<0.001)	0.44 (<0.001)	0.37 (<0.001)	0.29 (0.007)	0.33 (0.002)	0.47 (<0.001)
	69.29 (20.47)	75.96 (19.23)	Mental health	t1	0.30 (0.004)	0.37 (<0.001)	0.26 (0.020)	0.44 (<0.001)	0.53 (<0.001)	0.36 (<0.001)	0.18 (0.095)	0.19 (0.077)	0.51 (<0.001)
				t2	0.42 (<0.001)	0.52 (<0.001)	0.22 (0.041)	0.54 (<0.001)	0.47 (<0.001)	0.44 (<0.001)	0.42 (<0.001)	0.49 (<0.001)	0.58 (<0.001)
	30.52 (7.06)	43.07 (8.04)	Physical component summary	t1	0.21 (0.078)	0.07 (0.537)	-0.05 (0.715)	0.43 (<0.001)	0.38 (0.001)	0.43 (<0.001)	0.33 (0.006)	0.39 (0.001)	0.35 (0.003)
				t2	0.64 (<0.001)	0.64 (<0.001)	0.39 (<0.001)	0.78 (<0.001)	0.71 (<0.001)	0.71 (<0.001)	0.63 (<0.001)	0.68 (<0.001)	0.78 (<0.001)
51.09 (11.39)	51.86 (8.39)	Mental component summary	t1	0.30 (0.010)	0.39 (<0.001)	0.21 (0.098)	0.42 (<0.001)	0.51 (<0.001)	0.29 (0.014)	0.19 (0.124)	0.19 (0.125)	0.50 (<0.001)	
			t2	0.20 (0.084)	0.31 (0.008)	0.02 (0.883)	0.23 (0.045)	0.23 (0.045)	0.16 (0.161)	0.13 (0.264)	0.20 (0.081)	0.28 (0.016)	

KSS = Knee Society Score; SD = Standard deviation; Pre-op = Preoperative; Post-op = Postoperative mean scores respectively; Spearman's correlation coefficients (r) when comparing GNKSS domains and subdomains with the domains of German SF-36; t1 and t2 indicates pre-operative and post-operative respective r values

the KSS recognized this and has published a short version [17, 23]. This version should also be validated in a future study for German-speaking populations. Nonetheless, missing item percentages in the current study were similar to prior studies with 10.6% missing items in the German SF-36, 3% in the German WOMAC, and 1% in the GNKSS [31]. Second, only one center was involved in the study, although it is the only tertiary university hospital in its federal state of Brandenburg (population 2.5 million) and its patients reflect rural and urban populations. Theoretically, sample configurations should be conducted with recruitment from different areas and states of a country to decrease the risk of bias related to demographic and cultural factors. However, the results obtained in this current study, which are comparable with other well-designed validation studies, as well as the development study of the new KSS, make it less questionable whether the sample configuration was adequately conducted [13, 20, 31]. The proportion of women (60% [60 of 100]) reflects the published demographic features of German patients undergoing knee arthroplasty [33], and it also matches exactly the gender distributions in the development study of the new KSS [20]. Therefore, an additional analysis regarding the gender composition of our study was not deemed necessary.

The current study proved that the GNKSS has good construct validity by evaluating the convergent validity with the already validated German WOMAC and SF-36. Overall correlation of GNKSS and WOMAC total scores along with corresponding pain and total functional score demonstrated strong correlations, which were very similar to the findings by Kim et al. [13] and Van der Straeten et al. [31]. Correlation of the expectation domain of the new KSS on the other hand showed weak insignificant convergent correlations, whereas Kim et al. [13] found weak divergent correlations, where half of their results were also statistically insignificant. Other investigators either did not publish their data or did not use the WOMAC in evaluating their versions [10, 31]. These results were in line with our a priori hypothesis, because the expectation section of the KSS has no corresponding domain in the WOMAC.

Correlation of the pain and total activity scores of the KSS with the bodily pain and physical function domains of the SF-36 was strong and moderate for the preoperative group and strong for both domains for the postoperative group. Only two other studies in the literature published their comparative correlation results [10, 13]. Their study designs included either only pre- [13] or postoperative [10] patients. Our results regarding pre- as well as postoperative groups' pain domains showed strong correlations, whereas other authors reported either weak [13] or moderate [10] correlations. The seemingly divergent correlation result in the study of Hamamoto et al. [10] is likely the result of a calculation error we explain subsequently. Activity score

correlation was on the other hand moderate for the pre- and strong for the postoperative group in our study, whereas it was strong [13] for pre- and moderate [10] for the postoperative group in other studies. There may be several explanations for these results. Variable correlations between the same domains in pre- and postoperative groups were also observed and reported in the development of the new KSS study [20]. Because the SF-36 is a general health-related quality-of-life assessment tool and the GNKSS was developed to reflect patient status regarding primary TKA, strong correlations were not expected. Moreover, the only available comparative data are published in Asian countries, where their authors explained the variable results as related to cultural differences [13].

Our results also showed that the GNKSS was very responsive. The most responsive domain of the GNKSS was the symptom section; the total functional score and the total score also showed large changes. Both GNKSS and Korean versions [13] showed very similar large changes in symptom-oriented domains of all scoring tools: the symptom domain in the KSS, pain domain in the WOMAC, and bodily pain domain in the SF-36 as well as function-oriented domains; and total functional score of the KSS, function domain of the WOMAC, and physical function domains of the SF-36 (Supplemental Table, [Supplemental Digital Content 3](#)). As expected, the GNKSS proved to be more responsive to changes after TKA compared with the WOMAC and the SF-36. The new KSS was developed to reflect changes in patient status in relation to TKA, whereas the WOMAC was developed to highlight the status of patients with osteoarthritis without being specifically responsive to treatment, and the SF-36 was developed as a monitoring tool of overall well-being of patients [4, 7, 20].

The GNKSS also demonstrated excellent reliability with overall higher ICC scores compared with other validation studies of the new KSS, especially in satisfaction, total functional activity, and total score results [13, 31] (Supplemental Table, [Supplemental Digital Content 4](#)). Our results were excellent in all domains, whereas other studies also showed excellent results except in two domains in the Korean [13] and three domains in the Dutch version [31]. There may be several explanations for the excellent reliability seen here. First, in the current study, test-retest evaluation was made after a mean of 24 months postoperatively, whereas other investigators did it either preoperatively or 12 months postoperatively. Hence, the time interval could be a factor in the slightly different results for the ICCs. Second, cultural and linguistic factors could have affected test-retest evaluations, because we noted that previous German validation studies of other patient-reported outcome measures commonly stated higher scores for ICC in test-retest evaluations when compared with initial scoring tools, although they used

Table 4. Responsiveness of the German New Knee Society Score (GNKSS) compared with the WOMAC and SF-36

Questionnaire		Mean of change	SD	SRM (95% CI)
GNKSS				
Symptom		11.12	4.71	2.36 (1.97-2.76)
Satisfaction		15.01	9.11	1.65 (1.35-1.95)
Expectation		-4.6	3.24	-1.42 (-1.95 to -0.89)
Total functional score		27.86	16.3	1.71 (1.4-2.02)
Activities				
	Functional	7.21	7.87	0.92 (0.68-1.15)
	Standard	9.1	6.37	1.43 (1.11-1.75)
	Advanced	6.11	6.11	1 (0.73-1.27)
	Discretionary	5.05	3.91	1.29 (1.02-1.56)
Total score	49.94	26.71	1.87 (1.51-2.23)	
German WOMAC				
Pain		-36.55	21.27	-1.72 (-2.16 to -1.28)
Stiffness		-22.79	31	-0.74 (-0.99 to -0.48)
Function		-29.62	20.74	-1.43 (-1.88 to -0.98)
Total score		-29.95	20.18	-1.49 (-1.97 to -1.00)
German SF-36				
Physical function		29.66	23.05	1.29 (1.01-1.56)
Role-physical		32.19	44.67	0.72 (0.46-0.99)
Bodily pain		40.26	23.99	1.68 (1.42-1.94)
General health		5.81	16.61	0.35 (0.14-0.56)
Vitality		1.86	10.55	0.18 (-0.03-0.39)
Social function		14.72	24.49	0.60 (0.40-0.81)
Role-emotion		15.24	50.77	0.30 (0.05-0.55)
Mental health		5.81	17.06	0.34 (0.16-0.52)
Physical component summary		12.75	7.57	1.69 (1.29-2.09)
Mental component summary		-0.58	10.34	-0.06 (-0.32-0.21)

SRM values were graded as small, moderate, and large in means of change for values of 0.2 to 0.5, 0.5 to 0.8, and values > 0.8, respectively; WOMAC scores correlate negatively with GNKSS scores as a result of the reciprocal nature of these two scoring tools; SRM = standardized response mean, calculated as the mean change between preoperative and postoperative periods divided by the corresponding SD; CI = confidence interval.

similar time intervals for retesting [5, 6, 9, 19, 21, 25]. Cronbach's α values were calculated for both pre- and postoperative scores. Postoperative results were higher with α values between 0.92 and 0.94 compared with 0.80 to 0.88 preoperatively. These results were interpreted as good and excellent, respectively. In previously published adaptation and translations studies of the new KSS, either pre- or postoperative versions or mixed groups were examined [10, 13, 31]. Nevertheless, our findings are comparable to these results as absolute numbers (Table 6). No discussion of the difference between pre- and postoperative results was noted in review of prior validation studies, but the development study of the new KSS also showed higher α values in every domain and subdomain except the advanced activities subdomain postoperatively [20]. Overall, our results concerning the Cronbach's α coefficients were higher than in the development study and in other

adaptation and validation studies (Table 6). Results from both of these measurement properties proved that the GNKSS is reproducible.

We observed some common issues while analyzing other validation studies regarding the new KSS that may prove helpful to others considering such studies. During pretesting we realized that how patients' responses to symptoms section were noted and scored was prone to produce calculation errors. This section includes two 10-level scale pain questions and responses, which should not be carried to the score box directly adjacent to the scales. Not subtracting the patient's recorded answer from the maximum score of 10 results in disproportionate scores in this section, which causes higher (that is, better) scores although the patient reports more severe pain resulting from the reciprocal nature of the pain scale and symptom section score. Some prior published validation

Table 5. Reliability measurements of the German New Knee Society Score (GNKSS)

The new Knee Society Scores	Test 1 (n = 39)	Test 2 (n = 39)	Test-retest reliability		Internal consistency	
					Preoperative	Postoperative
	Mean ± SD	Mean ± SD	ICC	95% CI	CA	CA
Symptom/25 points	21.28 ± 4.14	21.13 ± 4.17	0.91	0.83-0.95	0.85	0.93
Satisfaction/40 points	29.69 ± 9.21	29.85 ± 9.51	0.92	0.85-0.96	0.83	0.93
Expectation/15 points	9.46 ± 2.76	9.59 ± 2.87	0.82	0.68-0.90	0.88	0.94
Total functional/100 points	71.23 ± 23.34	70.85 ± 22.76	0.97	0.94-0.98	0.80	0.92
Activities						
Functional/30 points	22.44 ± 9.37	22.08 ± 9.29	0.94	0.89-0.97	0.82	0.93
Standard/30 points	22.49 ± 6.49	22.05 ± 6.44	0.90	0.82-0.95	0.83	0.93
Advanced/25 points	14.56 ± 6.69	14.87 ± 6.68	0.92	0.84-0.95	0.83	0.93
Discretionary/15 points	11.74 ± 3.21	11.85 ± 3.30	0.94	0.88-0.97	0.85	0.94
Total/180 points	131.67 ± 37	131.41 ± 37.05	0.97	0.95-0.99	0.78	0.92

Test 1 = postoperative 2-year followup; Test 2 = 1 week after Test 1; reproducibility was accepted as excellent for an ICC value > 0.8; internal consistency was evaluated as fair, good, or excellent for Cronbach's α values of 0.7, 0.8, or 0.9, respectively; CI = confidence interval; CA = Cronbach's α ; ICC = intraclass correlation coefficient.

studies revealed that this possible error may have produced disproportionate results, as can be seen in the French and Japanese versions of the new KSS [10, 12]. Several validation studies have not shared their results either in absolute numbers or in percentages or direction of changes for domains of the new KSS; they have shared only the statistical results of comparisons made with other scoring tools [16, 24, 31]. To prevent the aforementioned error, we contacted the developers and updated this section by marking the pain scale section as "to be calculated by the patient" and the scoring box as "to be calculated by the physician."

When evaluating the results from the scoring tools, we highlighted correlations of corresponding domains of the used scoring tools; we also analyzed total scores from the

WOMAC and the new KSS in relation to each other. Because there is no total score calculation in the SF-36 and summary component scores should not be analyzed on their own, we did not analyze physical and mental summary scores in this manner. We did include them in the study for further observational and comparable value, although a number of validation studies used either only summary scores or made separate analyses using them [14, 15, 28].

The GNKSS is a valid, responsive, reliable, and consistent outcome measurement tool to be used in German populations to evaluate preoperative and postoperative TKA status, including patients' symptoms, expectations, satisfaction, and physical activities. Future studies sampling other German-speaking populations may increase the external validity of the GNKSS.

Table 6. Comparison of internal consistency results as Cronbach's α values

The new Knee Society Score domains	German		English (preliminary version)		Korean	Dutch
	Preoperative	Postoperative	Preoperative	Postoperative	Preoperative	Mixed
Symptom	0.85	0.93	N/A	N/A	0.92	0.96
Satisfaction	0.83	0.93	0.90	0.95	0.89	0.84
Expectation	0.88	0.94	0.79	0.92	0.89	0.91
Total functional score	0.80	0.92	N/A	N/A	0.90	0.93
Activities						
Functional	0.82	0.93	N/A	N/A	0.84	0.96
Standard	0.83	0.93	0.87	0.88	0.91	0.91
Advanced	0.83	0.93	0.88	0.84	0.91	0.87
Discretionary	0.85	0.94	0.72	0.82	0.83	0.86
Total score	0.78	0.92	N/A	N/A	0.93	0.90

Internal consistency was evaluated as fair, good, or excellent for Cronbach's α values of 0.7, 0.8, or 0.9, respectively; KSS = Knee Society Score; N/A = not available.

Acknowledgments We thank Volker Musahl for his contribution in the translation of the new KSS into German and Enes Ahmet Güven for his contribution in statistical evaluations of reliability and sample group representativeness.

References

1. Anthoine E, Moret L, Regnault A, Sebille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes*. 2014;12:176.
2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25:3186-3191.
3. Bellamy N. *WOMAC Osteoarthritis Index: A User's Guide, IV*. London, Ontario, Canada: Health Services, McMaster University; 2000.
4. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to anti-rheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15:1833-1840.
5. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Phys Ther*. 1999;79:371-383.
6. Bolton JE, Humphreys BK. The Bournemouth Questionnaire: a short-form comprehensive outcome measure. II. Psychometric properties in neck pain patients. *J Manip Physiol Ther*. 2002;25:141-148.
7. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, Westlake L. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305:160-164.
8. Bullinger M, Ware J. [The German SF-36 health survey translation and psychometric testing of a generic instrument for the assessment health-related quality of life] [in German]. *Z Gesund Wiss*. 1995;3:21.
9. Curr N, Dharmage S, Keegel T, Lee A, Saunders H, Nixon R. The validity and reliability of the occupational contact dermatitis disease severity index. *Contact Dermatitis*. 2008;59:157-164.
10. Hamamoto Y, Ito H, Furu M, Ishikawa M, Azukizawa M, Kuriyama S, Nakamura S, Matsuda S. Cross-cultural adaptation and validation of the Japanese version of the new Knee Society Scoring System for osteoarthritic knee with total knee arthroplasty. *J Orthop Sci*. 2015;20:849-853.
11. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. *Clin Orthop Relat Res*. 1989;248:13-14.
12. Kayaalp ME. Comment on: 'French adaptation of the new Knee Society Scoring System for total knee arthroplasty.' *Orthop Traumatol Surg Res*. 2018;104:733-734.
13. Kim SJ, Basur MS, Park CK, Chong S, Kang YG, Kim MJ, Jeong JS, Kim TK. Crosscultural adaptation and validation of the Korean version of the new Knee Society knee scoring system. *Clin Orthop Relat Res*. 2017;475:1629-1639.
14. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopaedics: a brief guide. *J Bone Joint Surg Am*. 2015;97:1628-1634.
15. Lins L, Carvalho FM. SF-36 total score as a single measure of health-related quality of life: scoping review. *SAGE Open Med*. 2016;4:2050312116671725.
16. Liu D, He X, Zheng W, Zhang Y, Li D, Wang W, Li J, Xu W. Translation and validation of the simplified Chinese new Knee Society scoring system. *BMC Musculoskelet Disord*. 2015;16:391.
17. Maniar RN, Maniar PR, Chanda D, Gajbhare D, Chouhan T. What is the responsiveness and respondent burden of the new Knee Society Score? *Clin Orthop Relat Res*. 2017;475:2218-2227.
18. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol*. 2003;56:730-735.
19. Naal FD, Impellizzeri FM, Torka S, Wellauer V, Leunig M, von Eisenhart-Rothe R. The German Lower Extremity Functional Scale (LEFS) is reliable, valid and responsive in patients undergoing hip or knee replacement. *Qual Life Res*. 2015;24:405-410.
20. Noble PC, Scuderi GR, Brekke AC, Sikorskii A, Benjamin JB, Lonner JH, Chadha P, Daylamani DA, Scott WN, Bourne RB. Development of a new Knee Society scoring system. *Clin Orthop Relat Res*. 2012;470:20-32.
21. Ofenloch RF, Diepgen TL, Popielnicki A, Weisshaar E, Molin S, Bauer A, Mahler V, Elsner P, Schmitt J, Apfelbacher C. Severity and functional disability of patients with occupational contact dermatitis: validation of the German version of the Occupational Contact Dermatitis Disease Severity Index. *Contact Dermatitis*. 2015;72:84-89.
22. Scuderi GR, Bourne RB, Noble PC, Benjamin JB, Lonner JH, Scott WN. The new Knee Society Knee Scoring System. *Clin Orthop Relat Res*. 2012;470:3-19.
23. Scuderi GR, Sikorskii A, Bourne RB, Lonner JH, Benjamin JB, Noble PC. The Knee Society Short Form reduces respondent burden in the assessment of patient-reported outcomes. *Clin Orthop Relat Res*. 2016;474:134-142.
24. Silva A, Croci AT, Gobbi RG, Hinckel BB, Pecora JR, Demange MK. Translation and validation of the new version of the Knee Society Score-The 2011 KS Score-into Brazilian Portuguese. *Rev Bras Ortop*. 2017;52:506-510.
25. Soklic M, Peterson C, Humphreys BK. Translation and validation of the German version of the Bournemouth Questionnaire for Neck Pain. *Chiropr Man Therap*. 2012;20:2.
26. Stratford PW, Kennedy DM, Woodhouse LJ, Spadoni GF. Measurement properties of the WOMAC LK 3.1 pain scale. *Osteoarthritis Cartilage*. 2007;15:266-272.
27. Stucki G, Meier D, Stucki S, Michel BA, Tyndall AG, Dick W, Theiler R. [Evaluation of a German version of WOMAC (Western Ontario and McMaster Universities) Arthrosis Index] [in German]. *J Rheumatol*. 1996;55:40-49.
28. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? *Qual Life Res*. 2001;10:395-404.
29. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM, de Vet HC. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42.
30. The Knee Society. The 2011 Knee Society Knee Scoring System© Licenced user manual. Available at: http://kneesociety.org/wp-content/uploads/2017/08/2011-KSS-User-Manual_FINAL_12-2012.pdf. Accessed January 10, 2018.
31. Van Der Straeten C, Witvrouw E, Willems T, Bellemans J, Victor J. Translation and validation of the Dutch new Knee Society scoring system. *Clin Orthop Relat Res*. 2013;471:3565-3571.
32. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston, MA, USA: The Health Institute, New England Medical Centre; 1993.
33. Wengler A, Nimptsch U, Mansky T. Hip and knee replacement in Germany and the USA: analysis of individual inpatient data from German and US hospitals for the years 2005 to 2011. *Dtsch Arztebl Int*. 2014;111:407-416.