

# Limitations of Gram staining for the diagnosis of infections following total hip or knee arthroplasty

ZHENGXIAO OUYANG<sup>1,2\*</sup>, ZANJING ZHAI<sup>1\*</sup>, AN QIN<sup>1\*</sup>, HAOWEI LI<sup>1</sup>,  
XUQIANG LIU<sup>1</sup>, XINHUA QU<sup>1</sup> and KERONG DAI<sup>1</sup>

<sup>1</sup>Department of Orthopedics, Shanghai Key Laboratory of Orthopedic Implants, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, Shanghai; <sup>2</sup>Department of Orthopedics, Hunan Cancer Hospital and The Affiliated Cancer Hospital of Xiangya School of Medicine, Central South University, Changsha, Hunan, P.R. China

Received June 4, 2014; Accepted February 4, 2015

DOI: 10.3892/etm.2015.2315

**Abstract.** The diagnosis of prosthetic joint infection (PJI) following total joint arthroplasty is difficult for clinicians to make decisions due to the similar symptoms presented by aseptic loosening and infection. Gram staining (GS) is a widely used test but its value remains controversial due to conflicting results in the diagnosis of PJI. The aim of the present study was therefore to evaluate the value of GS in the diagnosis of PJI. Searches using MEDLINE, EMBASE and OVID databases were conducted for data published between January 1990 and December 2013. Meta-analysis was used to pool the sensitivity, specificity, diagnostic odd ratios (DORs), area under the receiver-operating characteristic curve (AUC), positive-likelihood ratios (PLRs), negative-likelihood ratios (NLRs) and post-test probability. The heterogeneity and publication bias were assessed, and subgroup and meta-regression analyses were conducted. A total of 18 studies, including a total of 4,647 patients, were selected for analysis. The pooled sensitivity and specificity values for the diagnosis of PJI were 0.19 and 1.00, respectively. The AUC, PLR and NLR were 0.89, 41.6 and 0.82, respectively. Subgroup analyses indicated that the sensitivity/specificity for total hip arthroplasty was 0.14/0.99, whereas that for total knee arthroplasty was 0.14/1.00. Synovial fluid best reflected accurate GS-based diagnoses, with the highest DOR of 242, whereas tissue had the highest AUC of 0.96 (95% CI, 0.94-0.97). GS had a poor clinically acceptable diagnostic value for detecting PJI. These data do not support the routine use of GS, without additional proof of infection, for diagnosing PJI; instead, GS

could be used as an adjuvant tool to support the results of other investigations.

## Introduction

Prosthetic joint infection (PJI) occurs in 1-12% of surgical cases and is one of the most common complications associated with total hip arthroplasty (THA) and total knee arthroplasty (TKA), often leading to revision. PJI has been reported to be the most common cause of early failure and is associated with several adverse outcomes (1,2). Infection recurrence following re-implantation is associated with significant morbidity (3,4).

Although various techniques can be used for the diagnosis of PJI, including preoperative laboratory tests, radiological examination, nuclear medicine detection, intraoperative culture and histopathology (5-8), no gold-standard test for the diagnosis of PJI has been established, and the limited sensitivity and specificity of the tests that are available make the differentiation between PJI and other causes of prosthetic failure, such as metal allergy or aseptic loosening, challenging (2,9). PJI continues to cause difficulties for orthopedic surgeons, particularly when the clinical signs and regular serum inflammatory markers are not fully indicative of infection (10).

With no single serological test available, doctors evaluate and diagnose PJI predominantly through a combination of symptom evaluation, physical examination and results of joint aspirates, even tissue samples (11). Gram staining (GS) is a widely used test that is easily available. GS is also commonly used in the diagnosis of PJI, particularly in developing countries, as a result of its low cost, rapid turnaround time and ease of use; however, its value remains controversial due to conflicting results on the effectiveness of GS in the diagnosis of PJI (5,12-28). The present meta-analysis was therefore performed as one of a series of meta-analyses (6,7,29) in order to evaluate the detection validity of GS for the diagnosis of PJI and to provide evidence-based advice to clinicians.

## Materials and methods

The protocol used in the present meta-analysis was conducted as described in our previous studies (6,7,29) and based on recommendations in the methodological guidelines for

*Correspondence to:* Dr Xinhua Qu or Professor Kerong Dai, Department of Orthopedics, Shanghai Key Laboratory of Orthopedic Implants, Shanghai Ninth People's Hospital, Shanghai Jiaotong University School of Medicine, 639 Zhizaoju Road, Shanghai 200001, P.R. China  
E-mail: xinhua\_qu@126.com  
E-mail: krdai@163.com

\*Contributed equally

**Key words:** prosthetic joint infection, Gram staining, meta-analysis, total hip arthroplasty, total knee arthroplasty

conducting systematic reviews studying diagnostic accuracy (30) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (31).

**Search strategy.** Searches of the MEDLINE, EMBASE and OVID databases were conducted for articles published between January 1990 and December 2013. All searches were performed using the medical subject headings 'joint prosthesis', 'prosthesis infection', 'septic loosening', 'aseptic loosening', 'replacement' and 'arthroplasty' and the free text words 'gram', 'stain', 'intra-operative' and 'synovial fluid'. The reference lists of eligible studies and review articles were also examined.

**Selection of studies.** The abstracts of the studies were read by two investigators, and a standardized data extraction form was used to identify potentially eligible articles. Subsequent full-text analysis determined whether the studies were eligible for inclusion. Disagreements were resolved by discussion with a third investigator.

The inclusion criteria for the analysis were as follows: i) Collection of data on GS in combination with an accurate diagnosis of PJI based on visible purulence in the joint aspirate or at the surgical site, evidence of communication between the prosthesis and a sinus tract (fistula), histopathological or periprosthetic tissue findings of acute inflammation, or microbiological cultures simultaneously obtained from at least two periprosthetic tissue samples (the reference standard); ii) studies with sufficient data to enable the true-positive (TP), false-negative (FN), false-positive (FP) and true-negative (TN) values to be determined; and iii) inclusion of  $\geq 10$  patients. Discrepancies were resolved through discussion with the other investigators and consultation of the original articles.

**Data extraction and assessment of study quality.** Relevant data regarding the study designs and results were extracted independently by two investigators using a standardized form. Blinding to the journal name, the authors' names and affiliations and the year of the study publication was not performed, as a previous study has shown such a step to be unnecessary (32). Disagreements between the reviewers were resolved through consultation with another reviewer, who evaluated all discrepancies. The opinion of the majority was used for the analysis. The methodological quality of the studies included in the meta-analysis was independently assessed by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool (33), which has been specifically developed for use in systematic reviews assessing diagnostic accuracy.

Validity analyses were performed through the use of a standardized form to extract the following items from each study: A description of the study participants, the authors' names, country in which the study was performed, number of patients and infected patients, mean age, study design, patient enrolment, sample type, surgical site, type of blinding conducted and characteristics of the reference standard used.

**Statistical analysis.** For each study, a two-by-two contingency table, consisting of TP, FP, FN and TN results according to the GS values and the reference standard, was constructed. The sensitivity [TP/(TP + FN)], specificity [TN/(FP + TN)] and the diagnostic odds ratios (DORs) [(TP x TN)/(FP x FN)]

were calculated. To evaluate the efficacy of GS assays in the diagnosis of PJI, the sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), DOR, post-test probability and area under the receiver operating characteristic curve (AUC) were calculated (34).

Heterogeneity was evaluated using the likelihood ratio  $I^2$  index and  $\chi^2$  tests (35). The  $I^2$  index demonstrates the percentage of total variation across the studies as a result of heterogeneity. An  $I^2$  value of  $>50\%$  indicates that there is more heterogeneity among the studies than would be expected solely by chance. For the likelihood ratio  $\chi^2$  test, heterogeneity among studies was shown by P-values of  $<0.05$ . When heterogeneity was observed, the primary meta-analysis was conducted using a random effects model to generate a summary estimate for the test sensitivity with 95% confidence intervals (CI). Meta-regression analyses were performed to evaluate potential heterogeneity, and a Deeks' funnel plot asymmetry test was utilized to assess potential publication bias (36). The different study characteristics, including infected arthroplasties, specimens processed, publication year, reference standard, study design, patient enrolment and type of blinding performed, were evaluated through subgroup analyses. All the statistical analyses were performed using STATA version 12 software (StataCorp, College Station, TX, USA).  $P < 0.05$  was considered to be significant.

## Results

**Study selection.** The database search yielded 287 primary studies. Of these, 238 were excluded following reviews of the title and abstract, and 33 were excluded following review of the full article. Two additional studies were obtained from one of manuscripts evaluated (12). In summary, 18 articles involving a total of 4,647 patients fulfilled all the inclusion criteria and were considered in the analysis (5,12-28) (Fig. 1). A general consensus was reached among the investigators with regard to the included studies (Cohen's unweighted  $\kappa = 0.93$ ).

**Study description and quality.** Eighteen studies in which GS was performed were identified; all of these studies met the eligibility criteria. Table I lists the included studies and describes the baseline patient characteristics. The studies were conducted in six different countries (12 in the United States, two in Canada and one study each in China, Finland, France and the United Kingdom). The median number of patients in each study was 169 (range, 33-1,004), and the median age of the participants was 66 years (range, 62-72.8 years). A total of six studies prospectively enrolled patients (5,12,21,22,24,25), and 12 studies were retrospective database reviews (13-20,23,26-28). Patient recruitment was consecutive in five studies (17,21,22,25,27) and was not documented in the remaining 13 studies (5,12-16,18-20,23,24,26,28). Ten studies detected PJI on the hip and knee (5,12,18,19-21,23,24,26,27); three, on the hip (16,22,28); three, on the knee (13,17,25); one, on the elbow (14); and two, on the shoulder (14,21). All the eligible studies scored  $>9$  points using the QUADAS quality assessment tool, indicating that they were of moderate quality.

**Diagnostic accuracy.** The pooled sensitivity, specificity, DOR and AUC values obtained from the random effects model are

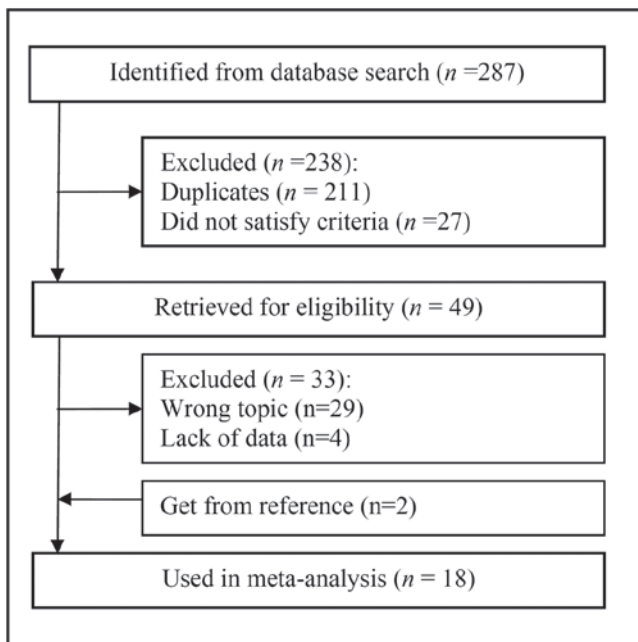


Figure 1. Flowchart for study selection.

shown in Fig. 2. The pooled sensitivity and specificity values for the detection of PJI using GS were 0.19 (95% CI, 0.12-0.27) and 1.00 (95% CI, 0.99-1.00), respectively. The pooled DOR for GS was 51 (95% CI, 18-140), and the pooled AUC was 0.89 (95% CI, 0.86-0.91). The inconsistency index indicated significant heterogeneity with respect to GS ( $I^2=75\%$ ,  $P<0.01$ ); as a result, meta-regression and subgroup analyses were performed to evaluate the potential sources of heterogeneity in the GS studies (Table II). The analyses of the sensitivity and specificity for the detection of PJI using GS indicated no effect with regard to the type of infected arthroplasty (knee versus hip versus knee plus hip), the publication year (prior to 2006 versus 2006 or later), the reference standard (simple versus multiple), the study design (perspective versus retrospective) or patient enrollment (consecutive versus not available). The sensitivities of GS for infected arthroplasty of the knee, hip and knee plus hip were 0.14 (95% CI, 0.08-0.23), 0.14 (95% CI, 0.09-0.20) and 0.19 (95% CI, 0.11-0.32), respectively, whereas the specificities were 1.00 (95% CI, 0.99-1.00), 0.99 (95% CI, 0.97-1.00) and 0.99 (95% CI, 0.98-1.00), respectively. The analysis also indicated that specimens from synovial fluid had a higher sensitivity than those from tissue swabs, tissue and synovial fluid plus tissue (0.30 vs. 0.14, 0.14 and 0.16, respectively;  $P<0.05$ ). Specimens from synovial fluid also yielded the highest DOR values (242 vs. 151, 22 and 36, respectively;  $P<0.05$ ). By contrast, the highest AUC of 0.96 (95% CI, 0.94-0.97) ( $P>0.05$ ) was noted in the tissue specimens. The application of blinding to a study affected the accuracy of the GS assay; low sensitivity (0.09 vs. 0.28,  $P<0.05$ ) and DOR (10 vs. 737,  $P<0.05$ ) were apparent when the pathologist was unaware of the clinical results.

**Evaluation of clinical utility.** The PLR and NLR of GS for the diagnosis of PJI were 41.6 (95% CI, 15.5-111.2) and 0.82 (95% CI, 0.75-0.89), respectively (Fig. 3). Likelihood ratios were used to simulate clinical scenarios using 25, 50 and 75% pre-test probabilities of PJI. Subsequent post-test probability

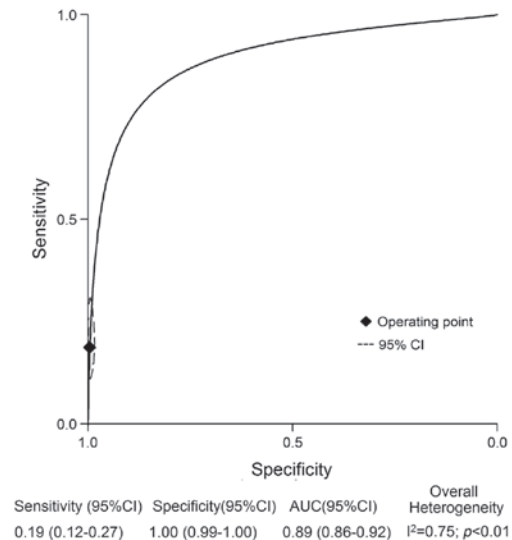


Figure 2. Summary receiver-operating characteristic curves for GS. Curves include a summary operating point for sensitivity and specificity on the curve and a 95% confidence contour ellipsoid. AUC, area under the receiver-operating characteristic curve; CI, confidence intervals.

was calculated and plotted on Fagan nomograms (Fig. 3). The post-test negative probability of PJI was 82% for the 25% pre-test probability, which could be considered sufficient to rule out PJI, and the post-test positive probability was 99% for the 75% pre-test probability, which could be considered sufficient for the diagnosis of PJI.

**Assessment of publication bias.** Potential publication bias was evaluated through the creation of Deeks' funnel plots by plotting the logDOR of the individual studies against their sample size. The funnel plots for GS are presented in Fig. 4. The regression test of asymmetry found evidence of a small-study effect for GS ( $P<0.01$ ).

## Discussion

In the present meta-analysis of 18 articles with a total of 4,647 patients, it was determined that GS could not be used alone for the diagnosis of PJI among patients who underwent THA or TKA, as the sensitivity and AUC, the NLR findings and the low clinical scenario-negative post-test probabilities demonstrated the poor clinical utility of GS to accurately diagnose PJI.

Although numerous preoperative and intraoperative tests have been employed, the diagnosis of PJI following THA or TKA remains a challenge. In contrast to aseptic loosening, the results of revision total joint arthroplasty of prosthetic infection can lead to high cost and even patient morbidity; despite this, none of the currently available tests have perfect sensitivity and specificity (1,2). Fluorodeoxyglucose positron emission tomography (FDG-PET) and antigranulocyte scintigraphy with  $^{99m}\text{Tc}$ -labeled monoclonal antibodies have been reported to be effective imaging modalities for PJI diagnosis, and two meta-analyses have demonstrated that the techniques exhibit acceptable diagnostic capability, with sensitivities of 0.82 and 0.83 and specificities of 0.87 and 0.80, respectively (37,38).

Table I. Characteristics of the 18 studies in a meta-analysis of the diagnosis of prosthetic joint infection using Gram staining.

First author, year (ref.)	Country	No. of patients	No of infections	Age (years)	Study design	Patient enrollment	Infected arthroplasties	Specimen	Blinding	Reference standard
Zywiel, 2011 (13)	United States	347	156	62	Retrospective	NA	Knee	Tissue swab	Yes	IOF, H, M
Schindler, 2011 (14)	France	62	48	68	Retrospective	NA	Knee, hip, shoulder, elbow	NA	NA	IOF, H, M
Oethinger, 2011 (15)	United States	269	84	NA	Retrospective	NA	NA	Synovial fluid and tissue	Yes	M
Johnson, 2010 (16)	United States	202	82	NA	Retrospective	NA	Hip	Tissue swab	NA	IOF, H, M
Morgan, 2009 (17)	United States	921	247	NA	Retrospective	Consecutive	Knee	Synovial fluid and tissue	NA	IOF, H, M, L
Ghanem, 2009 (18)	United States	1004	321	66	Retrospective	NA	Knee, hip	Tissue	NA	IOF, H, M
Trampuz, 2007 (5)	United States	331	79	68	Prospective	NA	Knee, hip	Sonicate fluid	NA	IOF, H, M
Parvizi, 2006 (12)	United States	70	39	68	Prospective	NA	Knee, hip	Synovial fluid and tissue	NA	IOF, M, L
Ko, 2005 (19)	China	40	9	72	Retrospective	NA	Knee, hip	Tissue	Yes	M
Banit, 2002 (21)	United States	121	21	64	Prospective	Consecutive	Knee, hip, shoulder	Tissue swab	NA	M
Viirolainen, 2002 (20)	Finland	68	21	72.8 <sup>a</sup> /65.1 <sup>b</sup>	Retrospective	NA	Knee, hip	Tissue	NA	M
Spanghehl, 1999 (22)	Canada	178	35	NA	Prospective	Consecutive	Hip	Tissue	Yes	IOF, H, M, L
Della Valle, 1999 (23)	United States	413	68	62.8	Retrospective	NA	Knee, hip	Tissue	Yes	IOF, H, M,
Atkins, 1998 (24)	United Kingdom	297	41	70	Prospective	NA	Knee, hip	Synovial fluid and tissue	Yes	H
Barrack, 1997 (25)	United States	67	20	65	Prospective	Consecutive	Knee	Synovial fluid	NA	H, M
Chimento, 1996 (26)	United States	169	32	NA	Retrospective	NA	Knee, hip	Tissue	Yes	H, M
Feldman, 1995 (27)	United States	33	9	62	Retrospective	Consecutive	Knee, hip	Synovial fluid	NA	M
Kraemer, 1993 (28)	Canada	55	13	60	Retrospective	NA	Hip	Tissue	NA	M

<sup>a</sup>Females; <sup>b</sup>males. H, histological examination; IOF, intraoperative finding; L, laboratory examination; M, microbiological; NA, not available.

Table II. Accuracy estimates from subgroup analyses.

Factor	No. of studies	No. of patients	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	DOR
<b>Infected arthroplasties</b>								
Knee and hip	9	2425	0.19 (0.11-0.32)	0.99 (0.98-1.00)	0.95(0.93-0.97)	38.1 (10.0-145.4)	0.81 (0.71-0.93)	47 (11-194)
Knee	3	1335	0.14 (0.08-0.23)	1.00 (0.99-1.00)	0.98 (0.96-0.99)	50.1 (12.6-198.6)	0.87 (0.79-0.95)	58 (14-244)
Hip	4	435	0.14 (0.09-0.20)	0.99 (0.97-1.00)	0.47 (0.43-0.52)	21.4 (4.8-95.1)	0.87 (0.81-0.92)	25 (5-112)
<b>Specimens processed</b>								
Tissue swab	3	670	0.14 (0.07-0.24)	1.00 (0.97-1.00)	0.86 (0.83-0.89)	130.9 (3.9-4391.4)	0.86 (0.78-0.95)	151 (4-5351)
Synovial fluid and tissue	3	1487	0.14 (0.10-0.20)	0.99 (0.99-1.00)	0.73 (0.69-0.77)	18.9 (7.7-45.9)	0.86 (0.82-0.91)	22 (9-56)
Synovial fluid	4	501	0.30 (0.17-0.48)	1.00 (0.88-1.00)	0.77 (0.73-0.80)	170.3 (2.1-13590.0)	0.70 (0.56-0.88)	242 (3-19836)
Tissue	5	1927	0.16 (0.08-0.29)	0.99 (0.98-1.00)	0.96 (0.94-0.97)	30.9 (7.4-129.7)	0.85 (0.75-0.96)	36 (8-163)
<b>Publication year</b>								
Prior to 2006	10	1441	0.14 (0.08-0.22)	0.99 (0.98-0.99)	0.99 (0.97-0.99)	13.6 (6.7-27.6)	0.87 (0.81-0.94)	16 (7-33)
2006 or later	8	3206	0.25 (0.15-0.39)	1.00 (0.98-1.00)	0.87 (0.84-0.89)	61.4 (15.6-241.8)	0.76 (0.64-0.89)	81 (20-334)
<b>Reference standard</b>								
Simple	7	883	0.15 (0.09-0.25)	0.99 (0.97-1.00)	0.83 (0.80-0.86)	21.9 (3.5-136.6)	0.85 (0.78-0.94)	26 (4-171)
Multiple	11	3764	0.20 (0.12-0.33)	1.00 (0.99-1.00)	0.87 (0.84-0.90)	55.2 (16.4-186.2)	0.80 (0.70-0.91)	69 (20-239)
<b>Blinding</b>								
NA	11	2934	0.28 (0.20-0.39)	1.00 (0.99-1.00)	0.73 (0.69-0.77)	528.4 (35.5-7859.4)	0.72 (0.63-0.82)	737 (52-10434)
Yes	7	1713	0.09 (0.06-0.14)	0.99 (0.98-0.99)	0.83 (0.79-0.86)	8.9 (4.6-17.2)	0.92 (0.88-0.96)	10 (5-19)
<b>Study design</b>								
Prospective	6	1064	0.26 (0.19-0.33)	0.99 (0.98-1.00)	0.89 (0.86-0.91)	34.6 (13.8-86.6)	0.75 (0.68-0.83)	46 (17-124)
Retrospective	12	3583	0.15 (0.08-0.26)	1.00 (0.95-1.00)	0.49 (0.44-0.53)	41.1 (3.3-506.0)	0.85 (0.77-0.95)	48 (4-599)
<b>Patient enrollment</b>								
Consecutive	5	1320	0.25 (0.18-0.32)	1.00 (0.98-1.00)	0.47 (0.43-0.51)	75.7 (11.8-485.8)	0.76 (0.69-0.84)	100 (15-679)
Not provided	13	3327	0.17 (0.10-0.28)	0.99 (0.98-1.00)	0.89 (0.86-0.91)	33.6 (11.0-102.3)	0.83 (0.75-0.93)	40 (13-127)
Overall studies	18	4647	0.19 (0.12-0.27)	1.00 (0.99-1.00)	0.89 (0.86-0.91)	41.6 (15.5-111.2)	0.82 (0.75-0.89)	51 (18-140)

AUC, area under the receiver operating characteristic curve; CI, confidence intervals; DOR, diagnostic odds ratio.



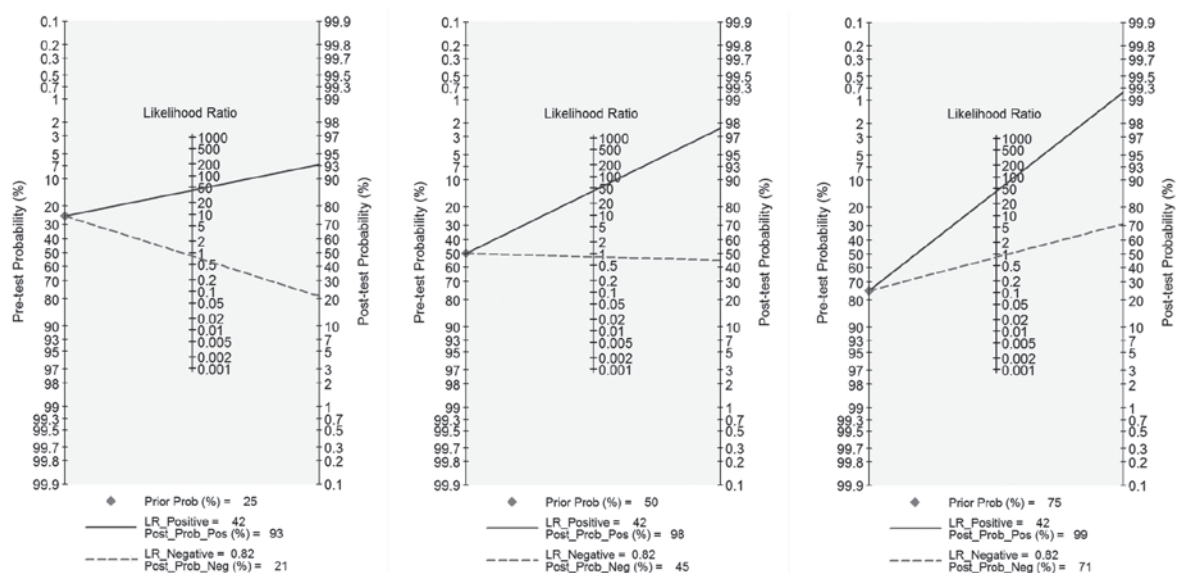


Figure 3. Pre-test probabilities and likelihood ratios for Gram staining. The post-test negative probability of PJI was 82% for the 25% pre-test probability, which could be considered sufficient to rule out PJI, and the post-test positive probability of PJI was 99% for the 75% pre-test probability, which could be considered sufficient for the diagnosis of PJI. PJI, prosthetic joint infection; LR, likelihood ratio.

There are, however, certain limitations to these diagnostic techniques for PJI diagnosis, including high expense, complex techniques and the requirement for skilled operators. Even the most common preoperative laboratory tests that are used for the diagnosis of PJI, such as white blood cell count (WBC), erythrocyte sedimentation rate (ESR) and serum C-reactive protein (CRP) levels (2,5,39), demonstrate low suitability for the diagnosis of PJI (40). A meta-analysis revealed that the accuracy of inflammation markers as indicators of PJI, as represented by DORs, was 13.1 for CRP, 7.2 for ESR and 4.4 for WBC (40).

GS, first described by Hans Christian Gram in 1884, has been widely used for the evaluation of bacterial infection through the results of staining (41). Whether the result is positive or negative depends on the morphological characteristics of the bacterium, such as the bacterial peptidoglycan layer and the outer membrane. GS is still widely used by laboratories and clinics all over the world, although there has been a significant development in diagnostic technology. In bacterial pneumonia, sepsis and bacteriuria, for example, the sensitivities and specificities are >90% (42,43). In the diagnosis of PJI, particularly in developing countries, GS is commonly used, possibly due to certain desirable characteristics, including the rapid turnaround time, convenience and cost-effectiveness; however, a number of studies have reported sensitivities of between 0 and 50%, which questions the application of GS in the diagnosis of PJI (17,18,23). Despite the low and variable sensitivity, numerous institutions continue to perform GS on all preoperative aspiration samples or intraoperative wound swab samples sent for microbiological analysis.

Consistent with the aforementioned reports (17,18,23), guidelines by the Infectious Diseases Society of America (IDSA) have recommended against using GS for the assessment of PJI (39,44); however, several factors can influence the final results, such as different criteria used to define a true infection and specimens from different sites. More than

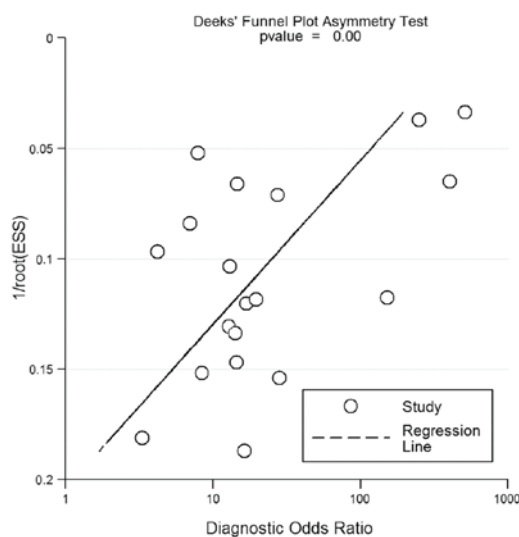


Figure 4. Meta-regression analyses of the sensitivity and specificity of GS. GS, Gram-staining; ESS, effective sample size.

three positive criteria for the diagnosis of PJI were reported by Morgan *et al* (17), in which a number of true infections could be missed. In comparison, only one positive intraoperative culture was sufficient in the study investigating the diagnosis of PJI following TKA by Banit *et al* (21), and the sensitivity was found to be 44%, which is relatively high compared with other reports (5,12-14,17,19,25-27). In the present meta-analysis, no difference was found in the sensitivity or specificity for the detection of PJI using GS between THA or TKA, publication year, reference standard, study design or patient enrolment; however, the type of specimen screened and whether blind analysis was performed or not did have an effect. Consistent with the IDSA guidelines (44) and a number of reports, these

results confirm that GS is a diagnostic method with low sensitivity and NLR for the diagnosis of PJI (5,12-28).

Notably, the low sensitivity of intraoperative GS has led to general discouragement regarding the use of the test for revision arthroplasty; however, positive findings are generally believed to have relatively high specificity. In the current meta-analysis it is therefore suggested that GS could be of value to help identify an organism to guide early antibiotic treatment in cases of re-implantation with a preoperative diagnosis of Gram-positive bacterial infection or gross purulence. In addition, GS may be useful when used as an adjuvant tool.

There are a number of limitations to the present study. Firstly, no established gold standard exists for the diagnosis of PJI. In the current meta-analysis, several reference standards were utilized among the studies, including clinical manifestation (purulence or fistula), laboratory studies (acute inflammation on histopathological examination or in blood tests) and microbiological growth (in periprosthetic tissues or sonication fluid culture). None of these techniques can be considered to be an optimal reference standard for the diagnosis of PJI, and misclassification bias, occurring due to an sub-optimal reference standard, may influence the predicted diagnostic accuracy of a tested method (37), generally resulting in an underestimation of the diagnostic accuracy. A second limitation in the current analysis was that the summary GS results exhibited high levels of statistical heterogeneity. This fact may diminish the strength of the conclusions that can be extracted from this meta-analysis. Thirdly, it was not clear in all of the studies whether a prospective study design was used, although the inclusion of a prospective study design, such as a covariate, compared with a bivariate model (prospective versus retrospective design) was not shown to significantly affect sensitivity or specificity. Finally, only a small number of studies mentioned the administration of antibiotics or the duration between the GS assessment and the confirmation of PJI; these factors may have had an effect on diagnostic accuracy, as antibiotic administration can increase the number of false negative results.

In combination, the results of this diagnostic accuracy meta-analysis indicate that GS in association with revision arthroplasty has low sensitivity, and that GS is therefore a poor choice for the diagnosis of PJI following knee and hip arthroplasty. Based on these data, we recommend that GS as a microbiological analysis should no longer be performed on the wound samples obtained when PJI is suspected.

## Acknowledgements

This study was supported by the Fund for Key National Basic Research Program of China (grant no. 2012CB619101), a major basic research grant from the Science and Technology Commission of Shanghai Municipality (grant no. 11DJ1400303), a scientific research grant from the National Natural Science Foundation for the Youth of China (grant no. 81201364), a scientific research grant for the Youth of Shanghai (grant no. ZZjdyx 2097), a scientific research grant from Zhejiang National Science Foundation (grant no. Y2110653) and an innovative research grant from the Shanghai Municipal Education Commission (grant no. 13YZ031).

## References

1. Clohisy JC, Calvert G, Tull F, McDonald D and Maloney WJ: Reasons for revision hip surgery: a retrospective review. *Clin Orthop Relat Res* 188-192, 2004.
2. Del Pozo JL and Patel R: Clinical practice. Infection associated with prosthetic joints. *N Engl J Med* 361: 787-794, 2009.
3. Hanssen AD, Trousdale RT and Osmon DR: Patient outcome with reinfection following reimplantation for the infected total knee arthroplasty. *Clin Orthop Relat Res* 55-67, 1995.
4. Mont MA, Waldman BJ and Hungerford DS: Evaluation of preoperative cultures before second-stage reimplantation of a total knee prosthesis complicated by infection. A comparison-group study. *J Bone Joint Surg Am* 82-A: 1552-1557, 2000.
5. Trampuz A, Piper KE, Jacobson MJ, *et al*: Sonication of removed hip and knee prostheses for diagnosis of infection. *N Engl J Med* 357: 654-663, 2007.
6. Qu X, Zhai Z, Wu C, *et al*: Preoperative aspiration culture for preoperative diagnosis of infection in total hip or knee arthroplasty. *J Clin Microbiol* 51: 3830-3834, 2013.
7. Qu X, Zhai Z, Li H, *et al*: PCR-based diagnosis of prosthetic joint infection. *J Clin Microbiol* 51: 2742-2746, 2013.
8. Qu X, Zhai Z, Liu X, Li H, Wu C, Li Y, Li H, Zhu Z, Qin A and Dai K: Evaluation of white cell count and differential in synovial fluid for diagnosing infections after total hip or knee arthroplasty. *PLoS One* 9: e84751, 2014.
9. Patel R, Osmon DR and Hanssen AD: The diagnosis of prosthetic joint infection: current techniques and emerging technologies. *Clin Orthop Relat Res* 55-58, 2005.
10. Kraay MJ, Goldberg VM, Fitzgerald SJ and Salata MJ: Cementless two-staged total hip arthroplasty for deep periprosthetic infection. *Clin Orthop Relat Res* 441: 243-249, 2005.
11. Leone JM and Hanssen AD: Management of infection at the site of a total knee arthroplasty. *J Bone Joint Surg Am* 87: 2335-2348, 2005.
12. Parvizi J, Ghanem E, Menashe S, Barrack RL and Bauer TW: Periprosthetic infection: what are the diagnostic challenges? *J Bone Joint Surg Am* 88 Suppl 4: 138-147, 2006.
13. Zywił MG, Stroh DA, Johnson AJ, Marker DR and Mont MA: Gram stains have limited application in the diagnosis of infected total knee arthroplasty. *Int J Infect Dis* 15: e702-e705, 2011.
14. Schindler M, Christofilopoulos P, Wyssa B, *et al*: Poor performance of microbiological sampling in the prediction of recurrent arthroplasty infection. *Int Orthop* 35: 647-654, 2011.
15. Oethinger M, Warner DK, Schindler SA, Kobayashi H and Bauer TW: Diagnosing periprosthetic infection: false-positive intraoperative Gram stains. *Clin Orthop Relat Res* 469: 954-960, 2011.
16. Johnson AJ, Zywił MG, Stroh DA, Marker DR and Mont MA: Should gram stains have a role in diagnosing hip arthroplasty infections? *Clin Orthop Relat Res* 468: 2387-2391, 2010.
17. Morgan PM, Sharkey P, Ghanem E, *et al*: The value of intraoperative Gram stain in revision total knee arthroplasty. *J Bone Joint Surg Am* 91: 2124-2129, 2009.
18. Ghanem E, Ketonis C, Restrepo C, Joshi A, Barrack R and Parvizi J: Periprosthetic infection: where do we stand with regard to Gram stain? *Acta Orthop* 80: 37-40, 2009.
19. Ko PS, Ip D, Chow KP, Cheung F, Lee OB and Lam JJ: The role of intraoperative frozen section in decision making in revision hip and knee arthroplasties in a local community hospital. *J Arthroplasty* 20: 189-195, 2005.
20. Virolainen P, Läähtenmäki H, Hiltunen A, Sipola E, Meurman O and Nelimarkka O: The reliability of diagnosis of infection during revision arthroplasties. *Scand J Surg* 91: 178-181, 2002.
21. Banit DM, Kaufer H and Hartford JM: Intraoperative frozen section analysis in revision total joint arthroplasty. *Clin Orthop Relat Res* 230-238, 2002.
22. Spanghel MJ, Masterson E, Masri BA, O'Connell JX and Duncan CP: The role of intraoperative gram stain in the diagnosis of infection during revision total hip arthroplasty. *J Arthroplasty* 14: 952-956, 1999.
23. Della Valle CJ, Scher DM, Kim YH, *et al*: The role of intraoperative Gram stain in revision total joint arthroplasty. *J Arthroplasty* 14: 500-504, 1999.
24. Atkins BL, Athanasou N, Deeks JJ, *et al*: Prospective evaluation of criteria for microbiological diagnosis of prosthetic-joint infection at revision arthroplasty. The OSIRIS Collaborative Study Group. *J Clin Microbiol* 36: 2932-2939, 1998.

25. Barrack RL, Jennings RW, Wolfe MW and Bertot AJ: The Coventry Award. The value of preoperative aspiration before total knee revision. *Clin Orthop Relat Res*: 8-16, 1997.
26. Chimento GF, Finger S and Barrack RL: Gram stain detection of infection during revision arthroplasty. *J Bone Joint Surg Br* 78: 838-839, 1996.
27. Feldman DS, Lonner JH, Desai P and Zuckerman JD: The role of intraoperative frozen sections in revision total joint arthroplasty. *J Bone Joint Surg Am* 77: 1807-1813, 1995.
28. Kraemer WJ, Saplys R, Waddell JP and Morton J: Bone scan, gallium scan and hip aspiration in the diagnosis of infected total hip arthroplasty. *J Arthroplasty* 8: 611-616, 1993.
29. Qu X, Huang X, Yan W, Wu L and Dai K: A meta-analysis of <sup>18</sup>F-FDG-PET-CT, <sup>18</sup>F-FDG-PET, MRI and bone scintigraphy for diagnosis of bone metastases in patients with lung cancer. *Eur J Radiol* 81: 1007-1015, 2012.
30. Devillé WL, Buntinx F, Bouter LM, *et al*: Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2: 9, 2002.
31. Liberati A, Altman DG, Tetzlaff J, *et al*: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 151: W65-W94, 2009.
32. Berlin JA: Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. *Lancet* 350: 185-186, 1997.
33. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM and Kleijnen J: The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 3: 25, 2003.
34. Moses LE, Shapiro D and Littenberg B: Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 12: 1293-1316, 1993.
35. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F and Botella J: Assessing heterogeneity in meta-analysis: Q statistic or I<sup>2</sup> index? *Psychol Methods* 11: 193-206, 2006.
36. Deeks JJ, Macaskill P and Irwig L: The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 58: 882-893, 2005.
37. Pakos EE, Trikalinos TA, Fotopoulos AD and Ioannidis JP: Prosthesis infection: diagnosis after total joint arthroplasty with antigranulocyte scintigraphy with <sup>99m</sup>Tc-labeled monoclonal antibodies - a meta-analysis. *Radiology* 242: 101-108, 2007.
38. Kwee TC, Kwee RM and Alavi A: FDG-PET for diagnosing prosthetic joint infection: systematic review and metaanalysis. *Eur J Nucl Med Mol Imaging* 35: 2122-2132, 2008.
39. Parvizi J, Adeli B, Zmistowski B, Restrepo C and Greenwald AS: Management of periprosthetic joint infection: the current knowledge: AAOS exhibit selection. *J Bone Joint Surg Am* 94: e104, 2012.
40. Berbari E, Mabry T, Tsaras G, *et al*: Inflammatory blood laboratory levels as markers of prosthetic joint infection: a systematic review and meta-analysis. *J Bone Joint Surg Am* 92: 2102-2109, 2010.
41. Popescu A and Doyle RJ: The Gram stain after more than a century. *Biotech Histochem* 71: 145-151, 1996.
42. Anevlavis S, Petroglou N, Tzavaras A, *et al*: A prospective study of the diagnostic utility of sputum Gram stain in pneumonia. *J Infect* 59: 83-89, 2009.
43. Søgaard M, Nørgaard M and Schønheyder HC: First notification of positive blood cultures and the high accuracy of the gram stain report. *J Clin Microbiol* 45: 1113-1117, 2007.
44. Osmon DR, Berbari EF, Berendt AR, *et al*: Infectious Diseases Society of America: Executive summary: diagnosis and management of prosthetic joint infection: clinical practice guidelines by the Infectious Diseases Society of America. *Clin Infect Dis* 56: 1-10, 2013.