Consensus on the definition and criteria for failure of surgical treatment in bacterial arthritis of a native joint: An international Delphi study

Alex B. Walinga[1,2,3,4] | Stein J. Janssen[1,2,3,4] | Arthur J. Kievit[1,2,3,4] | The International Panel of Clinical Experts | Corianne A. J. M. de Borgie[5] | Gino M. M. J. Kerkhoffs[1,2,3,4]

[1]Department of Orthopedic Surgery and Sports Medicine, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

[2]Amsterdam Movement Sciences, Musculoskeletal Health, Sport, Amsterdam, The Netherlands

[3]Academic Center for Evidence-Based Sports Medicine (ACES), Amsterdam, The Netherlands

[4]Amsterdam Collaboration for Health & Safety in Sports (ACHSS), IOC Research Center, Amsterdam, The Netherlands

[5]Department of Epidemiology and Data Science, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

Correspondence
Alex B. Walinga and Gino M. M. J. Kerkhoffs, Department of Orthopedic Surgery and Sports Medicine, Amsterdam University Medical Center, University of Amsterdam, Meibergdreef 9, 1105AZ, Amsterdam, The Netherlands.
Email: a.b.walinga@amsterdamumc.nl and g.m.kerkhoffs@amsterdamumc.nl

Funding information
None

Abstract

Purpose: The literature presents a wide range of success rates for a single surgical intervention of bacterial-septic-arthritis, and there is a lack of clear criteria for identifying treatment failure and making decisions about reintervention. This Delphi study aims to establish a consensus among an international panel of experts regarding the definition of treatment failure and the criteria for reintervention in case of bacterial arthritis.

Methods: The conducting and reporting Delphi studies (CREDES) criteria were used. Data from a systematic review was provided as the basis for the study. A list of 100 potential experts were identified. The study was designed and conducted as follows: (I) identification and invitation of an expert panel, (II) informing the participating expert panel on the research question and subject, and (III) conducting two or three Delphi rounds to reach consensus on explicit research items. Potential criteria were rated on a five-point Likert scale.

Results: Sixty orthopaedic experts from nine countries participated in this Delphi study, with 55 completing all three rounds. The mean experience as an orthopaedic surgeon was 15 years (SD ± 9). Strong (96%) consensus was reached on the definition of treatment failure: the persistence of physical signs of arthritis (e.g., pain and swelling) and/or systemic inflammation (e.g., fever and no improvement in CRP) despite surgical and antibiotic treatment. Furthermore, consensus (>80%) was reached on six criteria influencing the decision for reintervention; pain (81%), sepsis (98%), fever (88%), serum CRP (93%), blood culture (82%), and synovial fluid culture (84%).

Conclusion: The definition of treatment failure for bacterial arthritis after a single surgical intervention was established through a three-round Delphi study. Additionally, consensus was reached on six criteria that are helpful for determining the need for reintervention. This definition and these criteria

may help in the development of clinical guidelines, and will empower physicians to make more precise and consistent decisions regarding reintervention for patients, ultimately aiming to reduce over- and under-treatment and improve patient outcomes.

Level of Evidence: Level V.

KEYWORDS
bacterial arthritis, consensus, criteria, delphi study, failure, reintervention

INTRODUCTION

Bacterial arthritis of a native joint is a severe and potentially debilitating condition that requires timely and effective treatment. Surgical intervention and systemic antibiotics are the first-line treatment for bacterial arthritis [1]. This approach is effective for the majority of patients. The success rate of a single surgical intervention, however, varies widely in the literature, with reported reintervention rates ranging from 7% to 72% and a pooled failure rate of 26% [2]. This variation in failure rates might be attributed to the lack of a clear definition of treatment failure and when to decide on reintervention. If a single surgical intervention fails, some patients may require multiple reinterventions to eradicate the infection. The decision to perform and time these reinterventions can be challenging and depends on various factors, such as the severity of the infection, the presence of comorbidities, and the individual patient's response to treatment. Clinicians must carefully weigh the risks and benefits of additional interventions and use objective criteria to determine whether a reintervention is necessary or not. There is a need for clearer clinical guidelines regarding treatment failure and the decision to proceed with further interventions. This could help minimise over- and under-treatment.

Consequently, this Delphi study was initiated with the aim of reaching consensus among an international panel of clinical experts on (1) a clear definition of treatment failure for bacterial arthritis of a native joint following a single surgical intervention and (2) criteria that should be considered when deciding on the need for reintervention after a single surgical treatment.

MATERIALS AND METHODS

The Medical Ethical Committee of Amsterdam UMC, location AMC, granted approval for this study and determined that a formal ethical evaluation under the Medical Research Involving Human Subjects Act (WMO) was not required (reference number: W22_238 #22.293).

The Delphi method is commonly used in medical research to reach consensus on a topic with the aid of an expert panel. For this Delphi study, the recommendations for methodologic criteria and reporting for Delphi studies by Diamond et al. [3] and Jünger et al. [4] were followed. According to these 'conducting and reporting Delphi studies (CREDES)' criteria, reported by Jünger et al. [4], the study was designed and conducted as follows: (I) identification and invitation of an expert panel, (II) informing the participating expert panel on the research question and subject, and (III) conducting two or three Delphi rounds to reach consensus on explicit research items. Additionally, a steering committee was established: three orthopaedic surgeons with varying degrees of expertise in the treatment of bacterial arthritis (Gino M. M. J. Kerkhoffs, Stein J. Janssen and Arthur J. Kievit) along with an epidemiologist (Corianne A. J. M. de Borgie), and a clinical researcher (Alex B. Walinga). All the above-mentioned authors played pivotal roles in the design of this Delphi study rounds and its subsequent analyses.

Evidence review

A systematic literature search was performed up to January 2020, using the keywords: 'bacterial arthritis', native joint', and 'treatment failure' including synonyms and MeSH terms [2]. This resulted in 30 relevant studies that were utilised to create an overview of (1) the overall failure rate after surgical treatment for native joint bacterial arthritis, (2) prognostic factors associated with failure, and (3) the heterogeneity of definitions of failure among studies. This data was used as a basis for the development of the current Delphi study [2].

Panel and group selection methodology

First, a group of experts were invited in the field of orthopaedic surgery, specifically in treating bacterial arthritis. These experts were identified: (1) through the clinical network of co-authors, (2) through the study group who agreed to participate, and (3) by identifying clinicians through high-quality papers in the field of bacterial arthritis. This was done to ensure a representative, international and heterogeneous group. Patients were not involved in the design of this Delphi study.

Currently, there are no guidelines or recommendations on the optimal sample size for Delphi studies, nor are (too) small or large sample sizes clearly defined. Some authors believe that a sample size of 10–15 experts is sufficient for homogeneous participant groups, a larger sample size is necessary if various reference groups are involved. The steering committee has determined that a minimum of 20 evaluable respondents will be required for this study [5]. A list of 100 potential experts was identified, and the aim was to have at least one in five of them participate.

Consensus process

This Delphi study aimed to organise two separate rounds to reach consensus on (1) the definition of failure of a single surgical treatment of bacterial arthritis and (2) what criteria (e.g., physical or systemic signs) play a role in deciding on reintervention after a single surgical treatment for bacterial arthritis. Potential criteria were rated on a five-point Likert scale (1—*being not important* and 5—*being very important*). The survey was sent using Castor Electronic Data Capture software, which facilitates distribution of online questionnaires and data collection (Castor EDC). Consensus was defined as ≥80% agreement (i.e., ≥80% rated the item as either 'not important' [Likert score 1–3] or 'important' [Likert score 4–5]) among experts, whereas strong consensus was defined as 90%–99% agreement and a unanimous consensus as 100%. A third round was only conducted when there was no consensus on the definition of failure and/or when there was no consensus on at least five criteria that could determine whether a reintervention is needed. As recommended per guidelines—to avoid an infinite number of rounds—, an intentional maximum of three rounds was planned, regardless of whether or not consensus was reached. All nonresponders received an email reminder after 1 week, after 2 weeks, and once more after 3 weeks per round.

Potential experts were contacted by email to explain the study's purpose, provide background information, and asked if they wanted to participate.

Three weeks after the initial invitation, the first information and questionnaire were sent to all experts. In this round, background information regarding the study was provided in greater detail. This was followed by an online survey consisting of three parts; (1) gathering baseline characteristics (sex, age, country, hospital type and number of arthroscopies per year); (2) asking to provide a personal description of what they deemed to be the definition of failure of surgical treatment in patients with bacterial arthritis (free text, limited to 500 characters); (3) rating the usefulness of 18 primary clinical criteria in determining the need for surgical reintervention on a five-point Likert scale (Table 1). Experts were given the option to provide extra criteria through a free text option. Four weeks after the initial invitation of the first round, the second questionnaire was sent to all experts who completed the first round.

This second round consisted of an online survey comprising two parts: Part 1 involved the definition of surgical treatment failure, which was divided into three sections:

1. Experts were presented with a list of the three most commonly used definitions of surgical treatment failure from the first round. These definitions were re-categorised during a consensus meeting by three members of the Delphi steering committee (Alex B. Walinga, Stein J. Janssen and Arthur J. Kievit). Experts were asked to select one of these definitions or indicate that all three were insufficient.

TABLE 1 Clinical criteria were rated on a five-point Likert scale.

Physical signs	Systemic signs	Synovial fluid	Other
Pain	Fever	Aspect of a new synovial fluid arthrocentesis	Comorbidities
Range of motion	Sepsis	Synovial fluid culture	Abnormal imaging
Redness	Serum CRP	Synovial fluid WBC	Gout[a]
Joint temperature	Serum ESR		
Position of the joint[a]	Serum WBC		
Joint effusion	Serum procalcitonin[a]		
Heart rate	Serum D-dimer[a]		
Respiratory rate	Serum alpha defensin[a]		
	Blood culture		

Abbreviations: CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; WBC, white blood cell.
[a]Criteria that were suggested by the participants in the first round, and were added to the list for the second.

2. Experts were asked whether they believed that the 'requirement for reintervention' should be included in the definition.
3. Experts were also asked to select the earliest possible timeframe in days for identifying failure of the intervention.

Part 2 involved rating the usefulness of specific criteria in determining the need for surgical reintervention. This part followed the same design as the first round but focused on criteria that did not reach consensus previously. Furthermore, experts were asked to rate the new criteria (i.e., position of the joint, serum procalcitonin, serum D-dimer, serum Alpha Defensin, gout) suggested by experts in the first round.

Lastly, a third round will only be conducted if there is no consensus (<80%) on the definition of surgical treatment failure and/or if there is no consensus on a list of at least five criteria. The Delphi study ends 3 weeks after the initial invitation of the third round, regardless of whether consensus was reached.

Data collection and analysis

Baseline demographic data were presented as mean with standard deviation (SD) or median and interquartile range according to their distribution. Categorical variables were expressed as absolute numbers with percentages. Data were analysed using Stata 15.0 (StataCorp LP) and Excel 16.0 (Microsoft Corporation). The Fisher exact test was used to assess differences in criteria rating based on country of origin (Netherlands vs. non-Netherlands), and based on type of practice (academic vs. nonacademic). A two-tailed p-value below 0.05 was considered statistically significant for all tests.

RESULTS

Characteristics of the panel

A total of 100 orthopaedic surgeons received an invitation email to participate in the Delphi study, scheduled from 1 December 2022, to 31 March 2023. Of those contacted, 60 orthopaedic expert (60%) agreed to participate. Fifty-five experts (92%) completed all rounds (Figure 1). The mean experience as an orthopaedic surgeon was 15 years (SD ± 8.6); they perform a median of 50 (interquartile range [IQR]: 25–120) arthroscopies (regardless of indication) per year. Half of the experts work in an academic hospital (47%, $n = 26$) (Table 2).
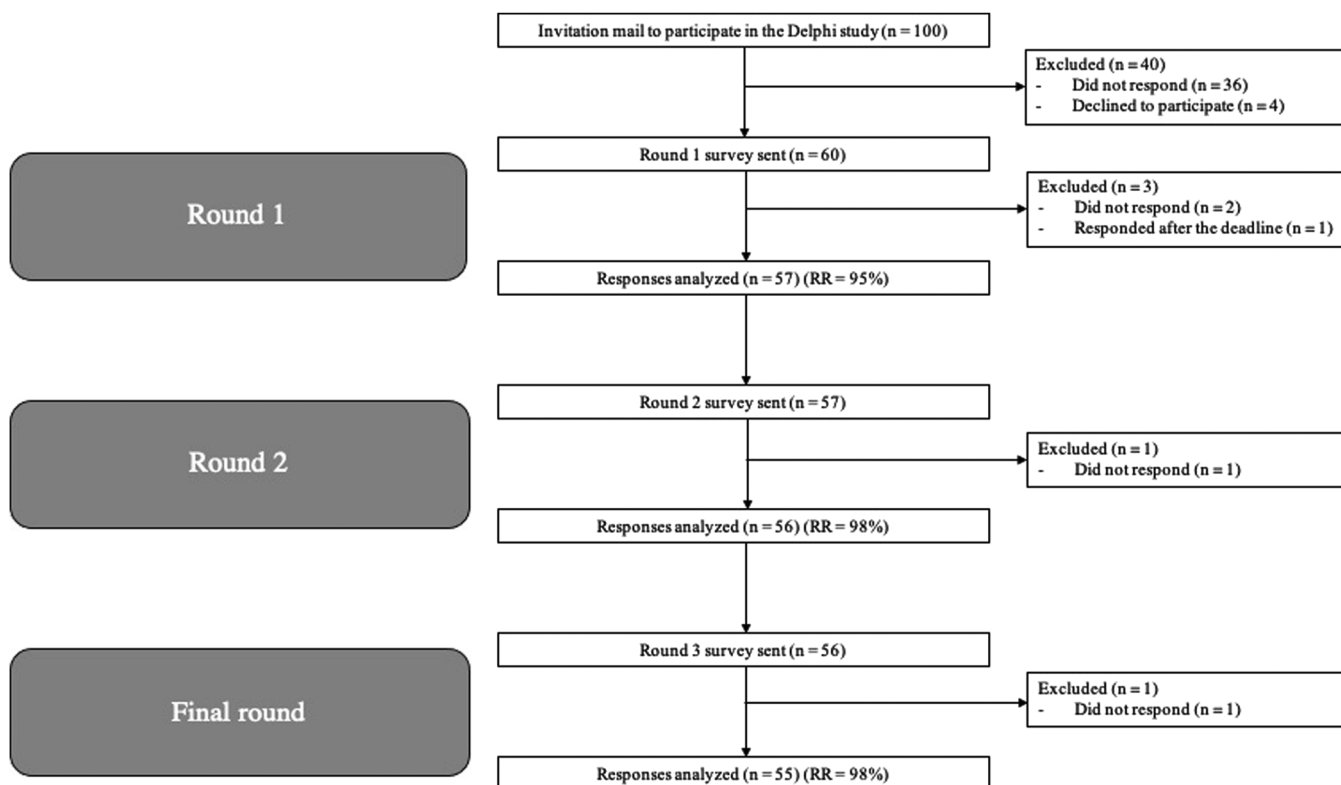

FIGURE 1 Flow diagram of experts and response rate (RR).

TABLE 2 Baseline characteristics of participating experts (*n* = 55).

Age, mean (SD)	48 (9)
Male, *n* (%)	52 (95)
Academic hospital, *n* (%)	26 (47)
Years of experience, mean (SD)	15 (9)
Arthroscopies per year, median (IQR)	50 (25–120)
Countries, *n* (%)	
Netherlands	43 (78)
Sweden	3 (5)
Italy	2 (4)
Spain	1 (2)
Portugal	1 (2)
Switzerland	1 (2)
Poland	1 (2)
China	1 (2)
Qatar	1 (2)

Abbreviations: IQR, interquartile range; n, number; SD, standard deviation.

Evidence review

Our literature review demonstrated an overall pooled failure rate of 26% (30 studies, 8586 native joints) with no difference (n.s.) between failure after arthroscopy (26%) versus arthrotomy (24%) [2]. The most commonly used definition of failure was based on (a combination of) clinical findings, elevated laboratory signs of systemic inflammation, and/or positive/purulent fluid analysis, requiring reoperation. The observed high heterogeneity in failure rates may be attributed to the variation in definitions used [2].

Delphi round 1

Fifty-seven out of 60 experts completed the first round and gave a definition of failure of surgical treatment in patients with bacterial arthritis of a native joint.

Strong consensus (≥90%) was reached on two criteria that play a role in deciding on reintervention after a single surgical treatment for bacterial arthritis; sepsis (98%) and serum CRP (93%). Consensus was also reached on: pain (81%) and synovial fluid culture (84%) (Table 3). Five new criteria were suggested by the experts, and were added to the list for the second round; position of the joint, gout, D-dimer, procalcitonin, and Alpha Defensin test.

TABLE 3 The criteria for which consensus was reached.

Strong consensus (90%–99%)	Consensus (80%–89%)
Sepsis	Pain
Serum CRP	Fever
	Blood culture
	Synovial fluid culture

Abbreviation: CRP, C-reactive protein.

TABLE 4 Top three re-categorised definitions for failure suggested by the participants in the first round.

Persisting physical signs of arthritis (e.g., pain and swelling) and systemic inflammation (e.g., fever and no improvement in CRP) despite surgical and antibiotic treatment.
Persisting physical signs of arthritis (e.g., pain and swelling) despite surgical and antibiotic treatment.
Persisting systemic inflammation (e.g., fever and no improvement in CRP) despite surgical and antibiotic treatment.

Abbreviation: CRP, C-reactive protein.

Delphi round 2

Fifty-six out of 57 experts completed the second round. Based on initial definitions (Supporting Information S1: Appendix A), the options were narrowed to three (Table 4). Seventy-seven percent agreed on the first definition, 5% on the second, and 9% on the third. However, 9% found all three definitions insufficient. With an agreement below 80%, no consensus was reached according to predetermined criteria.

In the first round, multiple experts mentioned the requirement for reintervention in their definitions. Based on their input, it was further inquired whether the experts believed that the words 'requirement for reintervention' should be included in the definition, to which 68% of the experts disagreed.

Moreover, the earliest possible timeframe in days was inquired for identifying failure of intervention; responses varied substantially with 2% selecting '1 day', 14% selecting '2 days', 30% selecting '3 days', 11% selecting '4 days', and 43% selecting '5 or more days'.

Due to a predetermined number of criteria (≥5) that were needed for a consensus, the criteria list (*n* = 19) was re-rated during round 2, including five criteria suggested by the experts. Consensus was reached on two additional criteria: fever (88%) and blood culture (82%). Two criteria were deemed not important based on consensus: D-dimer (93%) and procalcitonin (80%). With six criteria of consensus, no further re-rating occurred in round 3 (see Table 3).

TABLE 5 The definition with strong consensus (>90%).

Persisting physical signs of arthritis (e.g., pain and swelling) and/or systemic inflammation (e.g., fever and no improvement in CRP) despite surgical and antibiotic treatment.

Abbreviation: CRP, C-reactive protein.

Delphi round 3

Fifty-five out of 56 experts completed the third round. Based on the results and comments of the experts on the chosen definitions of the second round, the definition was modified to: 'Persisting physical signs of arthritis (e.g., pain and swelling) and/or systemic inflammation (e.g., fever and no improvement in CRP) despite surgical and antibiotic treatment'. Experts were asked to agree (yes or no). A strong consensus (96%) was reached on this definition (Table 5).

Moreover, there was no consensus yet on whether to include 'requirement for reintervention' in the aforementioned definition, with 64% disagreement. Also, no consensus was reached on specifying the earliest timeframe for failure identification: 0% for 'less than 2 days', 20% for '2–4 days', 44% for 'more than 4 days', and 36% selecting 'it is hard to say and therefore not applicable in the definition'.

Areas of disagreement

No consensus was reached on 17 of the 23 criteria and these were therefore not considered helpful in clinical practice in deciding whether a surgical reintervention is needed.

In round 3, 4% disagreed with the proposed modified definition. They remarked; (1) requiring additional surgical treatment has to be added to the definition and (2) the and/or option is confusing. In response to comment 1, requiring additional surgical treatment was not added because 64% of the experts already disagreed with this in the third round. In response to comment 2, due to the different comments of the experts in round two, the and/or was added in the modified definition of round three.

Thirteen experts commented on the reintervention timeframe, on which no consensus was reached; they noted that it depends on various factors: infection severity, bacteria virulence, and antibiotics effectiveness. However, unanimous consensus was reached that reintervention should not occur before 2 days post initial surgery.

Subanalyses based on country of origin and type of practice

Experts from nine countries participated across three rounds. The majority (78%) were from the Netherlands, while 22% represented other countries (Table 2). A subanalysis compared criteria ratings between experts from the Netherlands and experts from other countries. Differences were found in the first round for the joint temperature ($p = 0.024$) and synovial white blood cell count (WBC) ($p = 0.020$). In the second round, differences were found for joint temperature ($p = 0.003$), comorbidities ($p = 0.023$), and serum procalcitonin ($p = 0.001$). No differences were identified in the remaining criteria (n.s.). Experts from other countries assigned a higher degree of importance to joint temperature, synovial WBC levels, comorbidities, and serum procalcitonin compared to experts from the Netherlands.

Experts were also categorised by practice: 47% in academic hospitals, 51% in nonacademic hospitals, and 2% in both (analysed as academic vs. nonacademic). A subanalysis compared academic ($n = 26$) and non-academic experts ($n = 28$). No difference in first-round criteria was observed (n.s.). In the second round, difference was found in serum procalcitonin ($p = 0.018$), with no difference in the remaining criteria (n.s.). Academic experts emphasised higher importance on serum procalcitonin compared to nonacademic experts.

DISCUSSION

The most important finding of the present study was the strong consensus on the definition of treatment failure for bacterial arthritis of a native joint after a single surgical intervention: the persistence of physical signs of arthritis (e.g., pain and swelling) and/or systemic inflammation (e.g., fever and no improvement in CRP) despite surgical and antibiotic treatment. Additionally, consensus on six specific criteria (pain, fever, sepsis, serum CRP, blood culture, and synovial fluid culture) was also established to determine the need for reintervention after surgery.

Clinical implications

Patients with bacterial arthritis may require reinterventions, with a reported failure rate of approximately 26% in the literature [2]. Diagnosis is challenging due to the lack of clear standards and unreliable tests, making it even more difficult to determine if reintervention is necessary [6]. Failure to perform a reintervention in patients with persistent bacterial arthritis can lead to serious consequences, such as: joint damage, sepsis, the need for additional surgery, and even dead [7]. Therefore, it is important to minimise the risk of under-, but also overtreatment in these patients. The consensus definition of failure recommends that when a patient has persisting physical signs of arthritis (e.g., pain and swelling) and/or has persisting systemic

inflammation (e.g., fever or no improvement in CRP) despite surgical and antibiotic treatment after minimal 2 days of the initial surgery, diagnostic and treatment strategy needs then to be reconsidered, and reintervention might be necessary. Nonetheless, this remains a subjective judgement of a physician based on a constellation of symptoms and tests.

Consensus was reached on one physical criteria sign—pain—for deciding whether to do a reintervention or not. Pain should therefore be evaluated daily using a pain score of the affected joint [8]. When the pain score increases or remains persistently high, the physician should weigh this factor in deciding whether reintervention is needed or not.

Consensus was also reached on five systemic criteria signs—fever, sepsis, serum CRP, blood culture, and synovial fluid culture—for deciding whether to do a reintervention or not. While fever has limited specificity (57%) for bacterial arthritis, persistent or worsening postsurgery fever suggests ongoing infection [9]. Sepsis, a systemic response to infection with organ dysfunction, is associated with a higher reintervention rate [10–12]. Jung et al. [10] showed in a case-control study ($n = 137$) that patients with systemic sepsis had a significantly higher reintervention rate (51% vs. 24%, $p = 0.012$). This supports the use of systemic sepsis signs to determine the need for reintervention [10]. Serum CRP, with its high sensitivity for diagnosing bacterial arthritis, is a valuable screening test [6]. The baseline CRP level (i.e., preoperation) can serve as a benchmark to assess infection improvement or deterioration. After surgery, CRP levels usually increase, peaking at 2–3 days postsurgery. However, in cases of persistent infection, we expect persistently high CRP levels [13].

Although positive synovial fluid or blood cultures do not directly correlate with a higher reintervention risk, a positive joint culture itself confirms bacterial arthritis. Given the reported 26% failure rate, physicians must monitor patients with positive cultures for the potential persistence of the infection [2]. Furthermore, once the infection is confirmed, it is more likely to be a matter of persistence or recurrence rather than a noninfectious condition (e.g., gout). In contrast, bacterial arthritis cannot be ruled out, even with negative joint cultures, given their less than 100% sensitivity and potential for false negatives [9]. Blood cultures can complement inconclusive synovial fluid cultures, aiding in identifying the causative organism, and making them valuable for reintervention decisions in bacterial arthritis [6, 14].

It is felt that the criteria with consensus cannot be used independently for determining the need for reintervention. These criteria rely on expert opinion and should be validated in future cohort studies. The established definition and identified criteria for reintervention provide a valuable foundation for future research. Future prospective cohort studies, assessing these criteria across diverse patient groups, can offer valuable insights. By comparing clinical outcomes of patients meeting these criteria with those who do not meet these criteria, including the likelihood of reintervention, we can strengthen the evidence supporting their validity and use in clinical practice.

The study's primary strengths lie in the methodology employed to establish the new definition and clinical criteria for failure in patients with bacterial arthritis of a native joint. The current Delphi method is a structured approach, with a supporting review [2], that allows for the integration of multiple perspectives and can help to ensure that the final definition and criteria are supported by a diverse range of orthopaedic surgeons. Another strength of this study is the large number of participating experts completing all three rounds ($n = 55$, 92%).

However, the study has limitations and should be interpreted considering the following remarks: this study only included orthopaedic surgeons with experience in treating bacterial arthritis, excluding other physicians and stakeholders (e.g., infectious disease specialists, microbiologists, rheumatologists, and patients), which could have lead to a more heterogeneous group. However, since orthopaedic surgeons assess patients in clinic and make the ultimate decision for treatment and reintervention, we felt that such a sample would be most relevant for this Delphi study aim. Another limitation is the respondent population, consisting solely of experts from countries with advanced surgical options and ample antibiotic access. Concerns about external validity arise, especially regarding applying the findings to low-income countries with limited medical resources. Despite efforts for an international perspective, the expert panel represented only seven European countries and two non-European countries, with a majority from the Netherlands (78%). Though some differences in criteria rating were found based on country of origin and practice type, we do not feel that this should influence our conclusion as differences were small and might have been spurious due to the large number of statistical tests performed.

CONCLUSION

This Delphi study has defined treatment failure for bacterial arthritis and established a consensus on six specific criteria for identifying the need for reintervention. Both may help in development of clinical guidelines, and will empower physicians to make more precise and consistent decisions regarding reintervention for patients, ultimately aiming to reduce over- and undertreatment and improve patient outcomes.

THE INTERNATIONAL PANEL OF CLINICAL EXPERTS

Alessandro Carrozzo, Agnita Stadhouder, Arien J. Dijkstra, Arthur van Noort, Barend J. van Royen, Bernard G. Schutte, Changyu Huang, Christiaan J. A. van Bergen, Daniel Perez-Prieto, Denise Eygendaal, Dirk Jan Moojen, Edoardo Monaco, Gerard R. Schaap, Gijs T. T. Helmerhorst, Hubert J. Oostenbroek, Jan A. P. Geurts, Jesse W. P. Kuiper, Jon H. M. Goosen, Jon Karlsson, Joris J. W. Ploegmakers, Karsten D. Ottink, Laurens D. Pool, Lucien C. M. Keijser, Maaike P. J. van den Borne, Maarten V. Rademakers, Matthias U. Schafroth, Michael T. Hirschmann, Neel Desai, Niels H. Bech, Niels W. L. Schep, Olivier P. P. Temmerman, Pedro C. Diniz, Peter A. J. de Leeuw, Peter Z. Feczko, A. J. Peter Joosten, Peter Kloen, Peter A. Nolte, Peter A. A. Struijs, Pieter D'Hooghe, Pieter J. Emans, Rachid Saouti, Ritsert H. Boonstra, Roel de Haan, Rover Krips, A. Rutger Thomas, Rutger C. I. van Geenen, Stefan Keizer, Thorkell naebjornsson, Tomasz Piontek, Ton M. J. S. Vervest, Wim H. C. Rijnen, Ewout S. Veltman and one anonymous expert.

AUTHOR CONTRIBUTIONS

Alex B. Walinga: Conceptualisation; data curation; formal analysis; investigation; methodology; project administration; resources; supervision; validation; visualisation; writing—original draught; writing—review and editing; given final approval. **Stein J. Janssen:** Conceptualization; software; investigation; methodology; supervision; writing—original draught; writing—review and editing; given final approval. **Arthur J. Kievit:** Conceptualisation; methodology; supervision; writing—review and editing; given final approval. **Corianne A. J. M. de Borgie:** Methodology; writing—review and editing; given final approval. **Gino M. M. J. Kerkhoffs:** Conceptualisation; methodology; supervision; writing—review and editing; given final approval.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ETHICS STATEMENT

This study is in accordance with the ethical standards in the 1964 Declaration of Helsinki.

ORCID

Alex B. Walinga http://orcid.org/0000-0001-9441-632X

REFERENCES

1. Ravn, C., Neyt, J., Benito, N., Abreu, M.A., Achermann, Y., Bozhkova, S. et al. (2023) Guideline for management of septic arthritis in native joints (SANJO). *Journal of Bone and Joint Infection*, 8, 29–37.
2. Walinga, A.B., Stornebrink, T., Emanuel, K.S., Kievit, A.J., Janssen, S.J. & Kerkhoffs, G.M.M.J. (2023) Failure rates in surgical treatment in adults with bacterial arthritis of a native joint: a systematic review of 8,586 native joints. *Archives of Orthopaedic and Trauma Surgery*, 143, 6547–6559. Available from: https://doi.org/10.1007/s00402-023-04958-z
3. Diamond, I.R., Grant, R.C., Feldman, B.M., Pencharz, P.B., Ling, S.C., Moore, A.M. et al. (2014) Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, 67, 401–409.
4. Jünger, S., Payne, S.A., Brine, J., Radbruch, L. & Brearley, S.G. (2017) Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a methodological systematic review. *Palliative Medicine*, 31, 684–706.
5. Hsu, C. & Sandford, B.A. (2007) The Delphi technique: making sense of consensus. *Practical Assessment, Research, and Evaluation*, 12, 1–8.
6. Walinga, A.B., Stornebrink, T., Langerhuizen, D.W.G., Struijs, P.A.A., Kerkhoffs, G.M.M.J. & Janssen, S.J. (2021) What are the best diagnostic tests for diagnosing bacterial arthritis of a native joint?: a systematic review of 27 studies. *The Bone & Joint Journal*, 103-B, 1745–1753.
7. Abram, S.G.F., Alvand, A., Judge, A., Beard, D.J. & Price, A.J. (2020) Mortality and adverse joint outcomes following septic arthritis of the native knee: a longitudinal cohort study of patients receiving arthroscopic washout. *The Lancet Infectious Diseases*, 20, 341–349.
8. Baamer, R.M., Iqbal, A., Lobo, D.N., Knaggs, R.D., Levy, N.A. & Toh, L.S. (2022) Utility of unidimensional and functional pain assessment tools in adult postoperative patients: a systematic review. *British Journal of Anaesthesia*, 128, 874–888.
9. Margaretten, M.E., Kohlwes, J., Moore, D. & Bent, S. (2007) Does this adult patient have septic arthritis? *Journal of the American Medical Association*, 297, 1478–1488.
10. Jung, S.W., Kim, D.H., Shin, S.J., Kang, B.Y., Eho, Y.J. & Yang, S.W. (2018) Septic arthritis associated with systemic sepsis. *International Orthopaedics*, 42, 1–7. Available from: https://doi.org/10.1007/s00264-00017-03565-00264
11. Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M. et al. (2016) The third international consensus definitions for sepsis and septic shock (Sepsis-3). *Journal of the American Medical Association*, 315, 801–810.
12. Vincent, J.L., Opal, S.M., Marshall, J.C. & Tracey, K.J. (2013) Sepsis definitions: time for change. *The Lancet*, 381, 774–775.
13. Shetty, S., Ethiraj, P. & Shanthappa, A.H. (2022) C-reactive protein is a diagnostic tool for postoperative infection in orthopaedics. *Cureus*, 14, e22270.
14. Colston, J. & Atkins, B. (2018) Bone and joint infection. *Clinical Medicine*, 18, 150–154.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Walinga, A.B., Janssen, S.J., Kievit, A.J., The International Panel of Clinical Experts, de Borgie, C.A.J.M. & Kerkhoffs, G.M.M.J. (2024) Consensus on the definition and criteria for failure of surgical treatment in bacterial arthritis of a native joint: an international Delphi study. *Knee Surgery, Sports Traumatology, Arthroscopy*, 32, 235–242. https://doi.org/10.1002/ksa.12027