Defining Success and Failure In Prosthetic Joint Infections: A Meta-epidemiologic Study Toward A Core Outcome Set

Francesco Petri, Seyed Mohammad Amin Alavi, Cosmin Lucian Ciubotariu, Matteo Passerini, Fabio Borgonovo, Takahiro Matsuo, Haseeb Khawar, Ben A Marson, Anil Jagtiani, Mohammad Hassan Murad, Andrea Gori, Aaron J Tande, Elie F Berbari

Open Forum Infectious Diseases

MAJOR ARTICLE

Defining Success and Failure In Prosthetic Joint Infections: A Meta-epidemiologic Study Toward A Core Outcome Set

Francesco Petri,[1,2,a] Seyed Mohammad Amin Alavi,[3,a] Cosmin Lucian Ciubotariu,[1] Matteo Passerini,[1,2,4] Fabio Borgonovo,[1,2] Takahiro Matsuo,[2,5] Haseeb Khawar,[6] Ben A. Marson,[6] Anil Jagtiani,[7] Mohammad Hassan Murad,[2,8] Andrea Gori,[1,9,10] Aaron J. Tande,[2] and Elie F. Berbari[2]

[1]Department of Infectious Diseases, "Luigi Sacco" University Hospital, ASST FBF Sacco, Milan, Italy, [2]Division of Public Health, Infectious Diseases and Occupational Medicine, Department of Medicine, Mayo Clinic College of Medicine and Science, Mayo Clinic, Rochester, Minnesota, USA, [3]Faculty of Medicine, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran, [4]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy, [5]Department of Infectious Diseases, Infection Control and Employee Health, The University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA, [6]University Hospitals Nottingham NHS Trust, Trauma and Orthopedics, Nottingham, UK, [7]Department of Infectious Diseases, Kaiser Permanente Southern California, Fontana, California, USA, [8]Evidence-based Practice Center, Mayo Clinic, Rochester, Minnesota, USA, [9]Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy, and [10]Centre for Multidisciplinary Research in Health Science (MACH), Milan, Italy

Background. Prosthetic joint infection (PJI) represents an important complication of total joint arthroplasty, affecting approximately 2% of patients. However, considerable heterogeneity in outcome definitions renders cross-study and clinical comparisons challenging.

Methods. We performed a meta-epidemiological study and systematic review of cohort studies and randomized controlled trial (RCTs) reporting clear definitions of PJI outcomes. Relevant data regarding outcome definitions of treatment success or failure were extracted and categorized into 7 outcome domains: clinical signs and symptoms, imaging, microbiology, mortality, antibiotics-related, surgical-related, and histology. A network plot was generated to visualize domains combinations. Data on patient-reported outcome measures (PROMs) were also collected. Based on these results, we finally proposed a working definition of PJI core outcome set.

Results. The literature search identified 7585 references, leading to 461 studies. The majority were retrospective (416/461, 90.2%), with limited representation of prospective studies (37/461, 8.0%) and RCTs (8/461, 1.7%). Outcome reporting was highly heterogeneous, with the clinical criterion being the most frequently used domain, combined with microbiological, surgical, and mortality-related outcomes. In contrast, histopathological, imaging, and antibiotic-related criteria were rarely reported. Thirty-six PROMs were included in only 111 (24.1%) studies, most commonly the Knee Society Score (26%) and Harris Hip Score (23%).

Conclusions. We showed substantial heterogeneity in PJI outcome definitions, with inconsistent reporting across key domains, limiting comparability and evidence synthesis. Standardized, consensus-based outcomes incorporating patient-reported and antibiotic-related measures are needed, and the proposed core outcome set may support harmonized research and clinical practice aligned with the new PJI definition.

Keywords. core outcome set; meta-epidemiological study; outcome definitions; patient-reported outcome measures; prosthetic joint infection.

Total joint arthroplasty (TJA) is one of the most frequently performed orthopedic procedures worldwide. By 2030, the anticipated number of primary total knee arthroplasty and total hip arthroplasty surgeries in the United States is projected to be 3.48 million and 572,000, respectively [1]. Periprosthetic joint infection (PJI) affects approximately 2% of arthroplasty patients, a risk that has remained relatively constant over time [2].

There is considerable inconsistency in how outcomes are defined. For instance, while some studies classify the use of suppressive antibiotic therapy as treatment failure, others include it within the criteria for treatment success [3, 4]. These discrepancies make it challenging to compare outcomes across different treatment strategies and studies.

Various criteria have been proposed in the literature to define treatment success or failure in PJI, including 3 landmark studies (Table 1) that have significantly shaped the current understanding of treatment outcomes in PJI. In 2013, Diaz-Ledezma et al [5] proposed consensus criteria for defining treatment success, focusing on clinical healing, normalization of inflammatory markers, no need for ongoing antibiotic

Received 12 March 2026; accepted 05 June 2026; published online 21 June 2026

[a]equally contributed.

Correspondence: Cosmin Lucian Ciubotariu, MD, Department of Infectious Diseases, "Luigi Sacco" University Hospital, ASST FBF Sacco, Via G. B. Grassi Milan, Italy (cosmin.ciubotariu@unimi.it); Elie F. Berbari, MD, MBA, FIDSA, Chief of the Department of Medicine, Professor of Medicine, Mayo Clinic College of Medicine and Science, Mayo Clinic, Rochester, MN 74–20157 55905 (berbari.elie@mayo.edu).

Open Forum Infectious Diseases®

https://doi.org/10.1093/ofid/ofag364

Table 1. Characteristics of the Included Studies

Characteristic	Number of Studies (N = 461)
Study design	…
Retrospective	416 (90.2%)
Prospective	37 (8.0%)
RCTs	8 (1.7%)
Year of publication	…
2005–2009	18 (3.9%)
2010–2014	73 (15.8%)
2015–2019	143 (31.0%)
2020–2024	227 (49.2%)
Type of joint involved	…
Hip	83 (18.0%)
Knee	138 (29.9%)
Shoulder	2 (0.4%)
Other (elbow)	1 (0.2%)
Miscellaneous	237 (51.4%)
Size of cohort	…
≥50– < 100 patients	204 (44.2%)
≥100– < 200 patients	157 (34.0%)
≥200–1000 patients	96 (20.8%)
≥1000 patients	4 (0.9%)
Type of surgical approach	…
DAIR	90 (19.9%)
1-stage	26 (5.6%)
2-stages	178 (38.6%)
Other	9 (1.9%)
Miscellaneous	154 (33.4%)
Not mentioned	3 (0.7%)
Definition of PJI	…
2011[a]	81 (17.6%)
2014[b]	66 (14.3%)
MSIS 2018[c]	62 (13.4%)
EBJIS 2021[d]	11 (2.4%)
Other	129 (28.0%)
Not specified	111 (24.1%)
Source of the definition of PJI outcome[e]	…
Author-derived	264/438 (60.3%)
Peer-reviewed	174/438 (39.7%)
Delphi 2013[f,g]	75/174 (43.1%)
MSIS ORT 2019[f,h]	25/174 (14.4%)
Miscellaneous	76/174 (43.7%)
Defined success	265/438 (60.5%)
Defined failure	306/438 (69.9%)
Considered PROMs or functional outcomes	111 (24.1%)
Minimum follow-up	…
<2 y	192 (41.6%)
≥2 y	247 (53.6%)
Not mentioned	22 (4.8%)
Individual criteria for outcome definition[e]	…
Surgical-related domain	309/438 (70.5%)
Clinical domain	279/438 (63.7%)
Microbiological domain	247/438 (56.4%)
Mortality domain	191/438 (43.6%)
Antibiotics-related domain	135/438 (30.8%)
Histopathological domain	38/438 (15.5%)
Imaging domain	52/438 (11.9%)

[a]Parvizi J, Zmistowski B, Berbari EF, et al. New definition for periprosthetic joint infection: from the Workgroup of the Musculoskeletal Infection Society. Clin Orthop Relat Res. 2011;469(11):2992–2994. doi:10.1007/s11999-011-2102-9.

[b]Parvizi J, Gehrke T; International Consensus Group on Periprosthetic Joint Infection. Definition of periprosthetic joint infection. J Arthroplasty. 2014;29(7):1331. doi:10.1016/j.arth.2014.03.009.

[c]Parvizi J, Tan TL, Goswami K, et al. The 2018 Definition of Periprosthetic Hip and Knee Infection: An Evidence-Based and Validated Criteria. J Arthroplasty. 2018;33(5):1309–1314.e2. doi:10.1016/j.arth.2018.02.078.

[d]McNally M, Sousa R, Wouthuyzen-Bakker M, et al. The EBJIS definition of periprosthetic joint infection. Bone Joint J. 2021;103-B(1):18–25. doi:10.1302/0301-620X.103B1.BJJ-2020-1381.R1.

[e]23/461 studies provided only PROMs or functional outcomes without a quotable definition.

[f]Two studies considered both Delphi 2013 and MSIS ORT 2019 as their reference outcome for PJI.

[g]Diaz-Ledezma C, Higuera CA, Parvizi J. Success after treatment of periprosthetic joint infection: a Delphi-based international multidisciplinary consensus. Clin Orthop Relat Res. 2013;471(7):2374–2382. doi:10.1007/s11999-013-2866-1.

[h]Fillingham YA, Della Valle CJ, Suleiman LI, et al. Definition of Successful Infection Management and Guidelines for Reporting of Outcomes After Surgical Treatment of Periprosthetic Joint Infection: From the Workgroup of the Musculoskeletal Infection Society (MSIS). J Bone Joint Surg Am. 2019;101(14):e69. doi:10.2106/JBJS.19.00062.

Abbreviations: DAIR, debridement, antibiotics, and implant retention; EBJIS, European Bone and Joint Society; MSIS, Musculoskeletal Infection Society; ORT, outcome reporting tool; PJI, prosthetic joint infection; PROM, patient-reported outcome measure; RCT, randomized controlled trial.

therapy, and the absence of infection recurrence or additional surgical intervention. In 2019, the Musculoskeletal Infection Society (MSIS) introduced a 4-tier outcome reporting tool (MSIS ORT) that categorizes results based on infection control, antibiotic use, need for further surgery, and mortality [6]. Building on these efforts, John et al. in 2022 utilized a desirability of outcome ranking approach to establish a more detailed framework encompassing a range of outcomes—from complete cure to relapse, chronic infection managed with suppressive antibiotics, and death [7]. Previous studies have shown that applying the MSIS ORT criteria tends to result in a lower reported rate of successful outcomes among patients treated for PJI [8, 9].

Furthermore, several studies have utilized functional outcome measures [10–13]. More recently, studies have increasingly focused on patient-reported outcome measures (PROMs) to capture patients' perspectives on their functional status and quality of life following treatment. However, to date, there are no PROMs specifically built for PJI but only for functional outcomes and patients' perception after primary TJA. However, most currently used PROMs were developed for primary arthroplasty and do not capture key PJI-specific burdens, such as prolonged antibiotic exposure, treatment-related toxicity, repeated hospitalizations and surgeries, psychosocial distress, and long-term uncertainty about infection control.

In this study, we conducted a systematic review using a meta-epidemiological methodology. This methodology was previously used in other studies on other bone infections (native vertebral osteomyelitis; postoperative spine infection) [14, 15], to define the reported outcomes of PJI and provide evidence foundation

to new definition of diseases [16]. Additionally, we aimed to identify the functional assessment tools and PROMs previously utilized in the literature to provide a future unified framework.

METHODS

The current meta-epidemiological study and systematic review was conducted and reported in accordance with the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for meta-epidemiological methodological research [17] and disease definition in guideline development [18] (Supplementary Table 1). In brief, meta-epidemiological methodology refers to a systematic, study-level framework approach that examines how definitions, design features, reporting practices, or methodological choices vary across studies and how this variation may affect evidence interpretation, bias, comparability, and guideline development.

Protocol

A predefined protocol was developed prior to the initiation of the study. This can be publicly accessed at https://protocols.io/view/redefining-the-outcomes-of-prosthetic-joint-infect-hqx6b5xrf.

Eligibility Criteria

Inclusion criteria were as follows: (1) studies involving adult populations (≥18 years old); (2) clear and quotable definition of PJI outcomes, including successful outcomes or adverse outcomes such as failure, relapse, reinfection, recurrence, or their synonyms, as well as studies reporting functional outcome measures or PROMs; (3) inclusion of ≥50 patients; (4) publication year 2005 or later; (5) articles published in any language; and (6) study design limited to cohort studies and randomized controlled trials (RCTs). Exclusion criteria included: (1) studies involving pediatric populations; (2) case reports, review articles, systematic reviews, meta-analyses, study protocols, gray literature, and preprint materials. A minimum sample size of 50 patients was chosen to focus on studies more likely to reflect routine clinical practice and to avoid very small, potentially idiosyncratic cohorts. The start date of 2005 was chosen to comprehend a 20-year time-frame to ensure generalizability and to detect changes in the attribution of outcomes and evolution of PJI treatment effect on them.

The primary objective was to analyze and deconstruct the definitions of PJI outcome, identifying predefined combinations of definition elements and evaluating their distribution over time and by thematic grouping. The secondary objective was to examine a set of items thematically related to the diagnostic criteria and propose a working definition of PJI outcomes.

Information Sources and Search Strategies

A comprehensive search of multiple databases was conducted on 6 November 2024. The search was restricted to cohort studies and randomized controlled trials published from 2005 onward. Databases searched (and their content coverage dates) were Ovid MEDLINE(R) (1946+ including epub ahead of print, in-process, and other nonindexed citations), Ovid Embase (1974+), Ovid Cochrane Central Register of Controlled Trials (1991+), Ovid Cochrane Database of Systematic Reviews (2005+), and Scopus via Elsevier (1970+).

Search strategies were developed and executed by an experienced medical librarian in collaboration with the research team (LH). A combination of controlled vocabulary and keyword terms was used to ensure a comprehensive and precise search. The full search methodology is provided in the Supplementary Material.

Selection of Studies

The screening of abstracts and full texts was performed using the COVIDence Systematic Review Software (Veritas Health Innovation, Melbourne, Australia). Two independent reviewers (F. B. and S. M. A. A.) screened all titles, abstracts, and full-text articles. Each record was evaluated by both reviewers to ensure accuracy and consistency. In cases of disagreement, a third reviewer (F. P.) was consulted, or consensus was achieved through discussion among the reviewers. Details regarding inter-rater reliability are provided in the Supplementary Appendix.

Data Extraction

A data extraction form was initially tested in Excel using a sample of 60 studies. Six reviewers (C. L. C., F. B., F. P., H. K., M. P., and S. M. A. A.) participated in this pilot phase. Based on feedback and recommendations from the research team, the form was refined through an iterative process to minimize misunderstandings and resolve potential discrepancies. Following this, 5 reviewers (C. L. C., F. B., F. P., M. P., and S. M. A.A.) extracted relevant data from each included study, recording it in Excel spreadsheets. Risk of bias assessment was deemed unnecessary for this study, as it did not aim to quantify associations.

Data Synthesis and Analysis

Diagnostic criteria used to define PJI at baseline were recorded descriptively, but the primary focus of this study was outcome definitions used to define treatment success or failure.

A thematic synthesis of the outcome definitions was performed, allowing a single source definition to generate multiple output definitions if authors provided alternative criteria options within their research, as we shown in previous studies [14, 15, 19].

To cluster the outcome definitions, after the pilot analysis on the first 60 definitions, we established 7 predefined outcome domains used as a framework to define treatment success or failure for PJI based on common clinical practices: (1) clinical features (including symptoms, signs, and laboratory biomarkers), (2) imaging findings, (3) microbiological evidence, (4) mortality, (5) antibiotics-related factors, (6) surgical-related

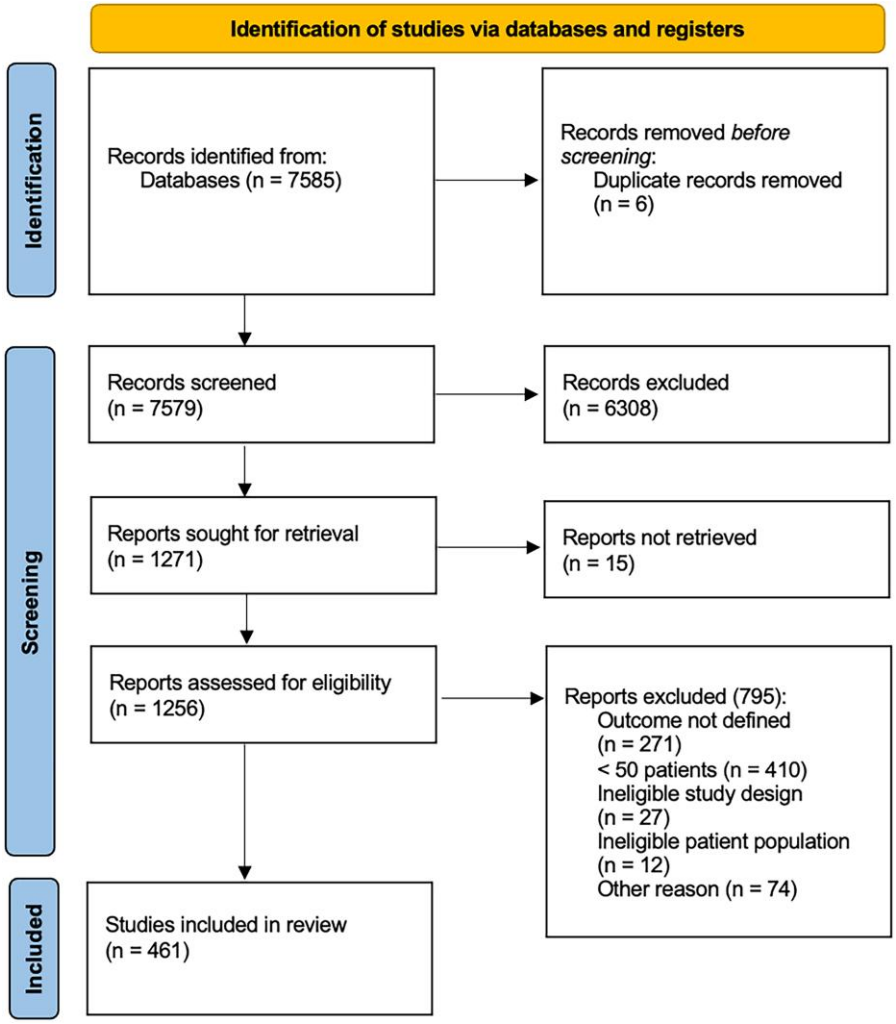

Figure 1. 2020 PRISMA flow diagram.

factors, and (7) histopathology. We also grouped the definitions encountered distinguishing into success and failure.

To graphically depict the outcomes of this study, we utilized multidimensional analytics with a network plot to represent combinations derived from the definitions. Each domain was treated as a node in the network. Co-occurrence of 2 domains within the same study outcome definition generated an edge between the corresponding nodes. Only pairwise combinations of domains were considered, as network edges represent binary relationships. The resulting network therefore reflects how frequently different outcome domains were jointly used across studies, with nodes representing outcome domains and edges representing their co-occurrence. We collected data on an Excel spreadsheet and built network graphs with an online tool (https://app.flourish.studio/projects, accessed on 5 September 2025). Figures were modified with Biorender (https://BioRender.com). Since no inferential analysis was performed, descriptive statistics were used, expressed as counts (percentage) and median (interquartile range [IQR]), when required.

RESULTS

The literature search yielded 7585 references, of which 461 were finally included, covering 87 021 patients with PJI. The characteristics and bibliography of the studies included can be found in the Supplementary Appendix (Supplementary Table 2). The study selection process is depicted in Figure 1 and Supplementary Table 3. Most definitions were described in retrospective observational studies (416/461, 90.2%). The remaining were in prospective studies (37/461, 8.0%), and RCTs (8/461, 1.7%). The median population of the PJI cohorts was 109 (IQR 76–178) patients.

Network analysis of outcome domains used in PJI studies (Figure 2) shows a predominance of clinical, microbiological, surgical, and mortality-related criteria. These 4 domains form the most densely connected cluster, reflecting their frequent co-occurrence in outcome definitions. In particular, the clinical criterion is the most central node, commonly paired with microbiology, surgery, and mortality. Conversely, histopathological

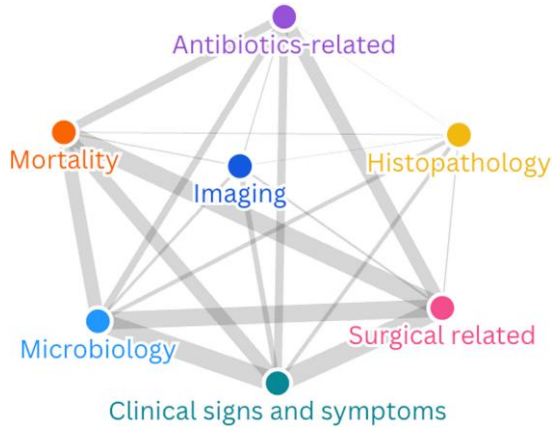

Figure 2. Network plot illustrating the co-occurrence of outcome criteria in PJI studies that provided a definition of outcome (overall, N = 438).

and imaging-based criteria appear infrequently, either alone or in combination, and are positioned at the periphery of the network. Antibiotic-related outcomes also show limited integration. The thickness of each edge is proportional to the number of studies reporting that specific combination of criteria, highlighting both common and underutilized domains in the literature. This approach captures how often domains co-occur but does not account for the relative importance or hierarchy of individual components within composite definitions.

Network analysis showed similar patterns on the distribution of the used thematic criteria for both success (Figure 3) and failure (Figure 4).

Among the 461 included studies, 111 (24.1%) reported at least 1 PROM or functional assessment tool. Considerable heterogeneity was observed in the instruments used. In

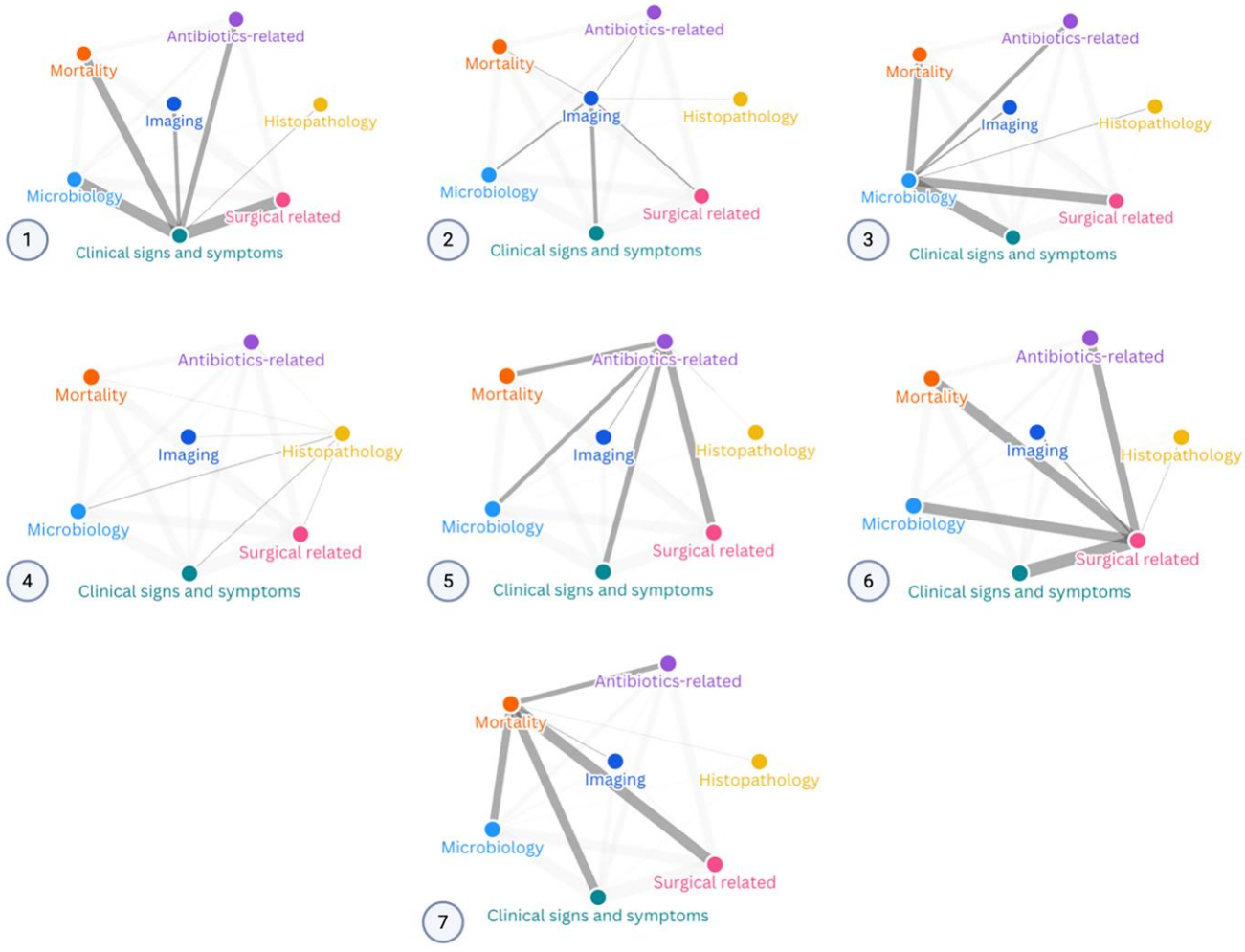

Figure 3. Network plot illustrating the co-occurrence of outcome criteria in PJI studies that provided a definition of success (n = 265). Edge thickness reflects the frequency of the co-occurrence of each criterion. Each plot highlights 1 criterion, namely (1) clinical signs, (2) imaging, (3) microbiology, (4) histopathology, (5) antibiotics-related, (6) surgical-related, and (7) mortality.

Figure 4. Network plot illustrating the co-occurrence of outcome criteria in PJI studies that provided a definition of failure (n = 306). Edge thickness reflects the frequency of the co-occurrence of each criterion. Each plot highlights 1 criterion, namely (1) clinical signs, (2) imaging, (3) microbiology, (4) histopathology, (5) antibiotics-related, (6) surgical-related, and (7) mortality.

fact, we identified 36 different PROMs or scores. Among those, we focused on the 15 most used across the included studies (Table 2).

Among the 111 studies that reported PROMs, the most frequently used instruments were the Knee Society Score (KSS) (26%) and Harris hip score (HHS) (23%), followed by the Western Ontario and McMaster Universities Arthritis Index (WOMAC) and the visual analog scale (VAS) (17% each). Other commonly used tools included the Oxford knee score (13%), EQ-5D and SF-12 (12% each). Hip-specific scores such as OHS (9%) and HOOS (5%) were less common. No PROM was specifically designed for patients with PJI.

Further analysis on the frequency of the verbatim words used to define outcomes for PJI are shown in a circle chart that can be found in Supplementary Appendix (Supplementary Fig. 1).

DISCUSSION

In this large meta-epidemiologic review of over 460 studies, we show that definitions of treatment success and failure in PJI remain highly heterogeneous, with wide variation in the domains used and frequent reliance on inconsistent composite outcomes. Our findings reveal the predominance of certain domains—namely clinical, microbiological, surgical, and mortality-related elements—while other relevant dimensions, such as histopathology, imaging, and antibiotic-related criteria, remain underutilized. These results underscore the urgent need for harmonization and standardization in outcome reporting for PJI, particularly for comparative effectiveness studies, guideline development, and patient-centered care. This need is especially relevant in light of the newly proposed intersocietal unified diagnostic definition of PJI [20].

Table 2. Distribution of the 15 PROMs Most Used in the Included Studies

PROM	Number of Studies (n = 111)
Knee society score (KSS)	29 (26.1%)
Harris hip score (HHS)	26 (23.4%)
Western Ontario and McMaster Universities Arthritis index (WOMAC)	19 (17.1%)
Visual analog score (VAS)[a]	19 (17.1%)
Oxford knee score (OKS)	14 (12.6%)
EuroQol-5D (EQ-5D)[a]	13 (11.7%)
12-Item short form health survey (SF-12)[a]	13 (11.7%)
Oxford hip score (OHS)	10 (9%)
Knee osteoarthritis outcome score (KOOS)	7 (6.3%)
Hospital for special surgery (HSS)	6 (5.4%)
Hip osteoarthritis outcome score (HOOS)	5 (4.5%)
University of California, Los Angeles (UCLA) activity score	5 (4.5%)
36-Item short form health survey (SF-36)[a]	5 (4.5%)
Hip society score	4 (3.6%)
Merle d'Aubigné and Postel method	3 (2.7%)

[a]These are tools not specifically addressing orthopedic performance indexes but general health-related quality of life measures.

Unlike Debbi et al [21], who compared predefined outcome tiers, our study maps how outcomes are originally defined across PJI studies. We analyze source definitions, domain co-occurrence, and PROM usage—providing broader, data-driven insights to inform future core outcome set (COS) development.

The most frequently coreported domains were clinical signs and symptoms (including inflammatory biomarkers), microbiological evidence, and surgery-related criteria. These were often used in composite definitions of both success and failure and frequently linked to mortality endpoints. Such combinations reflect common conceptualizations of treatment outcomes—eg, the resolution of infection signs, absence of culture-positive recurrence, and no need for further surgical intervention.

Conversely, our analysis shows that histopathology and imaging were rarely included, either alone or in combination with other domains. This may reflect both clinical practice limitations and methodological challenges: histopathology is not routinely available in all centers, and imaging often lacks specificity or standardized interpretative criteria in the PJI setting. Similarly, antibiotic-related criteria—such as the need for suppressive therapy, treatment duration, or intolerance—were surprisingly uncommon, despite their high relevance to patient quality of life and long-term management. This underreporting may lead to an incomplete understanding of therapeutic success, particularly in cases where surgical cure is not achieved, but infection control is maintained through long-term antimicrobial strategies.

The primary strength of this study lays in its systematic and reproducible methodology, including a preregistered protocol, a broad literature search across multiple databases, and a rigorous dual-reviewer screening and extraction process. By focusing on studies with ≥50 patients and spanning a 20-year period, we ensured both breadth and a degree of methodological quality, avoiding small, potentially underpowered or anecdotal reports.

Nonetheless, several limitations should be acknowledged. First, the classification of outcome criteria into domains, while grounded in clinical relevance and tested in a pilot phase, remains to some extent arbitrary. Studies often use vague, inconsistent, or overlapping definitions of outcomes, and the mapping of textual definitions to specific domains required interpretative judgment. We attempted to mitigate this through consensus-based coding and predefined operational definitions, but misclassification bias cannot be entirely excluded.

Second, only binary (present/absent) coding was used for individual domains and combinations. This precludes quantification of the weight or hierarchy of criteria within composite definitions –eg, whether failure was defined primarily by clinical relapse, or only when combined with revision surgery. Similarly, we considered only pairwise domain co-occurrence for network analysis, since network edges represent dyadic relationships. This necessarily excludes higher-order combinations (eg, triple or quadruple criteria definitions), which are common in many composite endpoints. While this simplification enhances visualization and interpretability, it inevitably reduces granularity.

Third, the vast majority of included studies were retrospective in nature (90.2%), with only a small proportion of prospective cohorts or RCTs. This reflects the general landscape of PJI research, which is dominated by observational designs due to ethical and logistical constraints in surgical infectious diseases. However, it also limits the quality and standardization of outcome reporting, as retrospective studies are more likely to use post hoc or center-specific definitions.

Fourth, many studies lacked an explicit statement distinguishing between "success" and "failure" outcomes, or used inconsistent thresholds. For example, some classified long-term suppressive therapy as a success, while others considered it a marker of failure [7, 21]. This inconsistency is not merely semantic—it has direct implications for reported outcome rates, benchmarking, and policy. It also introduces subjectivity in the synthesis process: depending on the definition applied, the same patient outcome could be labeled as success or failure.

Finally, our data extraction relied on manual review and spreadsheet-based recording, which limits scalability and introduces human error.

One of the most important insights from this review is the persistent ambiguity in how success and failure are defined across the literature. This lack of consensus has major implications for both clinical decision-making and research comparability. As shown in previous studies, applying different published criteria (eg, the 2019 MSIS ORT tool vs the 2013

Delphi-based criteria) can result in significantly different reported success rates from the same clinical data [8, 22, 23].

A standardized, reproducible definition of treatment success and failure in PJI is essential for benchmarking, risk adjustment, research, and clinical decision-making. Current heterogeneity hampers meta-analyses, guideline development, and the comparability of treatment strategies (eg, debridement, antibiotics, implant retention [DAIR] vs 2-stage exchange, different antibiotic regimens). Our findings highlight the need for a consensus-based, tiered definition incorporating both objective (clinical and microbiological) and patient-centered endpoints.

While prior frameworks (eg, 2019 MSIS ORT, Delphi 2013) offer a foundation, they lack consistent adoption, granularity, and adaptability to real-world scenarios—such as long-term suppressive therapy or persistent symptoms despite infection control. A robust definition should also address time-dependence, competing risks (eg, death vs reinfection), and stratification by surgical strategy and pathogen.

Notably, only 24.1% of studies reported PROMs, and most used tools designed for primary arthroplasty (eg, WOMAC, HHS, and KSS), not tailored to the specific burden of PJI. This underscores the need for a dedicated PROM instrument addressing domains like pain, mobility, antibiotic burden, and psychosocial impact.

Our review provides a foundation for a future COS for PJI, aligned with the Core Outcome Measures in Effectiveness Trials methodology [24]. Development and validation of a COS will require Delphi-based consensus and stakeholder engagement. In the interim, we encourage authors to clearly report which domains their outcome definitions include and adopt standard terminology for transparency and comparability.

In the absence of a validated PJI COS, we propose a working definition of PJI treatment success that includes: (1) resolution of clinical signs and symptoms, (2) no further infection-related surgical intervention, and (3) absence of microbiologically confirmed recurrence. To enable meaningful comparisons across clinical studies, registries, and policy frameworks, we suggest a 5-tier classification system:

1. Infection-free survival off antibiotics, with no reoperations and clinical recovery.
2. Infection control with ongoing suppressive antibiotic therapy, which duration should be reassessed at standardized timing by an integration of clinical and laboratory biomarkers, and supported where available by advanced imaging (eg, fluorodeoxyglucose positron emission tomography/computed tomography [FDG-PET/CT]). Further evidence is needed to reach definitive conclusions on these advanced tools.
3. Relapse or Reinfection: recurrence of infection with the same microbiologically confirmed organism or new infection with a different microbiologically confirmed organism, respectively.
4. Reoperation: any infection-related, unplanned surgical intervention (including DAIR, resection arthroplasty, arthrodesis, retained spacer, or amputation), consistent with 2019 MSIS ORT categories.
5. Death: all-cause or infection-related mortality.

We recommend outcomes be reported at standardized intervals, with a minimum follow-up of 2 years.

A key feature of the proposed tiered framework is that it captures PJI outcomes as a dynamic clinical pathway, rather than as a single static endpoint. Distinguishing infection control off antimicrobial therapy, infection control under suppressive antimicrobial therapy, and relapse or reinfection is clinically relevant because each state carries different implications for prognosis, diagnostic reassessment, antimicrobial management, and potential surgical intervention. Importantly, relapse or reinfection should also be reported according to whether it occurs off antibiotics or during suppressive therapy, as these scenarios may reflect different levels of disease control and may guide escalation toward further diagnostic work-up, treatment modification, or surgical strategies, potentially moving the patient into another tier. For this reason, the framework emphasizes repeated assessment at predefined, standardized time intervals, allowing patients' trajectories across tiers to be consistently documented throughout their PJI journey.

The proposed framework is not intended to impose uniform clinical practice across centers, as PJI definitions, diagnostic tools, and management strategies inevitably vary over time and across healthcare settings. Rather, it provides an adaptable structure for reporting outcomes consistently despite this heterogeneity, improving comparability across studies and highlighting underreported domains that may be incorporated where feasible.

We suggest this framework also serve as a springboard for the development of a PJI-specific PROM, grounded in both patient and clinician input. As an interim step, established tools such as the KSS [25] and HHS [26] could be used for hip and knee-specific domains. These may be integrated with global health scores such as EQ-5D [27] and SF-12 [28]. Importantly, PROMs must account for patient choice and acceptability, including satisfaction with suppressive therapy or retained spacers. Development of a PJI-specific PROM should incorporate qualitative input from the community of patients.

Future efforts should focus on 4 key areas: (1) endorsement and systematic implementation of a COS in both clinical care and PJI research to standardize outcome reporting; (2) development and validation of PJI-specific PROMs to capture patient-centered outcomes more accurately; (3) assessment of the performance and applicability of existing PJI definitions to revised joint arthroplasties; and (4) alignment of the proposed COS with the forthcoming unified 2025 definition of PJI to ensure consistency across diagnostic and outcome frameworks.

CONCLUSIONS

This meta-epidemiological review highlights the significant heterogeneity in outcome definitions for PJI and reveals a skewed emphasis on certain domains, with underreporting of others. The absence of standardized, consensus-based definitions of success and failure complicates clinical interpretation, undermines research synthesis, and may ultimately limit the quality of care delivered to patients with PJI.

Harmonizing outcome definitions—particularly through the inclusion of patient-reported and antibiotic-related measures—should be a priority for future research. To support this, we propose an interim COS as a foundation for advancing key goals: its systematic application in clinical care and research, the development of validated PJI-specific PROMs, and alignment with the forthcoming unified 2025 definition of PJI to ensure coherence between diagnostic and outcome frameworks.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. The authors acknowledge Leslie Hassett, M.L.S., AHIP, for conducting the literature search. No funding needs to be declared for this study.

Author contributions. F. P.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft; S. M. A. A.: Conceptualization, Data curation, Investigation, Writing – original draft; C. L. C.: Conceptualization, Data curation, Investigation, Writing – original draft; M. P.: Conceptualization, Investigation, Supervision, Writing – review &; editing; F. B.: Conceptualization, Investigation, Writing – review &; editing; T. M.: Conceptualization, Supervision, Writing – review &; editing; H. K.: Investigation, Writing – review &; editing; B. A. M.: Conceptualization, Supervision, Writing – review &; editing; A. J.: Conceptualization, Supervision, Writing – review &; editing; M. H. M.: Methodology, Supervision, Writing – review &; editing; A. G.: Supervision, Writing – review &; editing; A. J. T.: Conceptualization, Supervision, Writing – review &; editing; E. F. B.: Conceptualization, Supervision, Writing – review &; editing.

Data availability statement. The full dataset are available through the first and corresponding authors upon a reasonable request. ChatGPT was used for general proofreading of English language and not for analytical or generation of new text, according to the World Association of Medical Editors' recommendations on AI, chatbots. This study was previously presented at the European Bone and Joint Society Conference (EBJIS); September 11–13, 2025; Bologna, Italy (oral presentation) and at IDWeek 2025; October 19–22, 2025; Atlanta, GA, USA (poster presentation).

Potential conflicts of interest. All authors: No reported conflicts.

References

1. Tabaja H, Abu Saleh OM, Osmon DR. Periprosthetic joint infection: what's new? Infect Dis Clin North Am **2024**; 38:731–56.
2. Nelson SB, Pinkney JA, Chen AF, Tande AJ. Periprosthetic joint infection: current clinical challenges. Clin Infect Dis **2023**; 77:e34–45.
3. Biddle M, Kennedy IW, Wright PM, Ritchie ND, Meek RMD, Rooney BP. Improving outcomes in acute and chronic periprosthetic hip and knee joint infection with a multidisciplinary approach. Bone Jt Open **2021**; 2:509–14.
4. Barbero Allende JM, Gómez-Junyent J, Sorlí Redó L, et al. Description of reinfection of joint prosthesis after 2-stage replacement (infection of the 2nd stage prosthesis): a multicenter study. Enferm Infecc Microbiol Clin (Engl Ed) **2024**; 42: 354–60.
5. Diaz-Ledezma C, Higuera CA, Parvizi J. Success after treatment of periprosthetic joint infection: a Delphi-based international multidisciplinary consensus. Clin Orthop Relat Res **2013**; 471:2374–82.
6. Fillingham YA, Della Valle CJ, Suleiman LI, et al. Definition of successful infection management and guidelines for reporting of outcomes after surgical treatment of periprosthetic joint infection: from the workgroup of the musculoskeletal infection society (MSIS). J Bone Joint Surg Am **2019**; 101:e69.
7. Johns BP, Dewar DC, Loewenthal MR, et al. A desirability of outcome ranking (DOOR) for periprosthetic joint infection - a Delphi analysis. J Bone Jt Infect **2022**; 7:221–9.
8. Borsinger TM, Pierce DA, Hanson TM, Werth PM, Orem AR, Moschetti WE. Is the proportion of patients with "successful" outcomes after two-stage revision for prosthetic joint infection different when applying the musculoskeletal infection society outcome reporting tool compared with the Delphi-based consensus criteria? Clin Orthop Relat Res **2021**; 479:1589–97.
9. Zielinski MR, Ziemba-Davis M, Meneghini RM. Comparison of Delphi consensus criteria and musculoskeletal infection society outcome reporting tool definitions of successful surgical treatment of periprosthetic knee infection. J Arthroplasty **2024**; 39:2357–62.
10. Wildeman P, Rolfson O, Söderquist B, Wretenberg P, Lindgren V. What are the long-term outcomes of mortality, quality of life, and hip function after prosthetic joint infection of the hip? A 10-year follow-up from Sweden. Clin Orthop Relat Res **2021**; 479:2203–13.
11. Bunting AC, Costi K, Chimutengwende-Gordon M, et al. Staged revision hip arthroplasty with femoral impaction bone grafting for prosthetic joint infections: radiostereometric analyses and clinical outcomes at Minimum 5-year follow-up. J Arthroplasty **2023**; 38:2716–23.e1.
12. Ohlmeier M, Alrustom F, Citak M, Rolvien T, Gehrke T, Frings J. The clinical outcome of different total knee arthroplasty designs in one-stage revision for periprosthetic infection. J Arthroplasty **2022**; 37:359–66.
13. Schnetz M, Ewald L, Jakobi T, Klug A, Hoffmann R, Gramlich Y. Use of hinged implants for multi-stage revision knee arthroplasty for severe periprosthetic joint infection: remission rate and outcomes after a Minimum follow-up of five years. J Arthroplasty **2025**; 40:218–26.e1.
14. Petri F, Mahmoud OK, Zein SE, et al. Wide variability of the definitions used for native vertebral osteomyelitis: walking the path for a unified diagnostic framework with a meta-epidemiological approach. Spine J **2025**; 25:359–68.
15. Alavi SMA, Borgonovo F, Petri F, et al. Defining postoperative spinal infections: navigating the inconsistencies in diagnostic definitions. J Bone Jt Infect **2025**; 10: 451–7.
16. Shaw JD, Nelson SB, Simpfendorfer C, et al. New definition for postoperative spine infection (PSI): from the workgroup of the musculoskeletal infection society (MSIS) and the European Bone & Joint Infection Society (EBJIS). Spine (Phila Pa 1976) **2025**.
17. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. Evid Based Med **2017**; 22:139–42.
18. Kawtharany H, Azzam M, Murad MH, et al. Proposed framework for unifying disease definitions in guideline development. BMJ Evid Based Med **2025**; 30: 351–4.
19. Petri F, Mahmoud O, El Zein S, et al. It is time for a unified definition of native vertebral osteomyelitis: a framework proposal. J Bone Jt Infect **2024**; 9:173–82.
20. Wouthuyzen-Bakker M. Unified definition of PJI. Oral presentation at: European Bone and Joint Infection Society; 2025; Bologna, Italy.
21. Debbi EM, Khilnani T, Gkiatas I, et al. Changing the definition of treatment success alters treatment outcomes in periprosthetic joint infection: a systematic review and meta-analysis. J Bone Jt Infect **2024**; 9:127–36.
22. Tan TL, Goswami K, Fillingham YA, Shohat N, Rondon AJ, Parvizi J. Defining treatment success after 2-stage exchange arthroplasty for periprosthetic joint infection. J Arthroplasty **2018**; 33:3541–6.
23. Johns BP, Loewenthal MR, Dewar DC, Manning LA, Davis JS. Comparison of surgical treatments for hip and knee periprosthetic joint infections using the desirability of outcome ranking in a prospective multicentre study. J Bone Jt Infect **2025**; 10:73–84.
24. Prinsen CA, Vohra S, Rose MR, et al. Core Outcome Measures in Effectiveness Trials (COMET) initiative: protocol for an international Delphi study to achieve

consensus on how to select outcome measurement instruments for outcomes included in a 'core outcome set'. Trials **2014**; 15:247.
25. 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.
26. Mahomed NN, Arndt DC, McGrory BJ, Harris WH. The Harris hip score: comparison of patient self-report with surgeon assessment. J Arthroplasty **2001**; 16:575–80.
27. Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. Appl Health Econ Health Policy **2017**; 15:127–37.
28. Gandek B, Ware JE, Aaronson NK, et al. Cross-validation of item selection and scoring for the SF-12 health survey in nine countries: results from the IQOLA project. International Quality of Life Assessment. J Clin Epidemiol **1998**; 51: 1171–8.