Consensus position statement on advancing the classification of patients and tests of cure in studies of antibiotic treatment of complicated urinary tract infections

Authors: Truls E. Bjerklund Johansen, MD (Prof);[1] Christina Bahrs, MD;[2] Franck Bruyere, MD (Prof);[3] Tommaso Cai, MD;[4] Amelia Joseph, BMBS;[5] Béla Köves, MD;[6] David M. Livermore, PhD (Prof);[7] Alessandra Oliva, MD;[8] Alex Soriano, PhD;[9] Florian Wagenlehner, MD (Prof);[10] Per Henrik Zahl, PhD;[11] Manuel Etienne, MD (Prof);[12] cUTI consensus group

Affiliations:

1. Urology department, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; and Institute of Clinical Medicine, University of Aarhus, Aarhus, Denmark
2. Division of Infectious Diseases, Department of Medicine I, Medical University of Vienna, Vienna, Austria; and Institute of Infectious Diseases and Infection Control, Jena University Hospital/Friedrich-Schiller-University, Jena, Germany
3. Urologie, CHRU Tours, Tours, France
4. Department of Urology, Santa Chiara Regional Hospital, Provincial Health Care Agency (APSS), Trento, Italy
5. Department of Clinical Microbiology, Nottingham University Hospitals NHS Trust, Queens Medical Centre, Nottingham, UK
6. Urology Department, South-Pest Teaching Hospital, Budapest, Hungary
7. Norwich Medical School, University of East Anglia, Norwich, UK
8. Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy
9. Department of Infectious Diseases, Hospital Clínic of Barcelona, University of Barcelona, IDIBAPS, CIBERINF, Ciber in infectious diseases ISCIII, Madrid, Spain
10. Department of Urology, Pediatric Urology and Andrology, Justus Liebig University Giessen, Giessen, Germany
11. Division for Mental and Physical Health, Norwegian Institute of Public Health, Oslo, Norway
12. Department of Infectious Diseases, CHU Rouen, Rouen, France; and INSERM UMR 1311 DYNAMICURE, Univ Rouen Normandie, Rouen, France.

Corresponding author:

Truls E. Bjerklund Johansen

Urology department, Institute of Clinical Medicine, University of Oslo, Norway and Institute of Clinical Medicine, University of Aarhus, Denmark

Address: PO Box 1171, Blindern, 0318 Oslo

Telephone: +47 91841063

Email: tebj@uio.no

Running head: Consensus on studies for antibiotics in cUTI

Key words: complicated urinary tract infections; FDA guidance; randomized controlled trials; guideline adherence; risk of bias assessment; classification

Summary

Complicated urinary tract infections (cUTI) denote an important research field for new antibiotics against Gram-negative pathogens. There is, however, increasing concern that this disease entity is too vaguely defined, leading to heterogeneous study populations and risk of bias.

We analysed researchers' adherence to the US Food and Drug Administration (FDA) guidance on cUTI and assessed risk of bias using a three-step procedure: literature review of cUTI papers; assessment of the relative importance of risk factors for treatment failure, including statistical evaluation of how patients with risk factors might skew treatment effects; and a Delphi consensus process in a multidisciplinary group.

Our evaluation showed poor adherence to FDA guidance on cUTI and significant heterogeneity in the reporting of study-, patient-, and pathogen-characteristics, leading to a high risk of bias when interpreting and comparing study findings. We therefore question the concept of cUTI as a meaningful entity with its own study guidance.

Introduction

Urinary tract infections (UTIs) are common, ranging in severity from simple, uncomplicated infections to life-threatening sepsis.[1] Urosepsis accounts for about one-third of all sepsis cases.[2] Conditions that increase the risk of acquiring UTI, predispose to a more severe disease course, and/or of treatment failure and a worse outcome, are called complicating factors or risk factors.[3]

Patients with severe or complicated UTI (cUTI) are often given initial intravenous (IV) antibiotic treatment,[4] and UTI is a much-used model infection for studying the efficacy of new antibiotics.[1] Important reasons for this are that the spectrum of pathogens is limited and largely known, and the presence or absence of pathogens in the urinary tract can be easily assessed in most cases, as opposed to gastrointestinal or pulmonary infections, where it can be more challenging to identify the causative agent(s).[1,5]

The concept of cUTI was introduced by the Infectious Disease Society of America in 1992 for the evaluation of new anti-infective drugs in clinical studies.[6] The meaning of "complicated" was, however, vaguely defined and, in 2010, the European Section for Infections in Urology suggested to avoid dividing UTI into complicated and uncomplicated and instead describe UTIs by clinical severity grade, phenotyping of risk factors, and pathogen characteristics.[7] In 2015, this classification was adopted by the European Association of Urology guidelines panel on UTI.[8]

Patient populations with cUTI are very heterogeneous and the effect of an antibiotic might appear greater if studied in a population with less severe conditions and fewer (or different) risk factors. To ensure unbiased evaluation of antibiotics, we need careful descriptions of factors that influence treatment outcomes, such as the clinical condition, the patient, and the pathogen.

In 2018, the FDA published new guidance for industry on developing drugs for cUTI.[9] cUTI was defined as a clinical syndrome characterized by pyuria and a documented microbial pathogen on culture of urine or blood, accompanied by local and systemic signs and symptoms, including flank pain, back pain, and/or costo-vertebral angle pain or tenderness, and fever, chills, and malaise, occurring in the presence of a functional or anatomical abnormality of the urinary tract or in the presence of catheterization.[9] As fever is not a strict inclusion criterion, there is a risk that patients with less severe infections without fever may be included. To counteract this, the FDA recommends that at least 30% of patients enrolled have acute pyelonephritis.[9] Conversely, there is no limitation in the percentage of the trial population who can have more severe infections, with an increasing Sequential Organ Failure Assessment (SOFA) score.[10,11] All risk factors mentioned in the FDA definition belong to the urological domain,[9] but with no differentiation of the type and duration of

stents and catheters, or the severity of stone disease. The FDA definition fails to consider patient risk factors such as history of recurrent UTI or extra-urogenital risk factors, such as diabetes mellitus and immune deficiency.

According to the FDA guidance the main outcomes of studies on new antibiotics should be resolution of symptoms and evidence for reduction of pathogens.[9] The microbiological criterion for treatment success of $<10^3$ colony forming units (CFU)/mL on urine culture[9] might be interpreted as $\leq 10^3$ or $\leq 10^2$ CFU/mL. There is even a possibility that microbiological success might be interpreted as total absence of detectable micro-organisms (bacteriological cure); in any event, much will depend on the detection limit of the microbiological tests employed.

The time points for assessing the effect of treatment are related to treatment duration, including a switch to oral medication, which varies by antibiotic. The FDA states that the primary endpoint should be assessed after ~5 days of IV therapy,[9] with success re-evaluated at a test of cure (TOC) visit at least 5 days after the end of treatment (EOT).[9] A late follow-up visit to assess the sustainability of effect is recommended 21–28 days after randomization.[9] Hence, there are numerous criteria that might be interpreted in different ways, leading to bias in the evaluation of new antibiotics.

The aim of the present paper is to advance the classification of patients and tests of cure in studies of antibiotics for cUTI. Our primary objective was to review the most important recent publications on antibiotic trials in cUTI and to analyse the interpretation of and adherence to the FDA guidance. Our secondary objective was to perform a multidisciplinary consensus process on the importance of the clinical condition, patient risk factors, and microbiological criteria for assessment of treatment success in antibiotic trials for cUTI.

Methods

Literature review

Establishment of study group

A working group of urologists, infectious diseases specialists, and microbiologists was established to discuss criteria for studies on cUTI. Participants were identified based on an internal healthcare professional tiering tool as those with significant experience in cUTI (appendix p 4). The group held an online meeting in October 2023, via a virtual collaboration tool (Within3, Lakewood, Ohio, US) (appendix p 4).

Search strategy and selection criteria

To evaluate recent studies on cUTI, we conducted a literature search (appendix pp 4–5) to identify publications of randomized controlled trials (RCTs) of cUTI treatments. A spreadsheet with the FDA criteria for cUTI[9] and risk factors defined by the group was developed, and 16 papers identified by the literature search were evaluated for adherence (appendix p 4). Risk factors were assessed using the ORENUC classification.[7]

Risk of bias assessment

Weighting of risk factors

The group used its virtual collaboration tool to identify those risk factors that are most likely to cause clinical and microbiological failure in studies of cUTI.

Statistical evaluation

A statistical evaluation was performed of the interrelationship between the difference in the number of patients with significant risk factors between study arms and p-values for differences in treatment effects between study arms. We modelled the impact of four key parameters on the p-value of falsely rejecting a null hypothesis stating a difference between treatment arms (appendix p 6). The parameters were: absolute difference in number of patients with treatment success; number of patients in each arm with no effect of treatment; and the non-inferiority level.[12-14]

Delphi process

A modified, accelerated Delphi process was performed to assess the likelihood of bias[15] in clinical and microbiological outcomes if the clinical condition, patient risk factors, and microbiological aspects are not adequately considered in studies of new antibiotics for cUTI. Based on findings from the literature analysis, 26 issues were identified, defined by five signalling questions and grouped into four categories (clinical situation, patient risk factors, pathogen-related aspects, and study characteristics) (appendix pp 7–8). For each issue, participants assessed the risk of bias (low risk,

some concerns, or high risk). For each category, participants also assessed the risk of bias if two or more issues in each category were not adequately considered. Finally, participants assessed the risk of bias when comparing outcomes between arms within a single study, and the risk of bias when comparing outcomes between studies. The risk of bias domains were the randomization process; the measurement of outcome; and the reporting of the outcome.[16] To inform their evaluation, the participants were asked to use the methods in the reviewed papers, any other relevant evidence, and their own expert opinion. Consensus was defined as >75% agreement on the degree of risk of bias.

Participants in the Delphi process

The original working group was expanded to a consensus group with 22 participants with research experience in the field of cUTI (20 European, two North American): 13 specialists in urology, seven in infectious diseases, and two in microbiology. Participants received no financial compensation for their time spent on the consensus process.

Online platform

The Delphi process was run on an electronic platform (Within3, Lakewood, Ohio, US).

Consensus rounds

Two consensus rounds were held (appendix p 9). The first-round results were sent to participants with a summary of their comments. Results were displayed for the whole group, and for urologists and infectious diseases specialists separately. To be able to demonstrate differences in evaluations, the consensus process was closed after the second round without further attempts to reach consensus on individual issues.

Role of the funding source

The funder had a role in establishing the initial study group and online meeting. Juan Quevedo, MD (an employee of the funder) was a member of the independent study group and contributed to data collection, literature analysis, manuscript review, and agreed with the decision of the group to submit the manuscript for publication. All authors endorsed the decision to submit for publication and agreed to be accountable for the work.

Results

Evaluation of studies

Study characteristics

Sixteen RCTs were evaluated (appendix pp 10–11). Fourteen studies compared different IV/oral antibiotics,[17-30] one compared different durations of antibiotic therapy,[31] and one compared bacteriophage versus antibiotics.[32] Four papers specified that the definition of cUTI was consistent with FDA and/or the European Medicines Agency (EMA) guidance,[20,23,26,29] and two stated that study conduct/design was generally in accordance with FDA and/or EMA guidance;[22,30] the rest did not refer to specific guidance or classifications published by medical societies.

Setting and clinical presentation

One study specified the clinical setting as a community-acquired infection;[31] no other studies reported if the infection was community- or hospital-acquired. None of the studies detailed the proportion of patients with cystitis at baseline and only one study showed the proportion of patients with sepsis at baseline.[30]

All but one study[26] included a history of symptoms of cystitis (dysuria, increased urinary frequency, urinary urgency, lower abdominal pain and/or pelvic pain) in the list of possible inclusion criteria that defined cUTI, and all but one study[32] mentioned symptoms of pyelonephritis (flank pain) as a possible criterion for cUTI.

Only two studies included clinical findings suggesting cystitis (suprapubic tenderness based on physical exam) in the list of possible criteria indicating cUTI.[22,27] Thirteen studies listed subcostal tenderness as an anamnestic criterion indicative of pyelonephritis[18-30] and 12 studies included general symptoms such as nausea and vomiting.[17-20,22-25,27-30] All studies except for one included fever as a possible criterion for cUTI; the exception[32] enrolled patients with non-febrile UTI, excluding patients with a temperature >38°C. The definition of fever and ways of measuring it varied between trials (six used >38°C;[19,20,24,28-30] five used ≥38°C;[21-23,26,31] one each used >38.5°C[17] and ≥38.5°C;[18] one used fever as defined by investigator[27]; one used oral temperature >37.5°C or axillary temperature >37°C[25]). Two studies mandated that participants must have fever (see appendix p 12 for details).[21,31]

The proportion of patients with pyelonephritis in the studies varied from 42–83% (n=12 studies)[17-20,22-24,26-30] (Figure 1).

Study endpoints

Most studies (n=10) used composite primary endpoints of clinical and microbiological success (appendix pp 10–11).[19,20,22-24,26,28-31] Clinical response was defined as symptom resolution/improvement, with some definitions including no need for further antibiotics. There was significant variation in treatment duration and the timing of TOC (Figure 2). Of 12 studies with IV antibiotics as the investigational treatment, the duration of IV therapy in the test arm was 3–9 days (based on data from 11 studies reporting the mean/median/protocol-specified duration of IV therapy).[17-19,21-24,26-30]

Among 13 studies that specified the timing of a TOC visit,[17-20,22-30] seven gave the timing relative to study start (range: 14–23 days after treatment start)[19,20,22,24,28-30] and six gave the timing relative to EOT (range: 5–15 days after EOT).[17,18,23,25-27] Among these 13 studies, 11 evaluated outcomes at a later timepoint than TOC: seven specified the follow-up visit timing relative to treatment start (range: 21–35 days after treatment start)[19,20,22,24,28-30] and four specified the timing relative to EOT (range: 14–47 days after EOT).[17,18,23,26] An additional three studies did not specify a TOC timepoint (see appendix p 12 for details).

Efficacy findings

Among 13 studies that presented primary endpoints assessing either clinical response, microbiological response, or a composite of both (Table 1), the number of patients in the study arms in the primary endpoint analysis sets ranged from 7–449. Five studies assessed superiority of the test drug versus comparator for seven primary endpoints[22,23,26,29,32] (one study assessed three co-primary endpoints[23]). The criteria required to demonstrate superiority were the lower bound of the 95% confidence interval being >0 (two studies[22,23]) or ≥0 (one study[29]) (criteria were not specified for two studies[26,32]). Superiority criteria were met for all four of the primary endpoints that were composites of clinical and microbiological response; the absolute difference between treatment groups varied from 4·5–21·2 percentage points.[22,23,26,29] For three primary endpoints assessing microbiological response, superiority criteria were not met.[23,32] Ten studies assessed non-inferiority of the test drug versus comparator on the primary endpoint (non-inferiority margins varied from 10–35%).[19,20,22-24,26,29-32] Additional details are presented in Table 1 and appendix p 12.

Discordant findings between clinical and microbiological response

Nine studies reported the treatment effect between study arms for both microbiological and clinical outcomes at the TOC visit for the same analysis set, allowing assessment of whether discordant findings occurred between the two outcomes (appendix p 13)[19,20,22-24,26,27,29,30]. Five of these nine

studies (55·6%) reported discordant findings,[19,22,26,29,30] with a significant difference between treatment groups in microbiological response but not in clinical outcome. Overall, five studies asserted to asymptomatic bacteriuria as the explanation for patients who have clinical, but not microbiological success.[19,20,23,29,30]

Patient characteristics

Age, sex and race

Two studies included paediatric patients,[17,27] with one of these enrolling newborn babies;[27] the rest solely included adults. Fourteen studies enrolled male and female patients; 12 stated that pregnant females were excluded,[17-25,27,29,30] two did not mention pregnancy in their exclusion criteria nor as a risk factor,[26,28] and two restricted enrolment to male patients.[31,32] Details on race are in appendix p 14.

Risk factors (ORENUC criteria)

No studies reported inclusion of patients without risk factors or reported using "no risk factors" as an exclusion criterion. No studies specified lifestyle factors as the cause of recurrent UTI. Nine studies excluded patients with kidney transplant[17-20,22-24,29,30] and only one publication specifically stated that patients with kidney transplant could be included.[26] Fourteen studies excluded patients with kidney failure.[17-27,29-31] All publications (except two)[28,31] reported urological risk factors as indicators of cUTI among study inclusion criteria. The proportion of patients enrolled with catheters varied from 1–38% (n=7 studies) (appendix, p 15).[17-19,26,27,30,32] Two studies considered patients with catheters as non-eligible for inclusion,[25,31] and a further seven did not provide data on catheters.[20-24,28,29] Additional details on the reporting of risk factors are in the appendix (p 14).

Pathogen characteristics

Spectrum

Escherichia coli was the most reported pathogen in general (range: 26–92% of participants) (Figure 3). Across studies, 0–7% of patients had *Pseudomonas aeruginosa* infection and 3–20% had *Klebsiella pneumoniae* infection.

Antimicrobial resistance

Among studies that provided relevant information, the proportion of participants/isolates with multidrug resistant pathogens ranged from 15–36% (n=5 studies) and the proportion with extended-spectrum beta-lactamase-producing pathogens ranged from 2–31% (n=9 studies) (Figure 4).

Definition of susceptibility

Criteria used for susceptibility testing were reported for nine studies, with six using criteria developed by the Clinical and Laboratory Standards Institute (CLSI),[20,24,25,27,29,30] one using European Committee on Antimicrobial Susceptibility Testing (EUCAST) criteria,[21] and two using both CLSI and EUCAST criteria.[18,23]

Definition of microbiological response

Twelve studies had a primary endpoint that included assessment of microbiological response, either in isolation or as part of a composite endpoint. The criteria for assessing microbiological response varied, with a reduction in pathogens to $<10^3$ CFU/mL used in five studies,[19,20,22,29,31] one study using $\leq 10^3$ CFU/mL,[28] four studies using $<10^4$ CFU/mL,[18,24,30,32] one study using $\leq 10^4$ CFU/mL[26] and one study using both $<10^3$ CFU/mL and $<10^4$ CFU/mL in co-primary endpoints with different definitions.[23] For some of these studies, the term "microbiological eradication" was used to describe microbiological response, but none used "eradication" with its dictionary definition of total absence of the pathogen. Authors typically did not state the detection limit of the microbiological testing method used.

Definition of sustainability

Nine studies evaluated microbiological response as a criterion for sustainability of treatment effect (appendix p 16),[17,20,22-24,26,28-30] using varied time-points to assess sustainability (range: Day 21–35 for studies specifying the timing relative to study start; 14–36 days after EOT for studies specifying the timing relative to EOT),[17,22] and varied criteria to define microbiological response (reduction in colony counts to $<10^3$ CFU/mL,[20,22,23,29] $\leq 10^3$ CFU/mL,[28] $<10^4$ CFU/mL,[23,24,30] or $\leq 10^4$ CFU/mL;[26] one study provided no definition beyond "eradication [*sic.*] of pathogens"[17]).

Evaluation of bias

Weighting of risk factors

The presence of an indwelling urinary catheter or stent was regarded as the most significant patient-related risk factor for treatment failure, followed by anatomical abnormalities of the urinary tract causing drainage problems, and urinary stones (appendix p 17).

Statistical modelling

Our statistical modelling indicated that if zero, five, or ten patients with significant risk factors for treatment failure were unequally included in one arm in a study with 200 patients per arm, the p-value for non-inferiority of the test drug rises from 1%, to 12%, and 50%, respectively (appendix p 18). If there are 100 patients in each study arm, the effect of unequally including zero, five, or ten

patients with risk factors for treatment failure is that the p-value rises from 5%, to 20%, and 50%, respectively.

Delphi process

Twenty-two participants answered the Delphi survey. An overview of the results is shown in Figure 5 with additional details by specialty and round in the appendix (pp 19–24).

Clinical situation

There was consensus on a high risk of bias if clinical presentation and severity, and fever, were poorly reported and/or unbalanced between study arms. There was also consensus on a high risk of bias if two or more issues in this category were not satisfactorily described/balanced. Most participants voted for high risk of bias if the setting (nosocomial- or community-acquired UTI) was not adequately described or balanced between arms. Half of participants voted for a high risk of bias related to the general condition of the patient.

Patient-related issues

There was consensus on a high risk of bias related to urinary catheters/stents and stones being poorly reported or unequally distributed between study arms, in addition to history of bacterial prostatitis, evidence of immune suppression or diabetes mellitus, and antibiotic treatment within the previous 30 days; consensus was not reached on the risk of bias related to poor description/distribution of patients with history of symptomatic UTI in the previous 6 months, female sex, premenopausal women with history of recurrent UTI associated with sexual intercourse, and a history of obstipation.

Pathogen-related issues

There was 100% consensus on a high risk of bias in case of poor reporting/distribution between study arms of resistance to study drug, and if neither the spectrum of pathogens nor the occurrence of drug resistance was satisfactorily described/balanced.

Study characteristics

When comparing findings between studies, there was consensus that a discrepancy between studies in the definition of microbiological “eradication” would introduce a high risk of bias. There was also a high percentage of votes (without reaching the consensus threshold) for a high risk of bias if there was a discrepancy between studies in microbiological success criteria of ≥10 CFU/mL (e.g., using $\leq 10^3$ instead of $\leq 10^4$ CFU/mL). The majority of participants voted for high risk of bias if there were differences between studies in the length of time from EOT to TOC (73%; consensus threshold not

reached), and from EOT to assessment of sustainability (consensus reached). There was consensus on a high risk of bias if two or more issues in this category were present.

All categories

There was consensus that the presence of at least one issue in each of the categories related to clinical situation, patient, and pathogen would lead to a high risk of bias when interpreting the study findings. Likewise, there was consensus that comparing outcomes between studies would be subject to a high risk of bias if at least one issue in all four categories (clinical situation, patient, pathogen, and study characteristics) was present.

Discussion

Main findings

The concept of 'cUTI' was introduced in recognition of patient and pathogen factors that increase the risk of UTI recurrence and treatment failure. It was meant to be a suitable entity for studies of new antibiotics targeting Gram-negative pathogens, and cUTI is now often the setting for Phase 3 licensing trials.

Unfortunately, when reviewing recent antibiotic RCTs in cUTI, we found significant variation in the reporting of: (i) the health setting(s) where cUTI developed, (ii) clinical presentation and severity, (iii) patient and pathogen characteristics, (iv) definitions of clinical and microbiological 'cure', and (v) study characteristics in general. Our multidisciplinary group reached consensus that there is a high risk of bias for intra- and, especially, inter- study comparison of outcomes if key characteristics are unsatisfactorily reported, unequally distributed between treatment arms, or not reported at all.

Overall, recent RCTs on new antibiotics in cUTI are so heterogeneous that the highest level of evidence, based on systematic review and meta-analysis, cannot be achieved. Accordingly, we question the impact of the current FDA research guidance and the concept of 'cUTI' as a meaningful entity with its own study guidance.

Context and impact

To our knowledge this is the first evaluation of licensing trials of new antibiotics in cUTI to have: (i) compared the key characteristics of the study populations and (ii) to use Delphi methodology to evaluate the risk of bias if these characteristics are not satisfactorily described, or not described at all. Numerous previous authors have questioned the validity of the concept of 'cUTI' and called for new definitions to ensure that new antibiotics are evaluated in more homogeneous populations.[33-36] In 2023, authors of a systematic review of studies on therapeutic and prophylactic interventions in adult UTI concluded that there is wide variation in clinical and microbiological criteria for diagnosing UTI.[34] They supported an earlier recommendation[7] to abandon the overarching concept of 'cUTI'. A review of recent pivotal cUTI trials highlighted that, even when study designs follow regulatory guidelines, significant variation remained across trial populations.[37]

We identified shortcomings and variation in multiple key endpoints used to define success and failure, including the duration of study drug treatment, the period from EOT to assessment of effect, and the period from EOT to assessment of sustainability. Clinical and microbiological criteria for diagnosis and treatment success were unclear and variable in terms of the number of surviving uropathogens allowed to remain in a 'microbiological success'. The dictionary definition of

'eradication' (i.e. complete removal) is far stricter than how this word is commonly applied to describing microbiological outcomes in cUTI. In this same context, many studies showed discordant clinical and microbiological success rates, questioning the appropriateness of 'composite' success criteria, and the equal weighting of clinical and microbiological outcomes in treatment results. Kadry *et al.*[38] found higher rates of clinical failure at late follow-up if there is a discordance between clinical and microbiological success at TOC, which is a clinically relevant observation.

Shortcomings in the reporting of the clinical findings used as inclusion criteria, inconsistent definitions of fever, and vague definitions of clinical success call for the use of objective criteria, such as defervescence, normalization of leucocyte count, C-reactive protein, and procalcitonin. The word 'eradication' should be replaced by criteria requiring reduction of bacterial count to below agreed and standardized thresholds. Reporting the spectrum of pathogens and the rate of resistance to study antibiotics should be mandatory. The results of RCTs are not only needed for the registration of new antibiotics but also guide subsequent clinical decisions. Overcoming limitations in study characteristics and patient classification will therefore also improve clinical practice.

Strengths and weaknesses

It might be argued that the reviewed studies were too diverse to allow comparison. Some were performed before the publication of the most recent FDA guidance.[9] However, our objective was to explore studies on 'cUTI' and we were interested in displaying how diversely investigators interpreted the concept. Another limitation is that, to keep the search manageable, we focused solely on studies with results published in peer-reviewed literature and did not search congress presentations nor the grey literature which, by definition, is not indexed nor readily searchable. We are aware e.g. of unpublished trials of antibiotics that failed to demonstrate efficacy in cUTI, notably eravacycline.[39] Finally, in respect to pathogen characteristics, we acknowledge these may have been reported in secondary publications excluded from our analysis.

Our way of assessing risk of bias differed from Cochrane methodology for systematic reviews. Cochrane recommends using five domains[16] assessed independently by two investigators, often junior researchers. We did not specify so many domains, but instead specified 26 issues, and our assessment was made by 22 experts from three medical specialties. We not only present the final Delphi results, but also display the full variation between rounds and specialties. In compliance with Cochrane, each issue was rated as 'low', 'some concerns' or 'high' for risk of bias. The assessors' reasons for judgement were presented in free text comments in a transparent, independent, and confidential way. In the first consensus round, urologists tended to vote for a higher chance of bias related to patient risk factors than did infectious diseases specialists. This difference likely reflects

urologists, within their day-to-day practice, working to identify and minimize factors related to recurrence (ORENUC categories R and U). The composition of the panel might therefore be a bias itself, and a differently composed research panel performing a wider search might produce better evidence-based recommendations than developed here. The Europe-centred perspective of our panel may have influenced our weighting of risk factors compared with colleagues elsewhere, reflecting difference in practice and awareness.

Despite these limitations we contend that our issue-based evaluation of bias provides original, sound, and valuable criticism to improve and standardise the conduct of clinical trials in cUTI.

Consequences and conclusions

A triad of issues in cUTI research need further evaluation: poor adherence to research guidance; high risk of bias in clinical studies; and 'cUTI' inherently being a poor concept.

Improving adherence to guidance

Developing clearer guidance will facilitate adherence. This should start with obtaining consensus-based agreement, via a carefully composed, international, multidisciplinary expert group, on numerical protocol criteria (e.g. treatment duration(s), number of days between treatment start and assessment of response or sustainability) and on measures defining treatment success (such as time to defervescence and acceptable colony count thresholds at TOC). There is also a need for a better classification of clinical severity of UTI and the ESIU/EAU definition is a good starting point.[7]

The responsibility for adherence to research guidance lies with researchers, but also with reviewers and editors who evaluate and publish reports. Adherence should be monitored continuously. Improved guidance will not only enhance the quality of future research and increase the comparability of cUTI trials but will also improve clinical practice, which is guided by trial results.

Reducing risk of bias

Patient-related risk factors are the most discriminatory criteria within present 'cUTI' trial populations, and the proportions of patients with different risk factors may bias outcomes. Recruitment to future studies should start with assessment of clinical presentation form and severity, followed by evaluation of the patients' risk profile, preferably by phenotyping according to the ORENUC classification.[7] Special attention should be given to patient-related risk factors for recurrence (ORENUC category R), as these patients often have lifestyle factors as their main risk factor, such as fluid intake, voiding habits, frequency of sexual activity, or obstipation. Unless such aspects are addressed, patients will return to a lifestyle with an increased risk of recurrence, which may skew assessment of the sustainability of antibiotic-achieved cure.

Knowledge gaps remain regarding the impact of risk factors on the rate of recurrence and risk of progression to more-severe UTIs. We need to know more about the roles of non-catheter biofilms that harbour bacteria, preventing treatment success, and the importance of an immunologically impaired host response. Both remain grey zones in UTI research.[40] Pathogen species, strain, and particular resistance genes do impact outcome. Here, technologies are emerging that allow full genomic characterization of organism(s) within one hour, directly from urine.[41] Microbiological assessment by such methods during trial recruitment will reduce pathogen-related risks for bias.

New ways to analyse the impact of risk factors on study outcomes could entail a retrospective analysis of a broader range of studies using various *a priori* classifications, or re-analysis of individual study outcomes stratified according to risk factors. A caveat is that recruitment of patients according to stricter phenotyping and microbiological criteria replaces one big pool of patients with numerous smaller pools, reducing recruitment capacity within each hospital. Moreover, smaller sub-groups within a study will limit statistical power. A possible route forwards would be for future studies to be run by large, international specialist groups sponsored by pharmaceutical companies, replacing the current model, where trials typically are run by pharmaceutical companies themselves.

The concept of cUTI

If the concept of 'cUTI' is abandoned, complex decision algorithms will be required to obtain homogenous patient groups for studies. A potential way forwards lies in developing algorithms that simultaneously weight numerous variables. This calls for machine learning and use of artificial intelligence, which is already being tested in clinical decision-making for UTI treatment.[42] Notably, researchers in the Serpens study[43] are developing a machine-learning tool to predict outcomes of urosepsis based on multiple variables, including ORENUC criteria. In the shorter term, researchers can increase homogeneity in study arms by focusing on pyelonephritis patients without known risk factors, as a clearly defined clinical entity. As knowledge gaps are filled, new groups of patients with risk factors can be defined and studied. This would be in line with the strategy of the Infectious Diseases Society of America (IDSA) and the American Urologist Association (AUA), which have distinct guidelines on asymptomatic bacteriuria, uncomplicated-, recurrent- and catheter-associated UTI, but not for 'cUTI'.[44-47]

Concluding remarks

We hope this paper will push regulatory agencies and international medical societies to prioritize the establishment of an expert panel with a mandate to improve definitions, standards and phenotyping of patients for the betterment of evaluations of new antibiotics in UTI. UTI remains an important study field particularly for antibiotics against Gram-negative pathogens due to its clinical frequency,

462 and the strong likelihood of identifying the causative pathogen and of documenting microbiological
463 treatment effect.[1] Medical societies and regulatory bodies owe developers and patients clear
464 guidance on how to perform clinical studies and to improve clinical practice.

465

Contributors

TEBJ chaired the study group. The literature assessment was made by TEBJ, JQ and the medical writer (RL). PHZ and TEBJ performed the statistical evaluation. TEBJ designed and moderated the Delphi process. All authors contributed to the Delphi process. TEBJ had access to and verified all underlying data. Manuscript writing was led by TEBJ. All authors contributed to the development of the manuscript by critically reviewing the content. All authors endorsed the decision to submit for publication and agreed to be accountable for the work. Members of the cUTI consensus group are listed in the appendix (p 3).

Declaration of interests

AJ has received speaker fees from bioMerieux, Serosep, and The British Medical Journal; consultancy fees from Advanz Pharma, GlaxoSmithKline, Pfizer, and Global Access Diagnostics; grants from Pfizer; support for meeting attendance from Eumedica and GSK; and is a trustee for the British Society for Antimicrobial Chemotherapy.

AO has participated in advisory boards hosted by Advanz Pharma; and received honoraria from Advanz Pharma, MSD, InfectoPharm, and Pfizer.

AS reports receiving grants from Pfizer and Gilead Sciences; consulting fees and honoraria for lectures from Pfizer, MSD, Angelini, Shionogi, Gilead, and Menarini; and support for meeting attendance from Pfizer.

BK has participated in advisory boards hosted by OM Pharma; and is a member of the European Association of Urology guidelines panel on Urological Infections and unpaid chair of the UTISOLVE research group.

CB has participated in advisory boards hosted by Advanz Pharma, MiP Pharma, and GSK; and received honoraria for lectures from GSK and Mundipharma.

CGJ has received honoraria for lectures from Astellas Pharma, IBSA, and Medtronic; support for meeting attendance from Medtronic and IBSA; and participated in advisory boards hosted by Astellas Pharma, Innocon, and Aarhus Medical.

DML reports grants from Pfizer; personal fees from Adjutec, ADVANZ Pharma, AstraZeneca, bioMérieux, Centauri, GenPax, ParaPharm, Pfizer, Shionogi, Sumitovant, Summit, Thermofisher, Zambon, and Zuellig; shareholdings or options from GenPax, GSK, Merck, and Revvity; and nominated holdings in Arecor Therapeutics, Celadon Pharmaceuticals, Destiny Pharma, Genedrive, Genincode, Oxford BioDynamics, Optibiotix Health, Probiotix Health, SkinBiotherapeutics, and VericiDx under Enterprise Investment Schemes with no direct authority to trade holdings.

FJMN has received consulting fees from Meiji; honoraria from Advanz Pharma, Angelini Pharma, Gilead, GSK, Pfizer, Viatris, and ViiV; payment for expert testimony from Merck Sharp and Dome; support for meeting attendance from Advanz pharma, Gilead, Merck Sharp and Dome, and Pfizer; participated in advisory boards hosted by AstraZeneca, Advanz Pharma, Baxter, and Meiji; and has leadership roles for the Spanish Society of Infectious Diseases and Clinical Microbiology, the Latin-American Alliance on Infectious Diseases and Clinical Microbiology, and the Foundation SEIMC-GESIDA.

FW has participated in advisory boards hosted by VenatoRx and is a member of the UTISOLVE Research Group.

JCN reports being a consultant for Inmunotek and OM Pharma; and an invited speaker for Grand Rounds in Urology.

JK has received consulting fees from Bionorica and GSK; honoraria for presentations from Apogepha Arzneimittel GmbH, Bionorica, GSK, and Janssen Cilag GmbH; support for meeting attendance from Apogepha Arzneimittel GmbH and Janssen Cilag GmbH; has participated in advisory boards hosted by Bionorica, GSK, and Shionogi; and holds leadership/membership roles for the European Association of Urology, the German Society of Urology, and the UTISOLVE Research Group.

JMP has received speakers honoraria and support for meeting attendance from Astellas Pharma, Boston Scientific, Coloplast, Gebro Pharma, GSK, Lacer, Pierre Fabre, Qpharma, and Wellspect.

KGN is a consultant of Adamed Pharma, Bionorica, BioMerieux, GlaxoSmithKline, Inmunotek, Ingenion Medical, Johnson & Johnson, OM Pharma, and MIP/Rosen Pharma.

KSK has served as a consultant for GSK, Merck, Shionogi, Abbvie, Spero, Carb-X, and Biomeme; and participated in Data Safety Monitoring Board meetings for Venato-Rx, Meiji, and Contrafect.

ME reports honoraria for ad hoc expertise from Advanz Pharma and CORREVIO; and travel grants from ViiV Healthcare, MSD, Pfizer, and Gilead.

ML reports payment for manuscript writing from Pfizer.

TEBJ reports consultancy fees from Advanz Pharma (for chairing the cUTI consensus group and research for this manuscript), Sandoz, and Zambon; participation in advisory board meetings for Advanz Pharma and Zambon; and unpaid leadership roles with ESIU/EAU and UTISOLVE.

ZT is chair of the European Section of Infection in Urology/European Association of Urology and member of the UTISOLVE research group on UTI.

The other authors declared no conflicts of interest.

Acknowledgements

The authors would like to thank Juan Quevedo (JQ), MD, from ADVANZ PHARMA for his contribution to the project and Lucy Sims of Within3 for technical support with the online platform and Delphi process, funded by ADVANZ PHARMA. Medical writing support was provided by Ruth Lloyd (RL), PhD, of Herdwick Pharma Ltd. and funded by ADVANZ PHARMA.

Funding

This work was supported by ADVANZ PHARMA.

References

1. Wagenlehner FME, Bjerklund Johansen TE, Cai T, et al. Epidemiology, definition and treatment of complicated urinary tract infections. *Nat Rev Urol* 2020; **17**: 586-600.
2. Wagenlehner FME, Pilatz A, Weidner W, Naber KG. Urosepsis: Overview of the Diagnostic and Treatment Challenges. *Microbiol Spectr* 2015; **3**.
3. Bjerklund Johansen TE BH, Cek M, Grabe M, Tenke P, Wagenlehner FME, Naber KG. Critical review of current definitions of urinary tract infections and proposal of an EAU/ESIU classification system. In: Naber KG SA, Heyns CF, Matsumoto T, Shoskes DA and Bjerklund Johansen TE, ed. Urogenital Infections Arnheim: International Consultation on Urological Diseases (ICUD) and European Association of Urology 2010: 979-93.
4. European Association of Urology. Guidelines on Urological Infections. 2024. https://d56bochluxqnz.cloudfront.net/documents/full-guideline/EAU-Guidelines-on-Urological-Infections-2024.pdf (accessed December 17th 2024).
5. Burd EM, Hinrichs BH. Gastrointestinal Infections. *Molecular pathology in clinical practice* 2015; **19**: 707-34.
6. Rubin RH, Shapiro ED, Andriole VT, Davis RJ, Stamm WE. Evaluation of new anti-infective drugs for the treatment of urinary tract infection. Infectious Diseases Society of America and the Food and Drug Administration. *Clin Infect Dis* 1992; **15 Suppl 1**: S216-27.
7. Bjerklund Johansen TE, Botto H, Cek M, et al. Critical review of current definitions of urinary tract infections and proposal of an EAU/ESIU classification system. *Int J Antimicrob Agents* 2011; **38 Suppl**: S64-70.
8. European Association of Urology. Guidelines on Urological Infections. 2015. https://d56bochluxqnz.cloudfront.net/documents/EAU-Guidelines-Urological-Infections-2015.pdf (accessed December 12th 2024).
9. U.S. Food and Drug Administration. Complicated Urinary Tract Infections: Developing Drugs for Treatment - Guidance for Industry. 2018. https://www.fda.gov/media/71313/download (accessed February 5th 2024).
10. Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 2001; **286**: 1754-8.
11. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; **22**: 707-10.
12. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006; **145**: 62-9.
13. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, Group C. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006; **295**: 1152-60.
14. Fleming TR. Current issues in non-inferiority trials. *Stat Med* 2008; **27**: 317-32.
15. Cohchrane. RoB 2: A revised Cochrane risk-of-bias tool for randomized trials. 2024. https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials (accessed November 22nd 2024).
16. Higgins JPT SJ, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial [last updated October 2019]. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Cochrane, 2024. Available from www.training.cochrane.org/handbook. 2024.
17. Bradley JS, Roilides E, Broadhurst H, et al. Safety and efficacy of ceftazidime-avibactam in the treatment of children ≥3 months to <18 years with complicated urinary tract infection: Results from a phase 2 randomized, controlled trial. *Pediatr Infect Dis J* 2019; **38**: 920-8.
18. Connolly LE, Riddle V, Cebrik D, Armstrong ES, Miller LG. A multicenter, randomized, double-blind, phase 2 study of the efficacy and safety of plazomicin compared with levofloxacin in the

586 treatment of complicated urinary tract infection and acute pyelonephritis. *Antimicrob Agents*
587 *Chemother* 2018; **62**: e01989-17.
588 19. Dunne MW, Aronin SI, Das AF, et al. Sulopenem for the Treatment of Complicated Urinary
589 Tract Infections Including Pyelonephritis: A Phase 3, Randomized Trial. *Clin Infect Dis* 2023; **76**: 78-88.
590 20. Eckburg PB, Muir L, Critchley IA, et al. Oral Tebipenem Pivoxil Hydrobromide in Complicated
591 Urinary Tract Infection. *N Engl J Med* 2022; **386**: 1327-38.
592 21. Edlund C, Ternhag A, Skoog Ståhlgren G, et al. The clinical and microbiological efficacy of
593 temocillin versus cefotaxime in adults with febrile urinary tract infection, and its effects on the
594 intestinal microbiota: a randomised multicentre clinical trial in Sweden. *Lancet Infect Dis* 2022; **22**:
595 390-400.
596 22. Kaye KS, Belley A, Barth P, et al. Effect of Cefepime/Enmetazobactam vs
597 Piperacillin/Tazobactam on Clinical Cure and Microbiological Eradication in Patients With
598 Complicated Urinary Tract Infection or Acute Pyelonephritis: A Randomized Clinical Trial. *JAMA* 2022;
599 **328**: 1304-14.
600 23. Kaye KS, Bhowmick T, Metallidis S, et al. Effect of Meropenem-Vaborbactam vs Piperacillin-
601 Tazobactam on Clinical Cure or Improvement and Microbial Eradication in Complicated Urinary Tract
602 Infection: The TANGO I Randomized Clinical Trial. *JAMA* 2018; **319**: 788-99.
603 24. Kaye KS, Rice LB, Dane AL, et al. Fosfomycin for Injection (ZTI-01) Versus Piperacillin-
604 tazobactam for the Treatment of Complicated Urinary Tract Infection Including Acute Pyelonephritis:
605 ZEUS, A Phase 2/3 Randomized Trial. *Clin Infect Dis* 2019; **69**: 2045-56.
606 25. Li Y, Yin Y, Peng X, et al. A randomized, active-controlled, multicentre clinical trial to evaluate
607 the efficacy and safety of oral sitafloxacin versus levofloxacin in Chinese adults with acute
608 uncomplicated or complicated urinary tract infection. *Ann Med* 2021; **53**: 217-26.
609 26. Portsmouth S, van Veenhuyzen D, Echols R, et al. Cefiderocol versus imipenem-cilastatin for
610 the treatment of complicated urinary tract infections caused by Gram-negative uropathogens: a
611 phase 2, randomised, double-blind, non-inferiority trial. *Lancet Infect Dis* 2018; **18**: 1319-28.
612 27. Roilides E, Ashouri N, Bradley JS, et al. Safety and Efficacy of Ceftolozane/Tazobactam Versus
613 Meropenem in Neonates and Children With Complicated Urinary Tract Infection, Including
614 Pyelonephritis: A Phase 2, Randomized Clinical Trial. *Pediatr Infect Dis J* 2023; **42**: 292-8.
615 28. Wagenlehner F, Nowicki M, Bentley C, et al. Explorative Randomized Phase II Clinical Study of
616 the Efficacy and Safety of Finafloxacin versus Ciprofloxacin for Treatment of Complicated Urinary
617 Tract Infections. *Antimicrob Agents Chemother* 2018; **62**.
618 29. Wagenlehner FM, Gasink LB, McGovern PC, et al. Cefepime-Taniborbactam in Complicated
619 Urinary Tract Infection. *N Engl J Med* 2024; **390**: 611-22.
620 30. Wagenlehner FME, Cloutier DJ, Komirenko AS, et al. Once-Daily Plazomicin for Complicated
621 Urinary Tract Infections. *N Engl J Med* 2019; **380**: 729-40.
622 31. Lafaurie M, Chevret S, Fontaine JP, et al. Antimicrobial for 7 or 14 Days for Febrile Urinary
623 Tract Infection in Men: A Multicenter Noninferiority Double-Blind, Placebo-Controlled, Randomized
624 Clinical Trial. *Clin Infect Dis* 2023; **76**: 2154-62.
625 32. Leitner L, Ujmajuridze A, Chanishvili N, et al. Intravesical bacteriophages for treating urinary
626 tract infections in patients undergoing transurethral resection of the prostate: a randomised,
627 placebo-controlled, double-blind clinical trial. *Lancet Infect Dis* 2021; **21**: 427-36.
628 33. Nelson Z, Tarik Aslan A, Beahm NP, et al. Guidelines for the Prevention, Diagnosis, and
629 Management of Urinary Tract Infections in Pediatrics and Adults: A WikiGuidelines Group Consensus
630 Statement. *JAMA Netw Open* 2024; **7**: e2444495.
631 34. Bilsen MP, Jongeneel RMH, Schneeberger C, et al. Definitions of urinary tract infection in
632 current research: A systematic review. *Open Forum Infect Dis* 2023; **10**: ofad332.
633 35. Bilsen MP, Conroy SP, Schneeberger C, et al. A reference standard for urinary tract infection
634 research: a multidisciplinary Delphi consensus study. *Lancet Infect Dis* 2024; **24**: e513-e21.

635 36. Bonkat G, Wagenlehner F, Kranz J. Keep it Simple: A proposal for a new definition of
636 uncomplicated and complicated urinary tract infections from the EAU Urological Infections
637 Guidelines Panel. *Eur Urol* 2024; **86**: 195-7.
638 37. Portsmouth S, Bass A, Echols R, Tillotson G. Heterogeneity of Recent Phase 3 Complicated
639 Urinary Tract Infection Clinical Trials. *Open Forum Infect Dis* 2021; **8**: ofab045.
640 38. Kadry N, Natarajan M, Bein E, Kim P, Farley J. Discordant Clinical and Microbiological
641 Outcomes Are Associated With Late Clinical Relapse in Clinical Trials for Complicated Urinary Tract
642 Infections. *Clin Infect Dis* 2023; **76**: 1768-75.
643 39. Tetraphase Pharmaceuticals, Inc. U.S. Prescribing Information: XERAVA® (eravacycline). 2024.
644 https://xerava-assets.s3.amazonaws.com/prescribing-information.pdf (accessed December 19th
645 2024).
646 40. Bjerklund Johansen TE, Bonkat G, Cai T, Tandogdu Z, Wagenlehner F, Grabe M. Grey Zones in
647 the Field of Urinary Tract Infections. *Eur Urol Focus* 2016; **2**: 460-2.
648 41. Bellankimath AB, Chapagain C, Branders S, et al. Culture and amplification-free nanopore
649 sequencing for rapid detection of pathogens and antimicrobial resistance genes from urine. *Eur J Clin*
650 *Microbiol Infect Dis* 2024; **43**: 2177-90.
651 42. Cai T, Anceschi U, Prata F, et al. Artificial Intelligence Can Guide Antibiotic Choice in Recurrent
652 UTIs and Become an Important Aid to Improve Antimicrobial Stewardship. *Antibiotics (Basel)* 2023;
653 **12**: 375.
654 43. Tandogdu Z, Koves B, Ristovski S, et al. Urosepsis 30-day mortality, morbidity, and their risk
655 factors: SERPENS study, a prospective, observational multi-center study. *World J Urol* 2024; **42**: 314.
656 44. Bixler BR, Anger JT. Updates to Recurrent Uncomplicated Urinary Tract Infections in Women:
657 AUA/CUA/SUFU Guideline. *J Urol* 2022; **208**: 754-6.
658 45. Gupta K, Hooton TM, Naber KG, et al. International clinical practice guidelines for the
659 treatment of acute uncomplicated cystitis and pyelonephritis in women: A 2010 update by the
660 Infectious Diseases Society of America and the European Society for Microbiology and Infectious
661 Diseases. *Clin Infect Dis* 2011; **52**: e103-20.
662 46. Hooton TM, Bradley SF, Cardenas DD, et al. Diagnosis, prevention, and treatment of catheter-
663 associated urinary tract infection in adults: 2009 International Clinical Practice Guidelines from the
664 Infectious Diseases Society of America. *Clin Infect Dis* 2010; **50**: 625-63.
665 47. Nicolle LE, Gupta K, Bradley SF, et al. Clinical Practice Guideline for the Management of
666 Asymptomatic Bacteriuria: 2019 Update by the Infectious Diseases Society of America. *Clin Infect Dis*
667 2019; **68**: e83-110.

668

669 # Tables and Figures

670

671 **Table 1: Efficacy findings for studies with primary endpoints assessing clinical, microbiological or composite response**

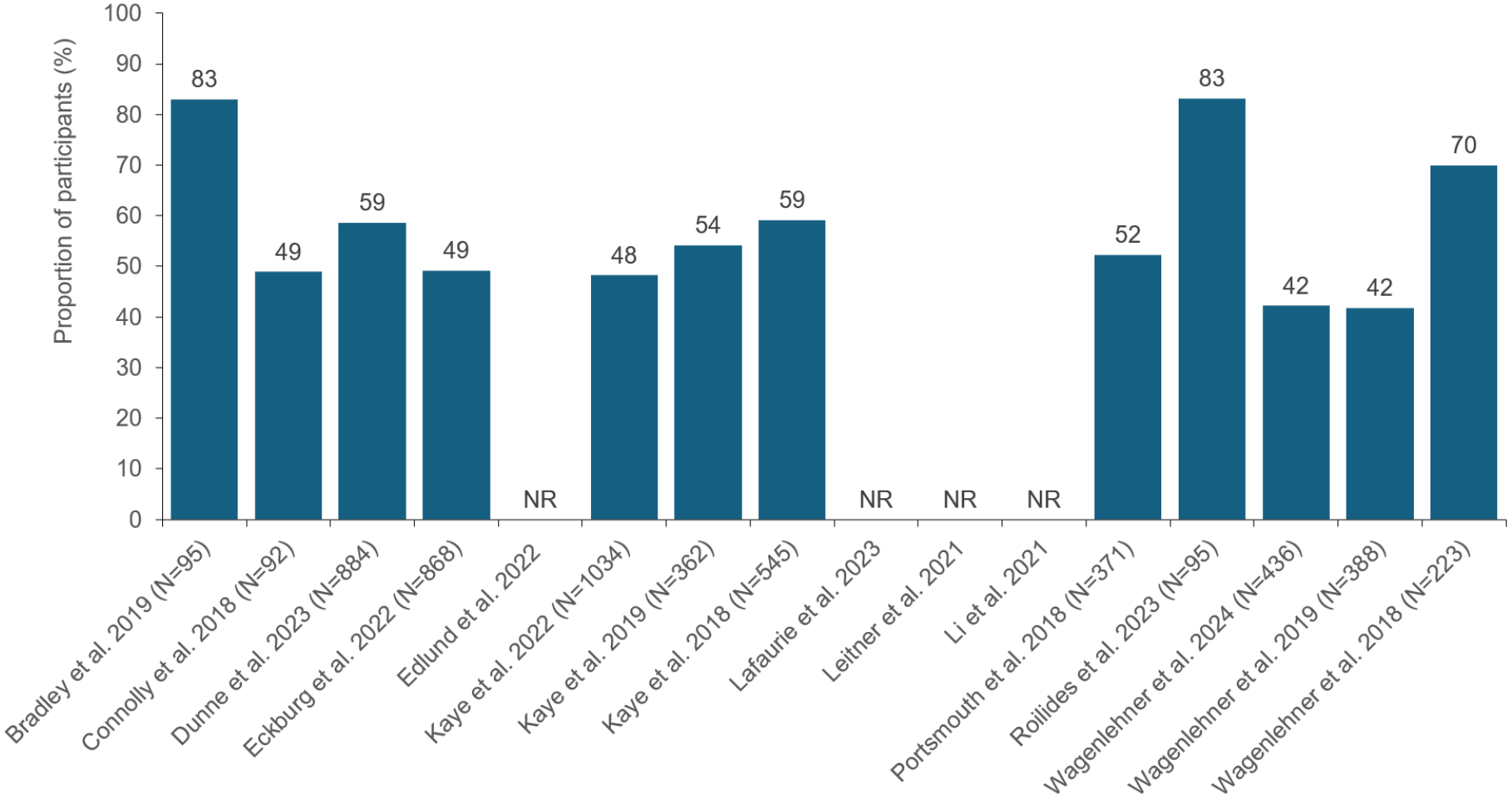
Study	Endpoint	Analysis set	Timepoint	Test	Comparator	Combined clinical and microbiological response, n/N (%)		Clinical response, n/N (%)		Microbiological response, n/N (%)		Treatment difference, percentage points (95% CI)*	Inferiority or superiority criteria met
						Test	Comparator	Test	Comparator	Test	Comparator		
Connolly et al. 2018[18]	Co-primary endpoint	mITT	TOC	Plazomicin 15 mg/kg IV	Levofloxacin IV	NA	NA	NA	NA	31/51 (61%)	17/29 (59%)	2.2 (–22·9, 27·2)	NA
				Plazomicin 10 mg/kg IV	Levofloxacin IV	NA	NA	NA	NA	6/12 (50%)	17/29 (59%)	NA	NA
	Co-primary endpoint	ME	TOC	Plazomicin 15 mg/kg IV	Levofloxacin IV	NA	NA	NA	NA	31/35 (89%)	17/21 (81%)	7.6 (–16·0, 31·3)	NA
				Plazomicin 10 mg/kg IV	Levofloxacin IV	NA	NA	NA	NA	6/7 (86%)	17/21 (81%)	NA	NA
Dunne et al. 2023[19]	Primary endpoint	Micro-mITT	TOC	Sulopenem IV†	Ertapenem IV†	301/444 (68%)	325/440 (74%)	NA	NA	NA	NA	–6.1 (–12·0, –0·1)	Non-inferiority criteria not met
Eckburg et al. 2022[20]	Primary endpoint	Micro-ITT	TOC	Tebipenem pivoxil hydrobromide (oral)	Ertapenem IV	264/449 (59%)	258/419 (62%)	NA	NA	NA	NA	–3.3 (–9·7, 3·2)	Non-inferiority criteria met
Kaye et al. 2022[22]	Primary endpoint	Micro-mITT	TOC	Cefepime/enmetazobactam IV	Piperacillin/tazobactam IV	273/345 (79%)	196/333 (59%)	NA	NA	NA	NA	21·2 (14·3, 27·9)	Non-inferiority and superiority criteria met
Kaye et al. 2019[24]	Primary endpoint	Micro-mITT	TOC	ZTI-01 IV	Piperacillin/tazobactam IV	119/184 (65%)	97/178 (54%)	NA	NA	NA	NA	10·2 (–0·4, 20·8)	Non-inferiority criteria met

Study	Endpoint	Analysis set	Timepoint	Test	Comparator	Combined clinical and microbiological response, n/N (%)		Clinical response, n/N (%)		Microbiological response, n/N (%)		Treatment difference, percentage points (95% CI)*	Inferiority or superiority criteria met
						Test	Comparator	Test	Comparator	Test	Comparator		
Kaye et al. 2018[23]	Primary end point for FDA	Micro-mITT	End of IV treatment	Meropenem-vaborbactam IV	Piperacillin-tazobactam IV	189/192 (98%)	171/182 (94%)	NA	NA	NA	NA	4.5 (0·7, 9·1)	Non-inferiority and superiority criteria met
	Co-primary endpoint for EMA	Micro-mITT	TOC	Meropenem-vaborbactam IV	Piperacillin-tazobactam IV	NA	NA	NA	NA	128/192 (67%)	105/182 (58%)	9.0 (–0·9, 18·7)	Non-inferiority criteria met; superiority criteria not met
	Co-primary endpoint for EMA	ME	TOC	Meropenem-vaborbactam IV	Piperacillin-tazobactam IV	NA	NA	NA	NA	118/178 (66%)	102/169 (60%)	5.9 (–4·2, 16·0)	Non-inferiority criteria met; superiority criteria not met
Lafaurie et al. 2023[31]	Primary endpoint‡	ITT	Week 6	7 days of antibiotics	14 days of antibiotics	64/115 (56%)	97/125 (78%)	NA	NA	NA	NA	–21·9 (–33·3, –10·1)	Inferiority criteria met
Leitner et al. 2021[32]	Primary endpoint	mITT	EOT or withdrawal	Intravesical pyophage	SoC antibiotics	NA	NA	NA	NA	5/28 (18%)	13/37 (35%)	Odds ratio (95% CI)§: 2·66 (0·79, 8·82)	Non-inferiority criteria met
				Intravesical pyophage	Intravesical placebo	NA	NA	NA	NA	5/28 (18%)	9/32 (28%)	Odds ratio (95% CI)§: 1·60 (0·45, 5·71)	Superiority criteria not met
Li et al. 2021[25]	Primary endpoint¶	Per-protocol	TOC	Sitafloxacin oral	Levofloxacin oral	NA	NA	27/33 (82%)	20/26 (77%)	NA	NA	4·9 (–16·0, 25·8)	NA

Study	Endpoint	Analysis set	Timepoint	Test	Comparator	Combined clinical and microbiological response, n/N (%)		Clinical response, n/N (%)		Microbiological response, n/N (%)		Treatment difference, percentage points (95% CI)*	Inferiority or superiority criteria met
						Test	Comparator	Test	Comparator	Test	Comparator		
Portsmouth et al. 2018[26]	Primary endpoint	mITT	TOC	Cefiderocol IV	Imipenem-cilastatin IV	183/252 (73%)	65/119 (55%)	NA	NA	NA	NA	18·58 (8·23, 28·92)	Non-inferiority criteria met and post hoc superiority criteria met
Wagenlehner et al. 2024[29]	Primary endpoint	Micro-ITT	TOC	Cefepime-taniborbactam IV	Meropenem IV	207/293 (71%)	83/143 (58%)	NA	NA	NA	NA	12·6 (3·1, 22·2)	Non-inferiority and superiority criteria met
Wagenlehner et al. 2019[30]	Co-primary endpoint	Micro-mITT	Day 5	Plazomicin IV	Meropenem IV	168/191 (88%)	180/197 (91%)	NA	NA	NA	NA	–3·4 (–10·0, 3·1)	Non-inferiority criteria met
	Co-primary endpoint	Micro-mITT	TOC	Plazomicin IV	Meropenem IV	156/191 (82%)	138/197 (70%)	NA	NA	NA	NA	11·6 (2·7, 20·3)	Non-inferiority criteria met
Wagenlehner et al. 2018[28]	Primary endpoint	Micro-ITT	TOC	5 days finafloxacin	10 days ciprofloxacin	45/64 (70%)	35/61 (57%)	NA	NA	NA	NA	NA	NA
				10 days finafloxacin	10 days ciprofloxacin	46/68 (68%)	35/61 (57%)	NA	NA	NA	NA	NA	NA

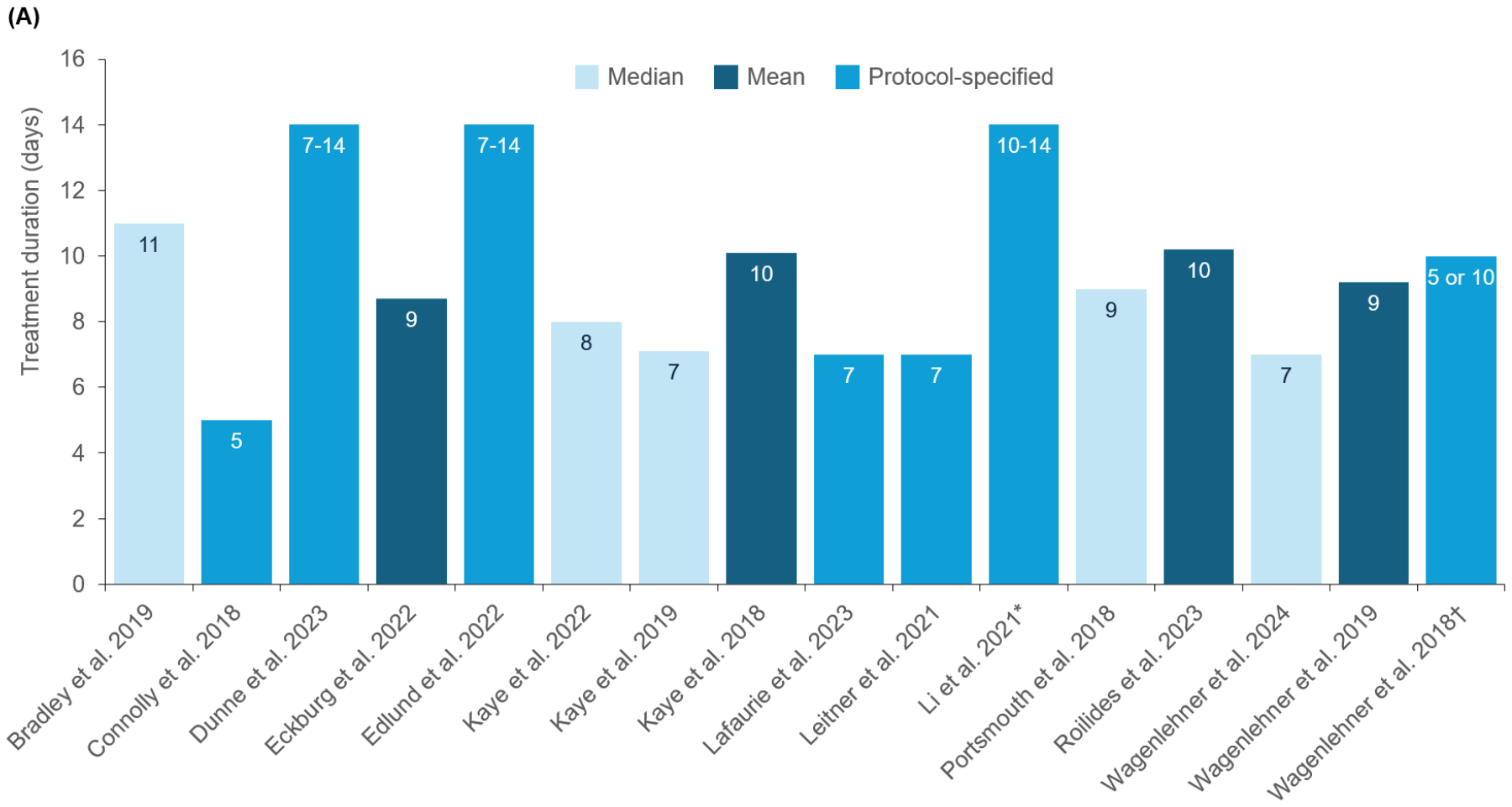
CI=confidence interval. ITT=intention-to-treat. IV=intravenous. ME=microbiologically evaluable. Micro=microbiological. mITT=modified intention-to-treat. NA=not assessed. SoC=standard of care. TOC=test of cure. UTI=urinary tract infection. *Unless otherwise indicated, treatment effect is treatment difference in percentage points and 95% confidence interval. †Sulopenem IV followed by oral sulopenem etzadroxil/probenecid or ertapenem IV followed by oral ciprofloxacin or amoxicillin-clavulanate. ‡Treatment success (clinical success, microbiological success, and absence of new antimicrobial treatment since the end of the antibiotic treatment for UTI [except if a new antimicrobial was prescribed for another infection and had no effect on the initial uropathogen]). §Adjusted logistic regression with pyophage as the reference. ¶Data shown for the subgroup of UTI patients with complicated UTI.

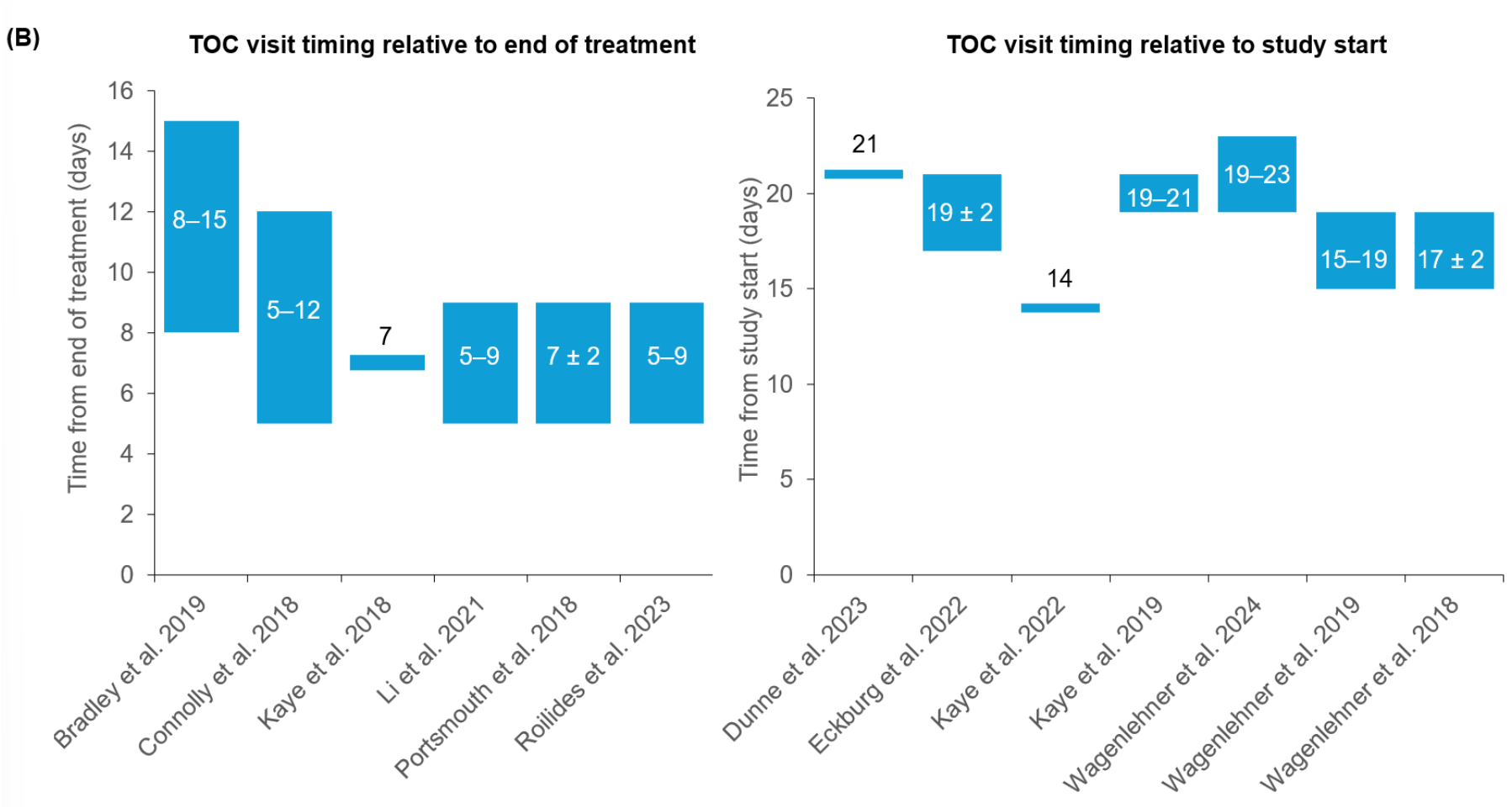
Figure 1: Proportion of participants with pyelonephritis



Data are presented for all treatment groups combined. Analysis sets used to report data varied between studies. NR=not reported.

Figure 2: Treatment duration (A) and timing of test of cure visit (B)





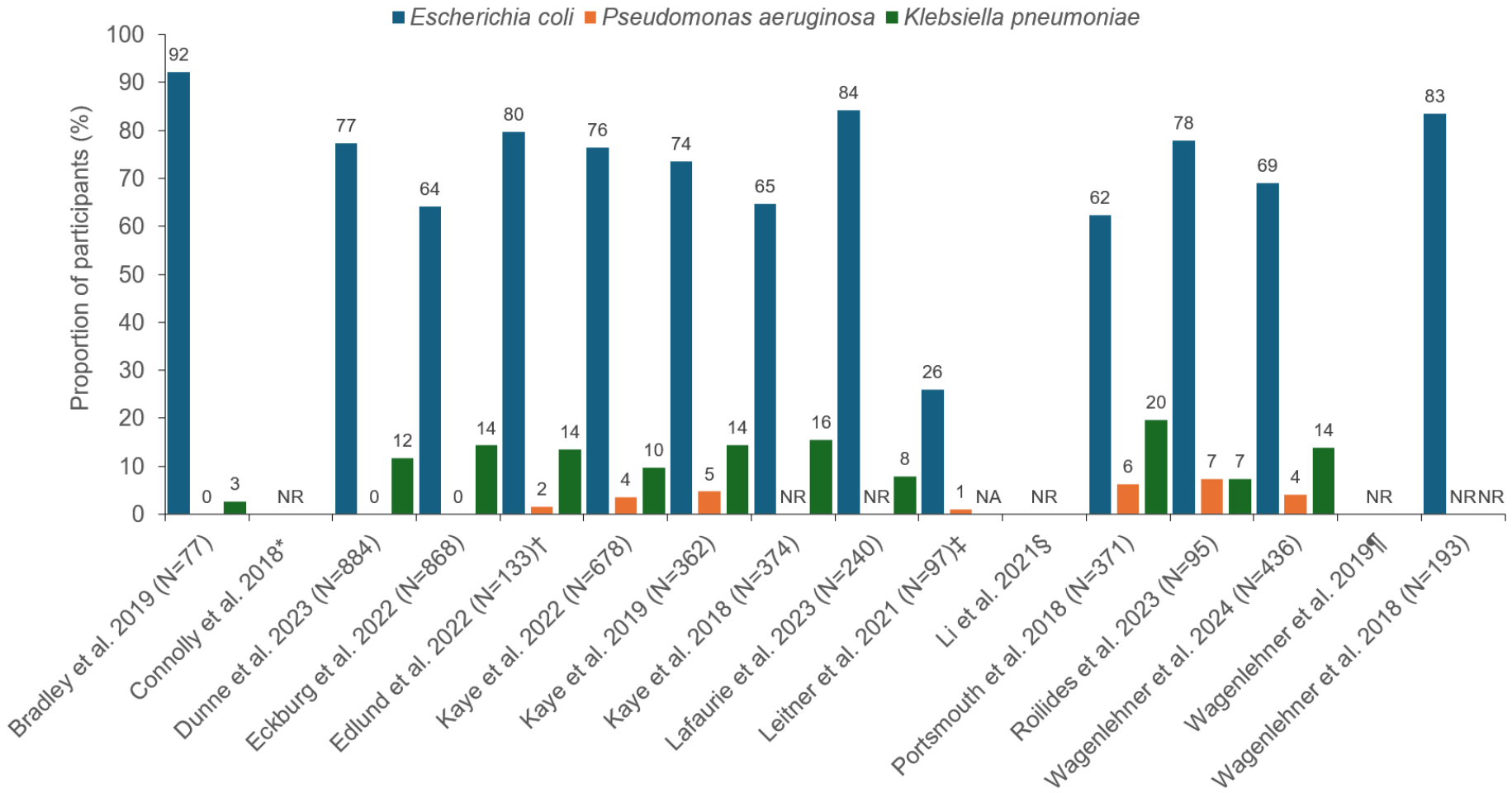
683

684 (A) Analysis sets used to report data varied between studies. Treatment durations are for both arms combined if available,[18,19,22,25,32] and for the
685 investigational treatment arm in all other cases; the duration is for all treatments (e.g., including both intravenous and oral step-down therapy, where this
686 was permitted). Durations are mean, median or protocol specified as indicated (protocol-specified durations are shown if the median/mean total duration of
687 all study treatments combined [including oral step-down] was not provided, or the studies had a fixed treatment duration). (B) TOC visit timing was specified

688 according to end of treatment or study start. Three studies are not shown as they did not specify timing of a TOC visit.[21,31,32] cUTI=complicated urinary tract
689 infection. TOC=test of cure. *For the subgroup with cUTI. †Treatment duration with investigational drug was 5 days in one arm and 10 days in another arm.

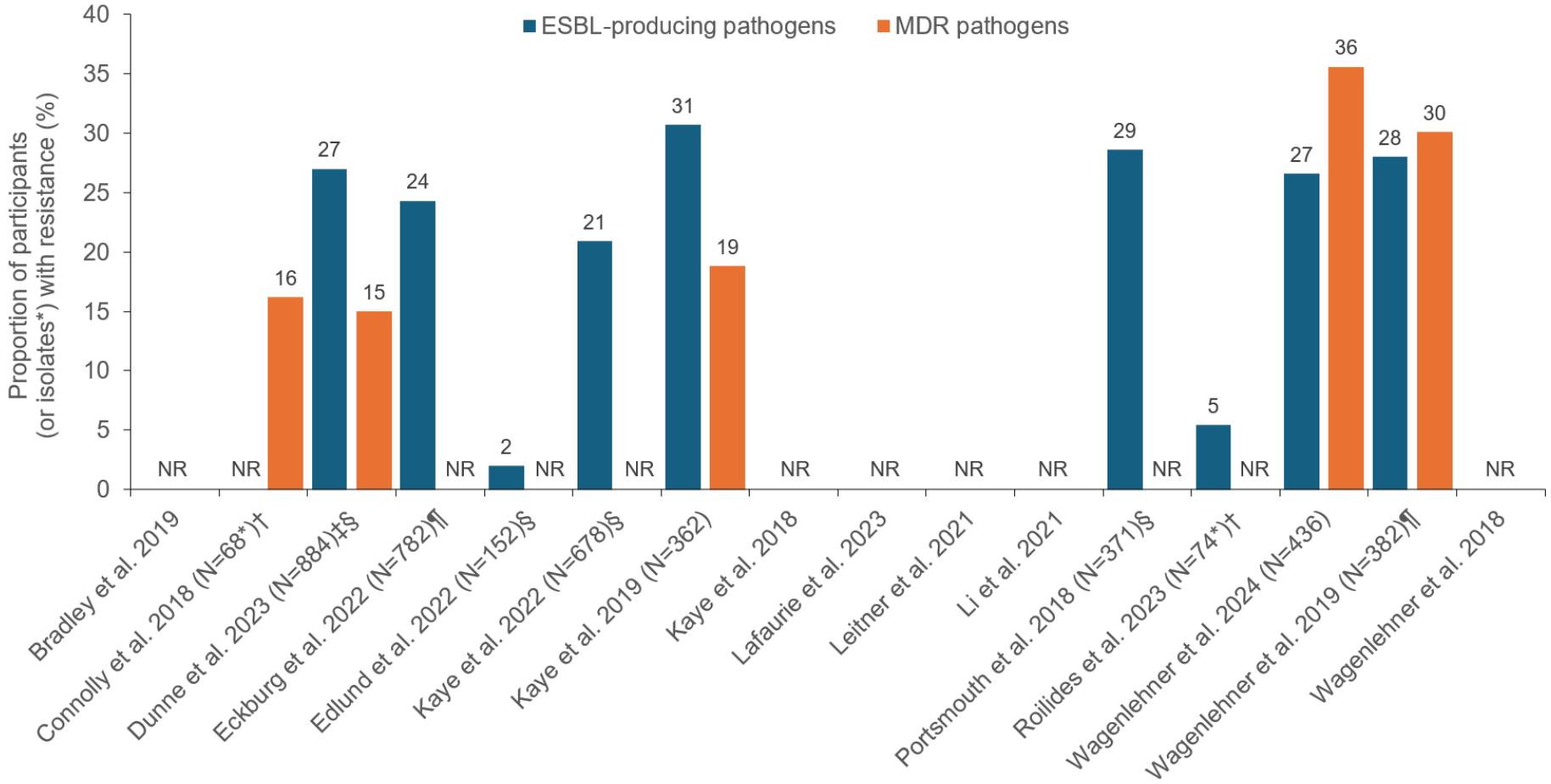
690

Figure 3: Proportion of participants with *Escherichia coli*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* pathogens



Data are presented for all treatment groups combined. Analysis sets used to report data on pathogens varied between studies (all included only patients who had pathogens isolated at baseline). NA=not applicable. NR=not reported. UTI=urinary tract infection. *69 *Escherichia coli* and 7 *Klebsiella pneumoniae* isolates reported among 92 patients. †Data shown for all *Klebsiella* spp. and *Pseudomonas* spp. ‡To be eligible for inclusion, patients had urine cultures that were positive for pathogens covered by the pyophage cocktail investigated in this study (i.e., *Enterococcus* spp., *E. coli*, *Proteus mirabilis*, *Pseudomonas aeruginosa*, *Staphylococcus* spp., and *Streptococcus* spp.); patients with other pathogens typical in UTI, such as *Klebsiella* spp., were not eligible. §23 *E. coli* isolates reported among 26 patients in the subgroup with complicated UTI. ¶270 *E. coli* and 76 *K. pneumoniae* pathogens reported among 388 patients.

Figure 4: Proportion of participants/isolates with resistant pathogens



Multidrug resistance was defined as resistance to at least one antibiotic from at least three different classes unless otherwise indicated. Data are presented for all treatment groups combined. Analysis sets used to report data on pathogens varied between studies. ESBL=extended-spectrum beta-lactamase. MDR=multidrug resistant. NR=not reported. *Indicates that data are proportion of isolates rather than patients. †Data are among Enterobacteriaceae

isolates[18] and *Escherichia coli* isolates only.[27] ‡The proportion of participants with MDR pathogens was not reported but data are shown for pathogens that specifically were ESBL-producers, fluoroquinolone non-susceptible, and trimethoprim-sulfamethoxazole non-susceptible. §ESBL data represent the proportion of patients with ESBL-producing Enterobacterales,[19,22] *Escherichia coli*,[21] and Gram-negative uropathogens.[26] ¶Among patients with Enterobacterales[20] and Enterobacteriaceae[30] only.

Figure 5: Delphi questionnaire and final consensus results

Signalling questions	Category	Issue number	Issues that are likely to affect the ability to draw reliable conclusions from studies	Risk of bias		
				Low	Some concerns	High
A. How would you assess the risk of bias in study outcomes in terms of clinical and microbiological success (including sustainability), if these issues (risk factors) were: not considered at all, or not equally distributed, or not clearly defined (i.e., fever) in study groups?	Clinical situation	1	The type and frequency of clinical presentation and severity, i.e., percentages of cystitis, non-febrile and febrile pyelonephritis, and urosepsis			100
		2	Whether the infection was nosocomial or community-acquired		41	59
		3	Fever		23	77
		4	The general condition of the patient (ASA status)	14	36	50
		5	If two or more of the issues in this category were not satisfactorily described		9	91
	Patient	6	Catheter or stent (including nephrostomy tube) in place at diagnosis, during treatment and during follow-up			100
		7	Presence of urinary stones anywhere in the urinary tract		14	86
		8	History of symptomatic UTI in the previous 6 months		32	68
		9	A history of bacterial prostatitis		18	82
		10	Evidence of immune suppression or diabetes		14	86
		11	Female sex	9	59	32
		12	Antibiotic treatment within the previous 30 days		23	77
		13	Premenopausal women who have a history of recurrent UTI associated with sexual intercourse	14	32	55
		14	A history of obstipation	45	36	18
		15	If two or more of the issues in this category were not satisfactorily described		18	82

Signalling questions	Category	Issue number	Issues that are likely to affect the ability to draw reliable conclusions from studies	Risk of bias		
				Low	Some concerns	High
B. How would you assess the risk of bias of study outcomes in terms of microbiological success, if these issues related to the pathogen were: not reported at all, or not equally distributed in study groups?	Pathogen	16	The spectrum of pathogens		18	82
		17	Resistance to study drugs			100
		18	If both issues in this category were not satisfactorily described			100
C. How would you assess the risk of bias when comparing studies if these differences in study characteristics were present?	Study characteristics	19	Difference between studies in the duration of treatment with study drug of more than 100% of average treatment duration (typically >3 days)		50	50
		20	The period from EOT to TOC assessment differs by >1 week between studies		27	73
		21	The observation period from EOT to assessment of sustainability of effect differs by >3 weeks between studies		14	86
		22	A discrepancy between studies in microbiological success criteria of ≥10 CFU/mL (using $\leq 10^3$ instead of $\leq 10^4$ CFU/mL)	5	27	68
		23	A discrepancy between studies in the definition of microbiological "eradication" (using $\leq 10^3$ CFU/mL or no definition at all)		9	91
		24	If two or more of the issues in this category were present		5	95
D. How would you assess the overall risk of bias in the outcomes of a study (in terms of clinical and microbiological success):	All categories	25	If at least one issue in each of the following categories was present (clinical situation, patient, and pathogen)?		5	95
E. How would you assess the overall risk of bias when comparing outcomes between studies (in terms of clinical and microbiological success):		26	If at least one issue in all four categories (clinical situation, patient, pathogen, and study characteristics) was present?		9	91

Numbers in circles in the last three columns indicate the percentage of participants who voted for each risk level during the final round of the Delphi process (out of 22 respondents). Consensus was achieved if there was >75% agreement on risk of bias for an issue (indicated by shaded rows). ASA=American Society of Anaesthesiologists. CFU=colony forming unit. EOT=end of treatment. TOC=test of cure. UTI=urinary tract infection.