# Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Li H-K, Rombach I, Zambellas R, et al. Oral versus intravenous antibiotics for bone and joint infection. N Engl J Med 2019;380:425-36. DOI: 10.1056/NEJMoa1710926

# SUPPLEMENTARY INFORMATION

## Table of Contents

# List of Investigators

### Principal Investigators

- Ila Aggarwal - NHS Tayside
- Caroline Barker - Norfolk and Norwich University Hospitals
- Gavin Barlow - Hull and East Yorkshire Hospitals
- Neena Bodasing - University of North Midlands Hospitals
- Jon Campion - Northampton General Hospital
- Tumena Corrah & Aula Abbara - London North West Healthcare
- Simon Ellis - Northumbria Hospitals
- Jonathan Folb - The Royal Liverpool and Broadgreen University Hospitals
- Julia Greig - Sheffield Teaching Hospitals
- Carolyn Hemsley - Guy's and St Thomas' Hospitals
- Susan Hopkins & Damian Mack - Royal Free London
- Neil Jenkins - Heart of England Hospitals
- Michael Kelsey - Whittington Health
- Parvez Moondi & Sunil Sharma - The Queen Elizabeth Hospital King's Lynn
- Elinor Moore - Cambridge University Hospitals
- Jonathan Sandoe - The Leeds Teaching Hospitals
- Catherine Sargent - Brighton and Sussex University Hospitals
- Matthew Scarborough - Oxford University Hospitals
- Uli Schwab - The Newcastle upon Tyne Hospitals
- Andrew Seaton - NHS Greater Glasgow and Clyde
- Rashmi Sharma - Blackpool Teaching Hospitals

- Bijayendra Singh - Medway NHS Foundation Trust
- Rebecca Sutherland & Claire Mackintosh - NHS Lothian
- Sentil Velayudham & Nick Nicolaou - Maidstone and Tunbridge Wells Hospital
- Simon Warren - Royal National Orthopaedic Hospital
- Martin Williams - University Hospitals Bristol

**Research Teams**

- K Bisnauthsing, S Brown, A Bullough, D Bunn, J Calderwood, R Carpio, K Cox, S Crossman, K Cullen, K Darnley, A Dunne, F Fitzgerald, L Flintoff, A Harin, MC Hoyle, J Marshall, G Menzies, S Pahary, G Pakou, A Querol-Rubiera, P Raymode, A Remegoso, H Reynolds, C Ridley, B Sandhu, H Sankey, E Saunders, L Shewan, K Smith, L Spoors, S Stafford, C Vallance, N Waddington, L Walford.

**Additional OVIVA trial collaborators**

- M McNally, B Atkins, D Stubbs, P McLardy-Smith, H Pandit, M Kumin, P Matthews, A Brent, J Lomas, A Woodhouse, R Gundle, M Rogers, A Taylor, B Angus, P Sendi, I Byren, A Berendt, CE Moran, A Briggs, E Khatamzas, T Wangrangsimakul, TH Wong, L Barrett, A Alvand, GE Thwaites

**Trial Committee Members**

- Data monitoring committee - N French (Chair), C Smith, M Llewelyn
- Trial steering committee - G Cooke (Chair), J Paul (Co-Chair), J Bostock, F Old
- Endpoint review committee – BA Lipsky (Chair), H Hughes, D Bose

**Trial Management Group**

- M Scarborough, HK Li, I Rombach, R Zambellas, AS Walker, P Bejon, C Cooper, C Scarborough, L Jefferies, K Holland, A Greshon. C Geue, N McMeekin

# Supplementary Methods

## Participants and Setting

### *Inclusion criteria*

The participant had to meet EACH of the following criteria:

1) A clinical syndrome comprising any of the following:
    a) localised pain, OR
    b) localised erythema, OR
    c) temperature >38.0ºC, OR
    d) a discharging sinus or wound
2) Willing and able to give informed consent
3) Aged 18 years or above
4) Had received 7 days or less of intravenous therapy after an appropriate surgical intervention to treat bone or joint infection (regardless of pre-surgical antibiotics) or, if no surgical intervention was required, the patient had received 7 days or less of intravenous therapy after the start of planned curative treatment for the relevant clinical episode
5) Life expectancy > 1 year
6) Acute or chronic bone or joint infection in one of the following categories:
    a) [1]native osteomyelitis (i.e., bone infection without metalwork) including haematogenous or contiguous osteomyelitis
    b) [2]native joint sepsis treated by excision arthroplasty;
    c) prosthetic joint infection treated by debridement and retention, by one stage revision or by excision of the prosthetic joint (with or without planned re- implantation);
    d) [3]orthopedic fixation device treated by debridement and retention, or by debridement and removal;
    e) [4]spinal infection

[1] *This group consisted of osteomyelitis of the extra-axial skeleton only; vertebral osteomyelitis was categorized separately under (e).*
[2] *Patients with native joint sepsis without associated osteomyelitis were not eligible as they would not ordinarily require at least six weeks of IV antibiotic therapy or excision arthroplasty.*
[3] *All participants in this group had a fixation device associated infection; prosthetic joints were categorized separately under (c)*
[4] *All participants in this group had vertebral osteomyelitis with or without associated discitis or soft tissue infection.*

***Exclusion criteria***

The participant was ineligible if he or she met ANY one of the following criteria:

1) *Staphylococcus aureus* bacteraemia on presentation or within the previous month
2) Bacterial endocarditis, either on presentation or within the previous month (NB: there were no study mandated investigations, so participants were not required to have echocardiograms, blood cultures, or any other investigations to exclude endocarditis in the absence of a clinical indication)
3) Any other concomitant infection which, in the opinion of the clinician responsible for the patient, required a prolonged course of intravenous antibiotic therapy (e.g., mediastinal infection or central nervous system infection)
4) Mild osteomyelitis, defined as bone infection which, in the opinion of the infection specialist responsible, required less than six weeks of antibiotic therapy
5) An infection for which there were no suitable antibiotic choices to permit randomization between the two arms of the trial (for example, where organisms were only sensitive to intravenous antibiotics)
6) Prior enrolment in the trial
7) Septic shock or systemic features requiring intravenous antibiotic therapy in the opinion of the treating clinician (the patient could be re-evaluated if these features resolved within seven days of the start of treatment episode)
8) Unlikely to comply with trial requirements following randomization in the opinion of the investigator
9) Clinical, histological or microbiological evidence of mycobacterial, fungal, parasitic or viral aetiology of the infection
10) Receiving an investigational medical product as part of another clinical trial

The use of antibiotic-loaded cement in spacers, bone substitutes or beads at the site of infection was not an exclusion criterion, but was recorded. Pregnancy, renal failure and liver failure were not exclusion criteria, provided that suitable antibiotic options could be identified for both IV and PO therapy prior to randomization.

***Diagnostic certainty at trial entry***

Baseline data for each participant were assessed for certainty of diagnosis of bone or joint infection based on the following criteria that were predefined in the protocol.

Definite evidence of infection at baseline was defined by one or more of the following:

a) isolating bacteria from 2 or more samples of bone/peri-prosthetic tissue, where the bacteria isolated from these samples were indistinguishable according to routine laboratory tests, including the antibiogram

b) a pathogenic organism (e.g. *Staphylococcus aureus* but not *Staphylococcus epidermidis*) on a single, closed, biopsy of native bone or peri-prosthetic tissue

c) diagnostic histology on bone/peri-prosthetic tissue

d) a draining sinus tract arising from bone/prosthesis or

e) frank pus adjacent to bone/ prosthesis

If any of these criteria were met, then the category “definitive” infection was applied without blinded, independent review.

Where these criteria were not met, three independent clinicians, uninvolved in recruitment or management of participants, were sent a redacted copy of the patient’s admission notes and laboratory results from the time of randomisation, and asked to apply the following criteria to determine “probable” or “possible” infection:

Infection was categorized as “probable” where microbiological sampling had not been undertaken, AND none of the other criteria for definite infection had been fulfilled AND any one of the following were met:

a) Radiological or operative findings of periosteal changes suggesting chronic osteomyelitis OR
b) Radiological findings suggesting vertebral infection OR
c) The development of a discharging wound after an orthopaedic procedure where prosthetic material had been implanted OR
d) The presence of deep pus close to but not adjacent to bone/prosthetic joint/orthopaedic device OR
e) The presence of peri-prosthetic necrotic bone OR
f) Rapid loosening of a joint prosthesis/orthopaedic device (i.e. leading to localized pain in less than 3 months since implantation) in the absence of a mechanical explanation for rapid loosening.

Infection was categorized as “possible” where microbiological sampling had been undertaken with negative results (according to criteria described above for “definite” infection) AND other criteria for definite infection were not fulfilled AND in addition one or more of the criteria listed a) to f) above was met.

Where the review committee members differed in their initial categorization of a participant, consensus was achieved by discussion or, if necessary, by majority opinion.

**Interventions**

Participants were required to start their randomized treatment strategy within seven days of the start of the treatment episode which, in most cases was defined by the date of surgical intervention; this was to allow sufficient time for availability of microbiological culture results and for recovery from surgery and anaesthesia. Prior to randomization, the choice of agents were a patient to be randomized to each arm was determined by a specialist in clinical infection based on clinically relevant variables including antimicrobial susceptibility, drug interactions, comorbidities, drug allergies, prior infections and local epidemiology; patients were ineligible if no suitable oral option could be identified. It is therefore likely that the most appropriate antibiotic regime, individualized for each participant, would have been selected according to criteria similar to those used in routine daily practice outside the context of a clinical trial. As antibiotic susceptibility informed part of the eligibility criteria (specifically, an exclusion criteria was ‘no treatment options available for either PO or IV therapy at the point of randomization’), we did not routinely collect data on adequacy of antimicrobial therapy on the case record forms.

Participants randomized to PO therapy were permitted up to five days of IV antibiotic therapy for intercurrent infections unrelated to the incident orthopaedic infection without meeting an endpoint or deviating from the protocol. Participants randomized to IV therapy were permitted adjunctive oral therapy, such as oral rifampicin, in order to reflect standard practice outside the context of a clinical trial. The infection specialist was permitted to alter the choice of antibiotic agent according to clinical need (e.g. due to side effects or emerging laboratory results). If suitable alternatives were available within the allocated strategy, the patient remained within protocol. If no alternative antibiotic agent was available within the randomized strategy, an early

exit from strategy (secondary endpoint) was recorded but the participant continued follow-up according to the trial protocol and was included in the 'intention to treat' analysis.

**Assessments**

Trial specific data were obtained either from direct face-to-face participant contact at predefined time points, or from their clinical case records provided that their routine clinical review occurred within a specified range: for day 42 (range 21 to 63), day 120 (range 70 to 180) and day 365 (range 250 to 420). If insufficient data were available from the clinical care record, the investigator arranged a telephone review with the participant or their general practitioner.

***Study oversight***

The study was conceived by Philip Bejon and designed by physicians and surgeons at the Oxford Bone Infection Unit. Data were collected by the investigators and associated site personnel, analysed by Ines Rombach and Sarah Walker and interpreted by the authors. The first, second and last author wrote the first draft of the manuscript. All authors participated in review and editing of the manuscript, approved the submitted version and vouch for its accuracy.

***Endpoint definitions***

Potential primary endpoints were identified through post-randomization prospective surveillance, and reviewed by an endpoint committee blind to the treatment group. Potential endpoints were identified and notified to the end-point committee according to the following protocol definition: 'Any post-randomization re-admission or return to theatre with signs or symptoms at the anatomical site of infection will be considered a potential endpoint. In addition, any signs or symptoms identified on review of the patient or their hospital notes at follow up visits that, in the opinion of the study clinician, may represent treatment failure will be considered a potential endpoint.'

The primary endpoint was: failure of infection treatment, where definite failure was indicated by one or more of the following:

a) isolating bacteria from 2 or more samples of bone/peri-prosthetic tissue, where the bacteria isolated from these samples were indistinguishable according to routine laboratory tests, including the antibiogram.

b) a pathogenic organism (e.g. *Staphylococcus aureus* but not *Staphylococcus epidermidis*) on a single, closed, biopsy of native bone or peri-prosthetic tissue

c) diagnostic histology on bone/peri-prosthetic tissue

d) a draining sinus tract arising from bone/prosthesis or

e) frank pus adjacent to bone/ prosthesis.

Secondary endpoints were:

1) SAEs, including death (i.e. all cause) according to treatment allocation;
2) Line complications (i.e. infection, thrombosis or other events requiring early removal or replacement of the line);
3) "Probable" or "possible" treatment failure as composites with definitive treatment failure. These were determined by a blinded endpoint committee according to the following criteria:

    a) Loosening of a prosthesis, confirmed radiologically; or

    b) Non-union of a fracture after 6 months, confirmed radiologically; or

    c) Superficial spreading erythema, treated as cellulitis with an antibiotic for >1 week, where results from deep tissue samples did not meet the primary endpoint as described above.

    Where appropriate deep tissue samples were sent for microbiology and the results of culture were negative, and either a), b) or c) were met, the endpoint was regarded as "possible". On the other hand, where deep tissue samples were not sent for microbiology, and either a), b) or c) were met, then the endpoint was regarded as "probable";
4) Early termination of the planned 6-week period of oral or IV antibiotics because of adverse events, patient preference or any other reason;
5) Resource allocation determined by: a) length of inpatient hospital stay; b) frequency of outpatient visits and c) antibiotic prescribing costs;
6) Quality of life evaluated by EQ-5D-3L questionnaire;
7) Oxford Hip and Knee Scores (where infection was in the hip or knee); and
8) Adherence to oral medication.

The study clinicians determined secondary endpoints 1, 2, 4 and 5. The blinded endpoint review committee determined primary endpoints and secondary endpoint 3, by reviewing relevant clinical notes redacted for personal details and any information

which might have betrayed the treatment allocation. In cases where potential treatment failures did not fit clearly into either definite, probable, possible or no treatment failure, the blinded endpoint committee achieved consensus by discussion or, if necessary, by majority opinion.

Participant questionnaires determined secondary endpoints 6 and 7. Secondary endpoint 8 was determined by questionnaire in all centres, and by MEMS at four sentinel sites.

**Choice of non-inferiority margin**

The choice of NI margin was discussed in detail at the time of trial design in terms of clinical relevance. There were few previous studies in this population which could reliably inform the choice but 10% non-inferiority margins have been recommended by licencing authorities for antibiotic trials[1,2] and are commonly used in other infections with failure rates of the order of 10% (or equivalently success rates of 90%).[3,4] Based on this, and the fact that we anticipated failure rates with IV antibiotics to be small in absolute terms (5%), such that an absolute difference was more relevant than a relative difference, we and the co-investigators agreed, following discussion with a wide range of clinicians, infection specialists and orthopaedic surgeons that a 5% non-inferiority margin for the risk difference on the absolute scale was reasonable clinically. Although an absolute increase of 5% translated to a relative increase of 100%, this was deemed clinically acceptable considering the low absolute risk i.e. corresponding to a worst case scenario of the upper limit of the 90% confidence interval suggesting that failure rates might be 10% in the PO arm (5% control group + 5% NI margin), and the other benefits to patients and the healthcare system that the PO arm was expected to deliver. Such potential benefits included circumventing the need for intravenous access devices, reduced frequency of hospital follow-up appointments for therapeutic drug monitoring and line reviews, reduced reliance on community IV nurses and fewer restrictions in activities of daily living and attendance at work. In addition, we anticipated that participants in the PO arm would be able to leave hospitals earlier, thereby reducing the pressure on hospital beds and the demand on health resources through introduction of a cost-saving intervention.

The original control group failure rate (5%) was based on pilot data from a limited number of participants from one specialised centre. At a planned interim analysis in

February 2015 after 601 patients had been randomized across multiple sites, the observed event rate in the IV arm was approximately 12.5%, higher than the "worst case" scenario envisaged for the original 5% NI margin. It can be inappropriate to maintain the relative or absolute risk difference after an observed change in the underlying event rate in the control group.[5] With this larger control group event rate, retaining the same 5% absolute non-inferiority margin would have corresponded to a much smaller relative difference (40% rather than 100%) which was considered too restrictive in light of the higher control group failure rate (meaning that more participants were already experiencing poor outcomes) and hence a slightly greater increase in the absolute percentage experiencing them on PO could be tolerated, given the potential benefits of PO therapy outlined above. The NI margin was therefore pragmatically amended to 7.5% on the absolute risk difference scale (corresponding to a 60% relative difference), being still more stringent than commonly used 10% margins in anti-infective trials, but only slightly larger than the original 5% absolute margin. This was based on its correspondence to a worst-case scenario of the upper limit of the 90% confidence interval suggesting that failure rates might be 20% in the PO arm (12.5% control group + 7.5% NI margin) compared to the original 10% (5% control group + 5% NI margin). By way of example, the Federal Drug Authority (FDA) provides a worked example in its most recent November 2016 guidance for antibiotic trials in community acquired bacterial pneumonia (CABP).[2] Historical observational studies estimated mortality rates of between 30-80% without antibiotics and 5-17% with antibiotics. Based on this, the FDA concluded that "Use of an NI margin of approximately 10% is therefore a valid approach for evaluating new treatments of CABP and would clearly represent an effect superior to no treatment as well as, based on clinical judgment, an appropriate clinical margin". For the serious infections eligible for this trial, historical data suggest that, without antibiotics, mortality would be between 5-25% and treatment failure rates likely >50%, compared to <20% with antibiotics.[6] The FDA guidance would therefore have suggested that a margin of 10% could have been considered for this trial, had it been intended for licensing. Retaining the original 5% non-inferiority margin with the higher control group event rate (12.5% vs original 5%) and the same alpha/power would also have increased the sample size from 1050 to 1668; this was unfeasible given the funding constraints and effect on trial duration. The increase in non-inferiority margin to 7.5% was approved as a protocol amendment by the Trial Steering, Data Monitoring and the Ethics Committees.

**Statistical Analyses**

The primary endpoint for ITT, mITT, per-protocol and sensitivity analyses were reported with 90% confidence intervals (following the sample size calculation) and also with 95% confidence intervals. All secondary and sub-group analyses were reported with 95% confidence intervals.

Subgroup analyses were based on the mITT population except for those concerning definite / probable / possible infection at baseline, which were performed in both the ITT and mITT populations. Odds ratios were obtained from logistic regression models using definitive treatment failure as the dependent variable, and randomized allocation, relevant subgroup and the interaction term as the only covariates.

Imputation for the primary endpoint used chained estimating equations with data augmentation and included the following variables identified as relevant in predicting outcomes by the Chief Investigator before the final analysis:

- Infection details at baseline were combined as follows and used as binary variables in the imputation model:
    - Chronic osteomyelitis debrided, no current implant or device OR spinal infection debrided
    - Chronic osteomyelitis as above, but not debrided OR spinal infection not debrided
    - Implant or device present and retained ("DAIR")
    - Removal of orthopaedic device for infection OR prosthetic joint implant removed
    - Prosthetic joint implant, 1-stage revision
- Whether or not antibiotic beads/ cement were used in the index operation
- Participants' comorbidity status (yes vs. no):
    - Diabetes
    - Peripheral vascular disease
    - Current smoker
    - Rheumatoid arthritis or systemic autoimmune disease
- *Staphylococcus aureus* present in samples taken before randomisation
- *Pseudomonas species* present in samples taken before randomisation
- Age
- Gender

20 imputations were combined using Rubin's rules.

*Internal Pilot study*

Randomization started in one centre (Oxford) as a pilot centre to demonstrate feasibility of recruitment. Data from the pilot study were never analysed on their own, but only as part of the main trial analyses.

**Health economic analysis**

Full formal cost-effectiveness analysis was beyond the scope of this manuscript and will therefore be reported separately. A summary of methods and results is presented here.

A within trial analysis was based on the resource use and Health Related Quality of Life (EQ-5D-3L) data collected during the trial. We considered the following cost categories: antibiotic costs, the costs of intravenous administration (including equipment and staff time) and inpatient stays through one year of follow-up. The British National Formulary was used for antibiotic costs (with adjustment for hospital pharmacy discounts) and standard NHS reference costs were used for hospitalizations, equipment and staff resource.

The total cost per participant in each intervention was summed and divided by the number of participants to calculate the mean cost per participant in each arm, along with the difference in means and 95% confidence interval. All resources were valued in GBP (pounds sterling) at the time of analysis, using the Health Service Cost Index where necessary to adjust for price changes from year to year.

The EQ-5D-3L instrument was used to estimate per-patient quality-adjusted life years (QALY) with adjustment for any differences between the groups in EQ-5D-3L at baseline. Non-parametric bootstrapping techniques were employed to confirm the robustness of the statistical analysis of cost, QALY and cost-per-QALY.

The mean QALY per participant for each arm was calculated by summing all participants' QALYs and dividing by the number of participants. The difference in the means along with 95% confidence intervals was calculated. The incremental cost-effectiveness ratio was calculated by dividing the difference in mean costs by the

difference in mean QALYs. Uncertainty in cost-effectiveness was represented on the cost-effectiveness plane and as confidence intervals for cost-effectiveness ratios.

Within trial non-surgical treatment costs relating to antibiotics, administration of antimicrobials (including equipment and staff time) and hospitalized bed-days (including readmissions) through one year of follow-up were an estimated £2,740 (~$3,480) less per patient in the PO arm (95%CI:£1,488, £3,992). There was a non-significant difference in quality-adjusted life-years of -0.008 (95%CI:-0.045,+0.031) favoring PO (Table S12;Figure S5).

As the results of the trial showed non-inferiority of PO therapy, we did not include post-randomization surgical costs and it was not considered necessary or useful to extrapolate the observed results beyond the clinical trial period in order to explore potential lifetime differences in cost-effectiveness.

**Bayesian analysis**

We performed supplementary analysis within a Bayesian framework to estimate the probability that PO was inferior to IV by different degrees.[7] We used a Bayesian binomial regression model to estimate the risk difference (and 90% credible interval) for definitive treatment failure between the treatment arms (i.e. PO vs. IV) in the modified intention-to-treat population (without imputation). Following the pre-specified primary analysis of the OVIVA trial, this analysis was not adjusted for any patient characteristics.

We ran the Bayesian model with both a neutral (uninformative / flat) prior, as well as with a sceptical (informative) prior to explore its effects on the model estimates. The uninformative prior was estimated by a normal distribution with a mean of 0 and a standard deviation (SD) of 100. The informative prior hypothesised a 10% risk difference in favour of the IV strategy: Normal (0.1, SD=0.025). This analysis was performed in Stata version 15 using the 'bayes: binreg' command.

For the two Bayesian binomial regression models using the neutral and sceptical priors, the probability that PO was inferior to IV was calculated for a range of thresholds which could be considered as varying non-inferiority margins. As reported in the main manuscript results, the probabilities of PO being 1%, 2% and 5% non-

inferior to IV using the neutral prior were 12.7%, 5.5% and 0.1% respectively. Figure S6 below plots the probabilities for PO being inferior to IV for a range of non-inferiority margins for Bayesian models using the neutral as well as the sceptical prior.

Regardless of the prior used, the probability that PO is worse than IV by 5% or more approaches 0. This provides reassurance that, regardless of the change to the non-inferiority margin, the trial robustly demonstrates non-inferiority of PO as compared to IV therapy, as assessed by definitive treatment failure at one year.

## Supplementary Figures

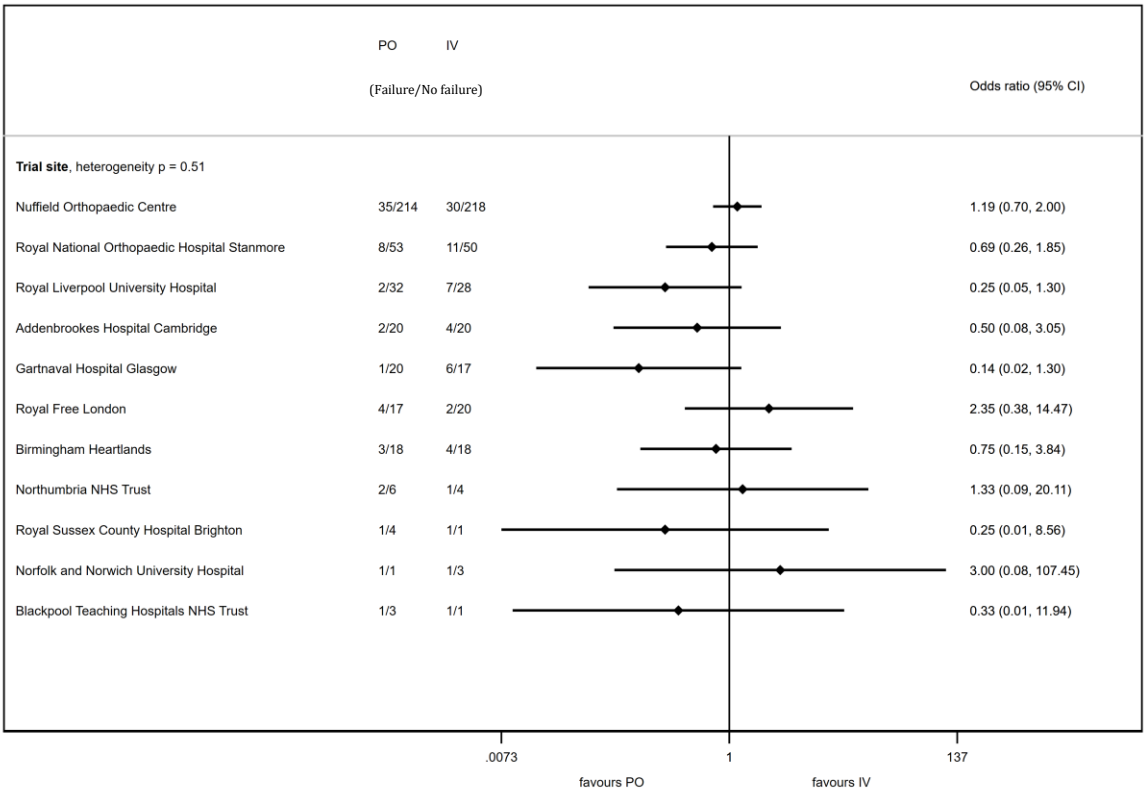
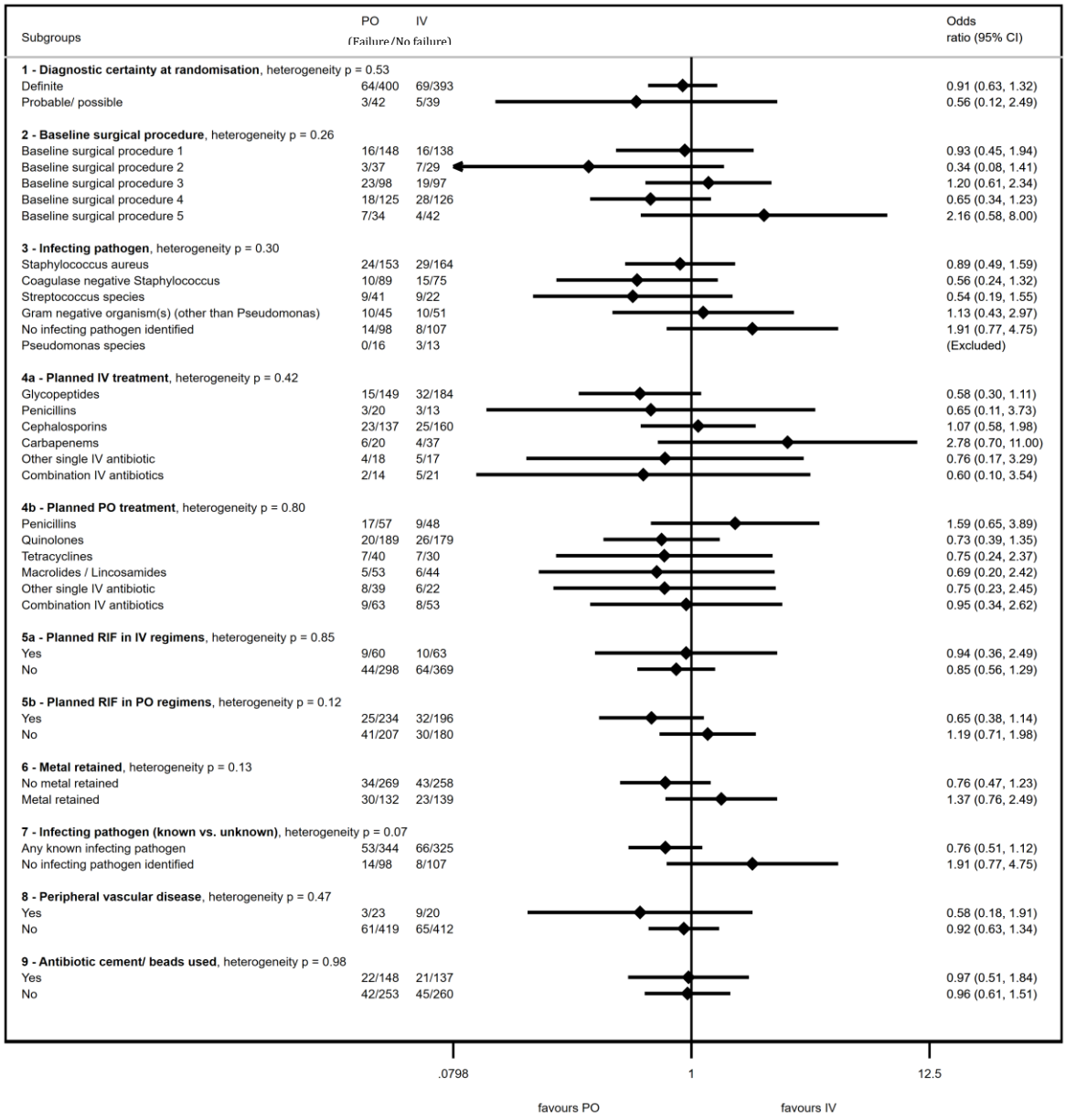



**Figure S1 : Forest plot of odds ratios (95% CI) for definitive treatment failure by trial site (PO/IV)**

A total of 26 sites contributed to recruitment with median (IQR) of 8(4-44) and range 1 to 512 participants per center. At sites not represented in this forest plot, there were no treatment failures in one or both arms of the trial so formal outcome comparison was not possible.

Figures for The Nuffield Orthopaedic Centre include 228 patients recruited during a single center internal pilot study.




## Figure S2: Forest plot of odds ratios (95% CI) for definitive treatment failure by subgroups (PO/IV)

1. Pre-specified subgroup analysis by diagnostic certainty. *Definite, probable and possible infection at baseline were determined by predefined criteria or, where necessary, by consensus through a blinded independent review committee.*
2. Pre-specified subgroup analysis by baseline surgical procedure. *Subgroups were as follows: (1) Osteomyelitis of axial or extra-axial skeleton debrided, no current implant or device; (2) Osteomyelitis as above, but not debrided; (3) Implant or device present and retained (i.e. "DAIR"); (4) Removal of prosthetic joint or fracture fixation device for infection; (5) Prosthetic joint implant, single-stage revision.*
3. Pre-specified subgroup analysis by infecting pathogen. *Other than Staphylococcus aureus, we did not systematically collect data to species level. 5 cases of S. lugdunensis were reported; all were included in the subgroup labelled Coagulase negative Staphylococci. Of 186 cases of Staphylococcus aureus for which antimicrobial susceptibilities were available, MRSA was identified in 19.(10.2%). Amongst 74 cases of Gram negative infection (other than Pseudomonas) where susceptibilities were available, 2 were extended spectrum beta-lactamase (ESBL) producers, and 35 represented species capable of expressing Amp C. There were no treatment failures in the PO arm of the Pseudomonas species subgroup and therefore comparison was not possible.*
4. *4a and 4b.* Pre-specified subgroup analyses by planned IV and PO treatment (excluding adjunctive oral rifampicin) as stated immediately prior to randomization.
5. *5a and 5b.* Pre-specified subgroup analyses by planned use of adjunctive oral rifampicin in the IV and PO arm as stated immediately prior to randomization.
6. Post-hoc subgroup analysis by metal retained vs no metal retained. *Excludes vertebral osteomyelitis and participants managed without surgical debridement. Prosthetic material other than metal was not considered in this analysis.*
7. Post-hoc subgroup analysis by infecting pathogen known or unknown.
8. Post-hoc subgroup analysis by peripheral vascular disease.
9. Post-hoc subgroup analysis by use of adjunctive local antibiotic therapy.

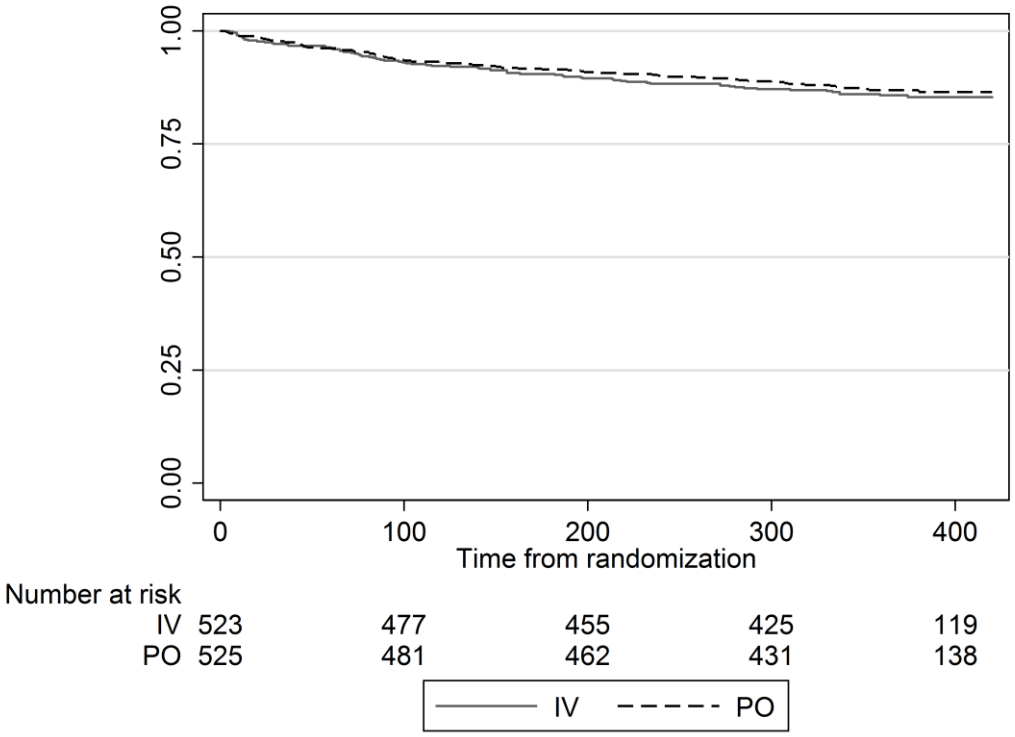



**Figure S3: Kaplan-Meier curves for time to treatment failure by randomized strategy**

*Hazard ratio p=0.57*

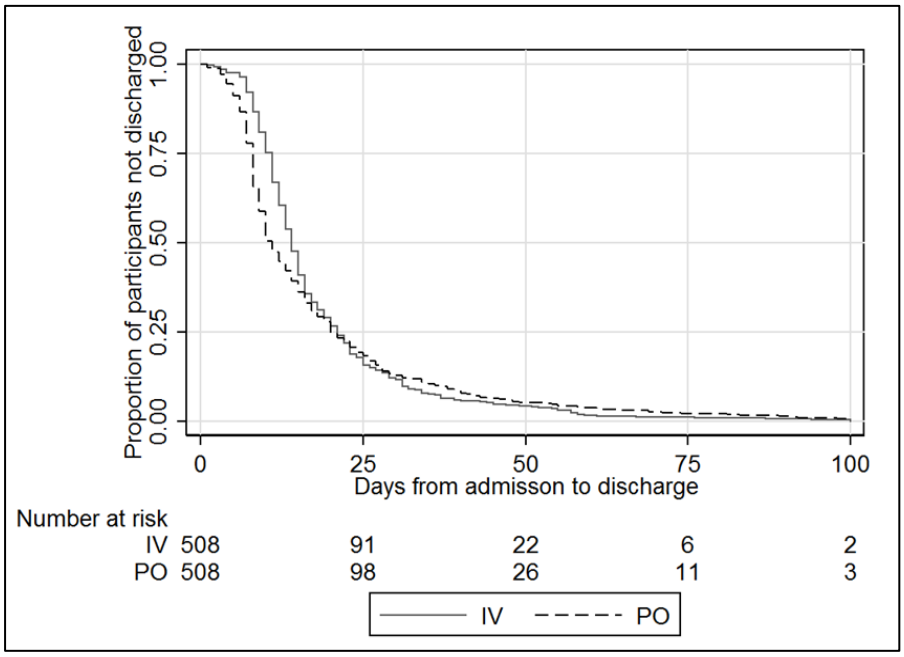



**Figure S4: Time from admission to discharge**

*p<0.001 (Ranksum test for difference in median time to discharge)*

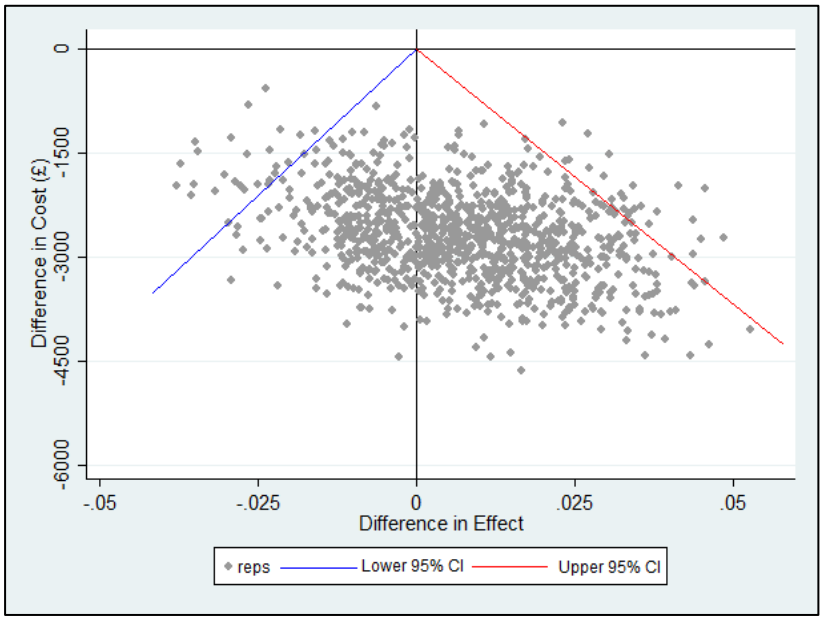



**Figure S5: Cost effectiveness plane showing 1000 bootstrap samples**

All fall in the Southern quadrants of the plane indicating that, in each case, PO antibiotic therapy cost less than IV therapy using data from the OVIVA trial. Samples in the South East quadrant represent higher QALYs than those in the South West quadrant; the sample distribution suggests that there remains uncertainty as to whether QALYs clearly favor the oral or intravenous arm.

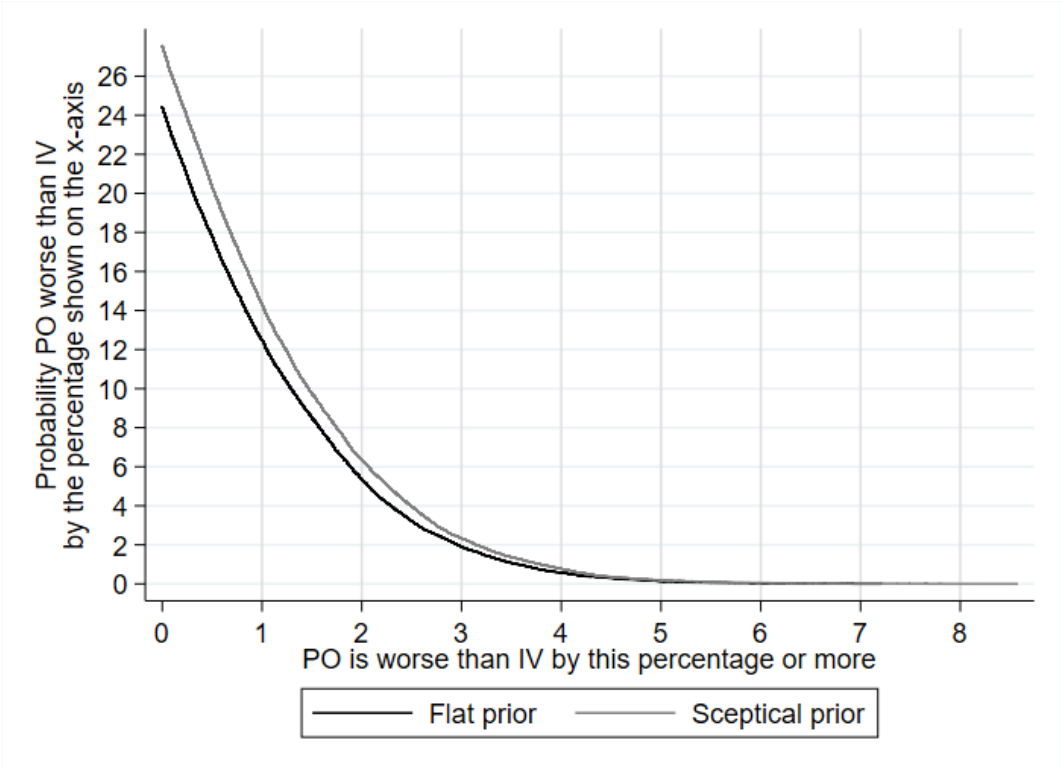



**Figure S6: Exploratory Bayesian analysis within the mITT population**

# Supplementary Tables

**Table S1a: Supplementary baseline information**

| | IV Antibiotic (N = 527) | PO Antibiotic (N = 527) | Total (N = 1054) |
|---|---|---|---|
| Clinical presentation* | | | |
| - localised pain | 397 (75.3%) | 403 (76.5%) | 800 (75.9%) |
| - localised erythema | 226 (42.9%) | 207 (39.3%) | 433 (41.1%) |
| - temperature > 38.0 C | 62 (11.8%) | 62 (11.8%) | 124 (11.8%) |
| - discharging sinus / wound | 296 (56.2%) | 285 (54.1%) | 581 (55.1%) |
| Anatomical site of infection* | | | |
| - spinal infection[a] | 37 (7.0%) | 35 (6.6%) | 72 (6.8%) |
| - upper limb infection | 43 (8.2%) | 59 (11.2%) | 102 (9.67%) |
| - lower limb infection[b] | 436 (82.7%) | 419 (79.5%) | 855 (81.1%) |
| - other area of infection | 12 (2.3%) | 14 (2.7%) | 26 (2.5%) |
| Operative details* | | | |
| - draining sinus arising from bone / prosthesis | 177 (33.6%) | 142 (26.9%) | 319 (30.3%) |
| - frank pus adjacent to bone / prosthesis | 179 (34.0%) | 186 (35.3%) | 365 (34.6%) |
| - intraoperative local antibiotics (cement or beads) | 165 (31.3%) | 179 (33.8%) | 343 (32.5%) |
| Baseline diagnosis determination* | | | |
| - clinical findings[c] (with or without other findings) | 285/527 (54.1%) | 273/527 (51.8%) | 558/1054 (52.9%) |
| - microbiological findings[d] (with or without other findings) where microbiology samples were submitted | 402/500 (80.4%) | 400/503 (79.5%) | 802/1003 (80.0%) |
| - histological findings[e] (with or without other findings) where histology samples were submitted | 266/310 (85.8%) | 277/326 (85.0%) | 543/636 (85.4%) |
| Baseline diagnosis determination by mutually exclusive category*: | | | |
| - clinical and microbiological and histological findings | 144 (27.3%) | 136 (25.8%) | 280 (26.6%) |
| - clinical and microbiological findings | 87 (16.5%) | 91 (17.3%) | 178 (16.9%) |
| - clinical and histological findings | 22 (4.2%) | 20 (3.8%) | 42 (4.0%) |
| - microbiological and histological findings | 79 (15.0%) | 91 (17.3%) | 170 (16.1%) |
| - clinical findings only | 32 (6.1%) | 26 (4.9%) | 58 (5.5%) |
| - microbiological findings only | 92 (17.5%) | 82 (15.6%) | 174 (16.5%) |
| - histological findings only | 21 (4.0%) | 30 (5.7%) | 51 (4.8%) |

| | | | |
|---|---|---|---|
| - other[f] | 50 (9.5%) | 51 (9.7%) | 101 (9.6%) |
| Comorbidities* | | | |
| - diabetes | 107 (20.3%) | 98 (18.6%) | 205 (19.5%) |
| - renal failure | 11 (2.1%) | 11 (2.1%) | 22 (2.1%) |
| - ischaemic heart disease | 43 (8.2%) | 45 (8.5%) | 88 (8.4%) |
| - peripheral vascular disease | 31 (5.9%) | 32 (6.1%) | 63 (6.0%) |
| - previous stroke or TIA | 19 (3.6%) | 22 (4.2%) | 41 (3.9%) |
| - dementia | 1 (0.2%) | 1 (0.2%) | 2 (0.2%) |
| - immunosuppressive medication | 28 (5.3%) | 17 (3.2%) | 45 (4.3%) |
| - known HIV infection | 1 (0.2%) | 3 (0.6%) | 4 (0.4%) |
| - rheumatoid arthritis or autoimmune disease | 47 (8.9%) | 38 (7.2%) | 85 (8.1%) |
| - current smoker | 61 (11.6%) | 79 (15.0%) | 140 (13.3%) |
| - malignancy (current or diagnosed <2 years) | 17 (3.2%) | 17 (3.2%) | 34 (3.2%) |

** Frequency and percentages are displayed*

[a] *This figure includes spinal and pelvic osteomyelitis*
[b] *Details of lower limb sites shown in table S1b*
[c] *Defined by a draining sinus tract arising from bone/prosthesis and/or frank pus adjacent to bone/ prosthesis at operation*
[d] *Defined by indistinguishable bacterial isolates from ≥ 2 deep tissue samples or a pathogenic organism from a single closed aspirate or biopsy*
[e] *Defined by characteristic inflammatory infiltrate or microorganisms on microscopy*
[f] *For 101 participants who could not be classified by the responsible infection specialist as definite infection according predefined protocol definitions, baseline diagnostic categorization was determined by an independent blinded assessment committee according to a range of clinical criteria, including radiological findings: 8 of these participants were classified as definite infection, 23 as probable infection, and 51 as possible infection. A further 19 participants had insufficient clinical information available to permit categorization according to predefined criteria but all were treated with antibiotics and were therefore included as possible infection at baseline*

**Table S1b: Information on location of infection where lower limbs affected**

| | **IV Antibiotic (N = 436)** | **PO Antibiotic (N = 419)** | **Total (N = 855)** |
|---|---|---|---|
| Hip* | 110 (25.2%) | 104 (24.8%) | 214 (25.0%) |
| Knee* | 133 (30.5%) | 115 (27.5%) | 248 (29.0%) |
| Foot* | 89 (20.4%) | 86 (20.5%) | 175 (20.5%) |
| Other area of lower limb infection* | 105 (24.1%) | 113 (27.0%) | 218 (25.5%) |

**Frequency and percentages are displayed*

*(Data on the location of infection in the lower limb were not available for four participants)*

**Table S2: Criteria defining primary endpoints (definitive treatment failure)**

| | IV Antibiotic (N = 74 ) | PO Antibiotic (N = 67 ) | Total (N = 141 ) |
|---|---|---|---|
| Definite treatment failure determination* | | | |
| - clinical findings[a] (with or without other findings) | 49 (66.2%) | 34 (50.7%) | 83 (58.9%) |
| - microbiological findings[b] (with or without other findings) where microbiology samples were submitted | 54 (73.0%) | 47 (70.1%) | 101 (71.6%) |
| - histological findings[c] (with or without other findings) where histology samples were submitted | 18 (24.3%) | 11 (16.4%) | 29 (20.6%) |
| Definite treatment failure determination by mutually exclusive category*: | | | |
| - clinical and microbiological and histological findings | 6 (8.1%) | 3 (4.5%) | 9 (6.4%) |
| - clinical and microbiological findings | 23 (31.1%) | 13 (19.4%) | 36 (25.5%) |
| - clinical and histological findings | 4 (4.5%) | 1 (1.5%) | 5 (3.6%) |
| - microbiological and histological findings | 8 (10.8%) | 5 (7.5%) | 13 (9.2%) |
| - clinical findings alone | 16 (21.6%) | 17 (25.4%) | 33 (23.4%) |
| - microbiological findings alone | 17 (23.0%) | 26 (38.8%) | 43 (30.5%) |
| - histological findings alone | 0 (0.0%) | 2 (3.0%) | 2 (1.4%) |

** Frequency and percentages are displayed*

[a] *Defined by a draining sinus tract arising from bone/prosthesis and/or frank pus adjacent to bone/ prosthesis at operation*
[b] *Defined by indistinguishable bacterial isolates from ≥ 2 deep tissue samples or a pathogenic organism from a single closed aspirate or biopsy*
[c] *Defined by characteristic inflammatory infiltrate or microorganisms on microscopy*

**Table S3: Results of imputations for patients with missing data in the ITT analysis**

| | **IV Antibiotic (N = 527)** | **PO Antibiotic (N = 527)** | **Total (N = 1054)** |
|---|---|---|---|
| Participants without observed data for the primary endpoint | 21 (4.0%) | 18 (3.4%) | 39 (3.7%) |
| Definitive treatment failures (out of all participants with observed primary endpoint data) | 74/506 (14.6%) | 67/509 (13.2%) | 141/1015 (13.9%) |
| Range of imputed numbers of definitive treatment failures out of those with missing data | 1 to 6 (out of 21) | 0 to 7 (out of 18) | |
| Range of % definitive treatment failures in the imputed datasets | 14.2% to 15.2% | 12.7% to 14.0% | |

**Table S4: Categorisation of endpoints as: 'definite' or 'probable' or 'possible' treatment failure** (complete cases population – i.e. the mITT population)

| | **IV Antibiotic * (N = 506)** | **PO Antibiotic * (N = 509)** | **Total * (N = 1015)** |
|---|---|---|---|
| Any treatment failure* | 80 (15.8%) | 77 (15.1%) | 157 (15.5%) |
| | **IV Antibiotic (N = 80)** | **PO Antibiotic (N = 77)** | **Total (N = 157)** |
| Endpoint category* | | | |
| Definite treatment failure | 74 (92.5%) | 67 (87.0%) | 141 (89.8%) |
| Probable treatment failure | 5 (6.3%) | 8 (10.4%) | 13 (8.3%) |
| Possible treatment failure | 1 (1.3%) | 2 (2.6%) | 3 (1.9%) |

***** *Frequency and percentages are displayed*

**Table S5: Patient reported outcome measures – treatment effects observed for EQ-5D-3L , OHS & OKS**

| | Day 14 | Day 42 | Day 120 | Day 365 |
|---|---|---|---|---|
| Treatment effect for the EQ-5D index PO vs. IV* | -0.003 (-0.052, 0.046), P=0.92 | 0.005 (-0.057, 0.066), P=0.88 | -0.032 (-0.1, 0.035), P=0.35 | -0.014 (-0.065, 0.038), P=0.61 |
| N | 596 | 631 | 554 | 533 |
| Treatment effect for the EQ-5D VAS PO vs. IV* | 0.206 (-3.243, 3.656), P=0.91 | 2.069 (-1.293, 5.431), P=0.23 | -1.64 (-6.082, 2.801), P=0.47 | -1.527 (-5.617, 2.562), P=0.46 |
| N | 571 | 610 | 533 | 514 |
| Treatment effect for the OHS PO vs. IV** | Not measured | Not measured | -2.416 (-8.503, 3.672), P=0.43 | -4.199 (-10.391, 1.992), P=0.18 |
| N | - | - | 137 | 109 |
| Treatment effect for the OKS PO vs. IV** | Not measured | Not measured | 5.766 (1.168, 10.364), P=0.01 | 4.407 0.366, 10.449), P=0.04 |
| N | - | - | 137 | 134 |

**The quantile regression models were adjusted for baseline EQ-5D-3L index/ VAS, gender, age, diagnosis of diabetes mellitus, ischaemic heart disease, peripheral vascular disease, previous stroke or TIA, use of immunosuppressive medication, rheumatoid arthritis, current smoking status, baseline surgical procedure, and infecting pathogen.*

*** The quantile regression models were adjusted for baseline OHS/ OKS, gender, age, diagnosis of diabetes mellitus, ischaemic heart disease, rheumatoid arthritis, current smoking status, baseline surgical procedure, and infecting pathogen.*

**Table S6: Self-reported adherence with antibiotics at day 14 & 42 using the Morisky Adherence Measure** [8] *(Maximum score is 8)*

| | **IV Antibiotic (N = 72†)** | **PO Antibiotic (N = 303)** | **Total (N = 375)** |
|---|---|---|---|
| Adherence score* (Day 14) | 8 (8, 8), (5, 8) | 8 (7, 8), (1, 8) | 8 (8, 8), (1, 8) |
| Adherence categories[a] (Day 14) | | | |
| High adherence | 49 (68.1%) | 207 (68.3%) | 256 (68.3%) |
| Medium adherence | 20 (27.8%) | 71 (23.4%) | 91 (24.3%) |
| Low adherence | 2 (2.8%) | 18 (5.9%) | 20 (5.3%) |
| Missing[b] | 1 (1.4%) | 7 (2.3%) | 8 (2.1%) |
| | **IV Antibiotic (N = 80)** | **PO Antibiotic (N = 323)** | **Total (N = 403)** |
| Adherence score* (Day 42) | 8 (7, 8), (4, 8) | 8 (7, 8), (0, 8) | 8 (7, 8), (0, 8) |
| Adherence categories[a] (Day 42) | | | |
| High adherence | 54 (67.5%) | 166 (51.4%) | 220 (54.6%) |
| Medium adherence | 21 (26.3%) | 117 (36.2%) | 138 (34.2%) |
| Low adherence | 3 (3.6%) | 25 (7.7%) | 28 (7.0%) |
| Missing[b] | 2 (2.5%) | 15 (4.6%) | 17 (4.2%) |

* *Median, IQR and range are displayed*

[a] *The 8 item Morisky Adherence scale is a self-report measure used as a screening tool to identify poorly adherent patients who are considered at risk of adherence related treatment failure. High adherence is defined by a Morisky score of 8; medium adherence is defined by a score of 6 or 7; low adherence is defined by a score of <6.*
*Frequencies (and percentages) are dispalyed.*

[b] *Adherence scores were considered missing if not all eight questions were answered*

† *For participants randomized to IV therapy, Morisky questionnaires were routinely requested only from patients who had been taught to self-administer their medication. Questionnaire returns were not required in cases where a healthcare professional was responsible for the administration of antibiotics.*

**Table S7: Compliance with PO antibiotics using Medication Event Monitoring System (MEMS)** [9]

| Compliance* (%) | Frequency (N = 62^) | Cumulative percentage (%) | Total number of doses missed (N = 4060) |
|---|---|---|---|
| 100 | 32 | 51.6 | 0 |
| 99 | 6 | 61.3 | 6 |
| 98 | 11 | 79.0 | 15 |
| 96 | 4 | 85.5 | 8 |
| 95 | 3 | 90.3 | 9 |
| 89 | 1 | 91.9 | 9 |
| 86 | 1 | 93.5 | 12 |
| 84 | 1 | 95.1 | 13 |
| 82 | 1 | 96.7 | 10 |
| 53 | 1 | 98.3 | 26 |
| 45 | 1 | 100 | 46 |

* *Expressed as number of doses taken/number of doses anticipated (%)*

^ *The MEMS subset included only consenting PO participants from four sentinel sites; it was not possible to subject IV therapy to MEMS technology*

**Table S8: Summary of intended agents specified before randomization for use if patient was subsequently to be randomized to IV**

| Intended IV agent | Randomized to PO* (N = 411) | Randomized to IV * (N = 506) | Total * (N = 917) |
|---|---|---|---|
| Glycopeptides | 164 (39.9%) | 216 (42.7%) | 380 (41.4%) |
| Penicillins | 23 (5.6%) | 16 (3.2%) | 39 (4.3%) |
| Cephalosporins | 160 (38.9%) | 185 (36.6%) | 345 (37.6%) |
| Carbapenems | 26 (6.3%) | 41 (8.1%) | 67 (7.3%) |
| Other single IV antibiotic | 22 (5.4%) | 22 (4.3%) | 44 (4.8%) |
| ^Combination IV antibiotics | 16 (3.9%) | 26 (5.1%) | 42 (4.6%) |

**Frequency and percentages are displayed*

^*There were 9 different planned IV combination therapies in the IV arm, the commonest being teicoplanin + ceftriaxone (10 participants) and teicoplanin + ertapenem (6 participants)*

**Table S9: Summary of intended regimes (excluding rifampicin) specified before randomization for use if patient was subsequently to be randomized to PO**

| Intended PO agent | Randomized to PO* (N = 507) | Randomized to IV* (N = 438) | Total * (N = 945) |
|---|---|---|---|
| Penicillins | 74 (14.6%) | 57 (13.0%) | 131 (13.9%) |
| Quinolones | 209 (41.2%) | 205 (46.8%) | 414 (43.8%) |
| Tetracyclines | 47 (9.3%) | 37 (8.5%) | 84 (8.9%) |
| Macrolides / Lincosamide | 58 (11.4%) | 50 (11.4%) | 108 (11.4%) |
| Other single PO antibiotic | 47 (9.3%) | 28 (6.4%) | 75 (7.9%) |
| ^Combination PO antibiotics | 72 (14.2%) | 61 (14.0%) | 133 (14.1%) |

**Frequency and percentages are displayed*

*^There were 19 different planned oral combination regimens in the PO arm, the commonest being ciprofloxacin + clindamycin (19 participants) and ciprofloxacin + doxycycline (17 participants); these figures do not take account of adjunctive rifampicin which was analysed separately.*

**Table S10: Overview of actual antibiotics (excluding rifampicin), as defined by agents used for more than one week during the initial six-week treatment period**

| | Participants randomized to IV Antibiotic* (N = 521) | Participants randomized to PO Antibiotic* (N = 523) | Total* (N = 1044) |
|---|---|---|---|
| Glycopeptides[a] (IV) | 214 (41.1%) | 22 (4.2%) | 236 (22.6%) |
| Penicillins (IV) | 38 (7.3%) | 11 (2.1%) | 49 (4.7%) |
| Cephalosporins (IV) | 173 (33.2%) | 8 (1.5%) | 181 (17.3%) |
| Carbapenems (IV) | 41 (7.9%) | 5 (1.0%) | 46 (4.4%) |
| Other single IV antibiotic | 35 (6.7%) | 2 (0.4%) | 37 (3.5%) |
| Combination IV antibiotics | 35 (6.7%) | 6 (1.1%) | 41 (3.9%) |
| Penicillins (PO) | 8 (1.5%) | 83 (15.9%) | 91 (8.7%) |
| Quinolones[b] (PO) | 33 (6.3%) | 191 (36.5%) | 224 (21.5%) |
| Tetracyclines[c] (PO) | 4 (0.8%) | 57 (10.9%) | 61 (5.8%) |
| Macrolides / Lincosamide [d] (PO) | 10 (1.9%) | 68 (13.0%) | 78 (7.5%) |
| Other single PO antibiotic (PO) | 10 (1.9%) | 54 (10.3%) | 64 (6.1%) |
| Combination PO antibiotics (PO) | 13 (2.5%) | 87 (16.6%) | 100 (9.6%) |

*The categories in this table were not mutually exclusive; 149 participants fell into more than one category and the data do not take account of adjunctive rifampicin which was analysed separately.*
**Frequency and percentages are displayed*

[a] *Glycopeptides were either teicoplanin or vancomycin*

[b] *Quinolones were ciprofloxacin in all but two cases, one each of moxifloxacin and levofloxacin. Of 191 participants in the oral arm who were prescribed quinolones, 160 (83.8%) were also prescribed rifampicin at some point during the trial.*
[c] *Doxycycline was the only tetracycline antibiotic prescribed.*
[d] *Macrolides were clarithromycin (4 cases) and erythromycin (2 cases); clindamycin was the only lincosamide used.*

**Table S11: Actual rifampicin use in 1049 participants**

| Observed rifampicin use[a] | Randomized to IV Antibiotic* (N=523) | Randomized to PO Antibiotic* (N=526) | Total* (N=1049) |
|---|---|---|---|
| No rifampicin use | 310 (59.3%) | 233 (44.3%) | 543 (51.8%) |
| <2 weeks[b] | 21 (4.2%) | 36 (6.8%) | 57 (5.4%) |
| 2 to 6 weeks[b] | 72 (13.8%) | 92 (17.5%) | 164 (15.6%) |
| >6 weeks[b] | 120 (22.9%) | 165 (31.4%) | 285 (27.2%) |

**Frequency and percentages are displayed*

[a] *The most commonly prescribed doses of rifampicic were 300mg BD (388 prescriptions) and 450mg BD (133 prescriptions).*
[b] *Based on the longest continuous period of use.*

**Table S12: Multiple imputation results for total non-surgical treatment costs and QALYs[10,11] through 1 year of follow up**

| Results | Randomized to IV Antibiotic* | Randomized to PO Antibiotic* | Difference | 95% CI |
|---|---|---|---|---|
| Costs | £13,274 (£446) | £10,534 (£453) | £2,740 (£638) | £1,488 to £3,992 |
| QALYs | 0.537 (0.013) | 0.545 (0.015) | -0.008 (0.019) | -0.045 to 0.031 |

** Means and standard errors are displayed*

## Supplementary References


1. US Department of Health and Human Services Food and Drug administration. Center for Drug Evaluation and Research (CDER). Guidance for Industry, Human Immunodeficiency Virus-1 Infection: Developing Antiretroviral Drugs for Treatment. . 2013;(Accessed at http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM355128.pdf).
2. US Department of Health and Human Services Food and Drug administration Center for Drug Evaluation and Research (CDER). Center for Biologics Evaluation and Research (CBER). Guidance for Industry. Non-Inferiority Clinical Trials. . 2010;(Accessed at http://www.fda.gov/downloads/Drugs/.../Guidances/UCM202140.pdf).
3. Bernard L, Dinh A, Ghout I, et al. Antibiotic treatment for 6 weeks versus 12 weeks in patients with pyogenic vertebral osteomyelitis: an open-label, non-inferiority, randomised, controlled trial. Lancet (London, England) 2015;385:875-82.
4. Lee CH, Steiner T, Petrof EO, et al. Frozen vs Fresh Fecal Microbiota Transplantation and Clinical Resolution of Diarrhea in Patients With Recurrent Clostridium difficile Infection: A Randomized Clinical Trial. JAMA 2016;315:142-9.
5. Mauri L, D'Agostino RB, Sr. Challenges in the Design and Interpretation of Noninferiority Trials. The New England journal of medicine 2017;377:1357-67.
6. Trueta J, Agerholm M. Acute haematogenous osteomyelitis; diagnosis and treatment. Postgrad Med J 1948;24:229-40.
7. Laptook AR, Shankaran S, Tyson JE, et al. Effect of Therapeutic Hypothermia Initiated After 6 Hours of Age on Death or Disability Among Newborns With Hypoxic-Ischemic Encephalopathy: A Randomized Clinical Trial. JAMA 2017;318:1550-60.
8. Morisky DE, Ang A, Krousel-Wood M, Ward HJ. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens (Greenwich) 2008;10:348-54.
9. Farmer KC. Methods for measuring and monitoring medication regimen adherence in clinical trials and clinical practice. Clin Ther 1999;21:1074-90; discussion 3.
10. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing... presumed at random: cost-analysis of incomplete data. Health Econ 2003;12:377-92.
11. Faria R, Gomes M, Epstein D, White IR. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. Pharmacoeconomics 2014;32:1157-70.