DEFINING ANTIBIOTIC TREATMENT DURATION AND ITS IMPACT ON ANTIMICROBIAL RESISTANCE

Mo Yin

St Hugh's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Clinical Medicine

Michaelmas 2021

Abstract

Reducing antibiotic treatment duration is a key stewardship intervention to mitigate antimicrobial resistance. In this thesis, I first address methodological issues introduced by non-adherence in non-inferiority trials, the commonest design for antibiotic duration randomised controlled trials. I conducted a simulation study and found that the probability of concluding non-inferiority when the treatment efficacy is actually inferior can be increased to as high as 0.1 from the acceptable 0.05 when adherence is relatively high at 90%. The simulations also highlighted the importance of anticipating patterns of non-adherence and accounting for this in power calculations. These findings were incorporated in the design of the "Reducing antibiotic treatment duration for ventilator-associated pneumonia (REGARD-VAP)" trial. The REGARD-VAP trial randomised patients with VAP to either a short (three to seven days) or a standard-of-care duration (eight days or more). The primary outcome was mortality or pneumonia recurrence by day 60. From May 2018 to July 2021, 340 patients were enrolled from seven hospitals in Nepal, Thailand, and Singapore. The second interim analysis, which included the first 231 patients, showed non-inferiority of the short duration arm based on a pre-defined Fleming-Harrington-O'Brien boundary. Adjusted analysis with inverse probability weighting suggested that a short duration was potentially superior to long duration. Lastly to examine shortened antibiotic duration as a strategy to reduce antimicrobial resistance in the hospital setting, I constructed agent-based models that incorporated within- and between-host dynamics of susceptible and resistant bacterial carriage in response to clinically-indicated antibiotic treatment. I found that shortening antibiotic duration is most effective at reducing resistance carriage in high transmission settings, and when resistant bacteria rapidly increase in abundance under antibiotic selection pressure and rapidly decrease when treatments are stopped amongst the treated individuals. A meta-analysis of antibiotic duration randomised trials showed an estimated odds ratio of 1.05 (80% credible interval 0.90 to 1.23%) with one additional day of antibiotic treatment. I conclude by discussing the future directions of the above work and the REGARD-VAP trial network.

DEFINING ANTIBIOTIC TREATMENT DURATION AND ITS IMPACT ON ANTIMICROBIAL RESISTANCE

Mo Yin

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Clinical Medicine

Michaelmas 2021

Acknowledgements

This thesis encapsulates the milestones of my work during the past four years, and offers only a glimpse of the countless precious moments during this tremendous journey. I owe these to my teachers, colleagues, friends and family who kept me thriving through even the most difficult times.

My sincere gratitude to Professor Ben S. Cooper for always rendering support, unreservedly sharing his wisdom and the tenacity to ensure that we produce the most robust and applicable work. I would like to thank Professor Paul A. Tambyah for his unwavering belief and always opening doors when the obstacles seemed insurmountable. Many thanks to A/Professor Direk Limmathurotsakul for the thoughtful advice which will continue to be a guiding light throughout my career.

I am extremely grateful for the collaborators and research teams at the REGARD-VAP study sites, many of whom suffered COVID-19 but stood by the trial through this extraordinary time. This work would not be possible without the vital inputs and support from my colleagues, many of whom have become life-long friends. My heartfelt thanks to Cherry Lim, Professor Mavuto Mukaka, James A. Watson, Professor Nicholas J. White, Professor Eoin T. West, A/Professor Hsu Li Yang, Jacob Armstrong, Thomas Crellen, Patrick Musicha, Mathupanee Oonsivilai, Anastasia Hernandez-Koutoucheva, Ricardo Sempedro, Jiraboon Tosanguan, Rene Niehus, Thi Mui Pham and Natcha Dankittipong.

Lastly, I would like to thank my friends and family for their love and tolerance, and for providing a home I always look forward to returning to.

Abstract

Reducing antibiotic treatment duration is a key stewardship intervention to mitigate antimicrobial resistance. In this thesis, I first address methodological issues introduced by non-adherence in non-inferiority trials, the commonest design for antibiotic duration randomised controlled trials. I conducted a simulation study and found that the probability of concluding non-inferiority when the treatment efficacy is actually inferior can be increased to as high as 0.1 from the acceptable 0.05 when adherence is relatively high at 90%. The simulations also highlighted the importance of anticipating patterns of non-adherence and accounting for this in power calculations. These findings were incorporated in the design of the "Reducing antibiotic treatment duration for ventilator-associated pneumonia (REGARD-VAP)" trial. The REGARD-VAP trial randomised patients with VAP to either a short (three to seven days) or a standard-of-care duration (eight days or more). The primary outcome was mortality or pneumonia recurrence by day 60. From May 2018 to July 2021, 340 patients were enrolled from seven hospitals in Nepal, Thailand, and Singapore. The second interim analysis, which included the first 231 patients, showed non-inferiority of the short duration arm based on a pre-defined Fleming-Harrington-O'Brien boundary. Adjusted analysis with inverse probability weighting suggested that a short duration was potentially superior to long duration. Lastly to examine shortened antibiotic duration as a strategy to reduce antimicrobial resistance in the hospital setting, I constructed agent-based models that incorporated within- and between-host dynamics of susceptible and resistant bacterial carriage in response to clinically-indicated antibiotic treatment. I found that shortening antibiotic duration is most effective at reducing resistance carriage in high transmission settings, and when resistant bacteria rapidly increase in abundance under antibiotic selection pressure and rapidly decrease when treatments are stopped amongst the treated individuals. A meta-analysis of antibiotic duration randomised trials showed an estimated odds ratio of 1.05 (80% credible interval 0.90 to 1.23%) with one additional day of antibiotic treatment. I conclude by discussing the future directions of the above work and the REGARD-VAP trial network.

Contents

List of Figures

List of Abbreviations

AMR . . . . . Antimicrobial resistance

ASP . . . . . Antimicrobial stewardship programme

ATS . . . . . American Thoracic Society

BSI . . . . . Bloodstream infection

CDC . . . . . Centers for Disease Control and Prevention

CONSORT . . Consolidated Standards of Reporting Trials

CRF . . . . . Case report form

CI . . . . . Confidence interval

CLSI . . . . . Clinical and Laboratory Standards Institute

COPD . . . . . Chronic obstructive lung disease

COVID-19 . . Coronavirus disease 2019

CrI . . . . . Credible interval

DIC . . . . . Deviance information criterion

DSMC . . . . . Data Safety and Monitoring Committee

EMA . . . . . European Medicines Agency

EUCAST . . . European Committee on Antimicrobial Susceptibility Testing

FDA . . . . . Food and Drug Administration

FiO2 . . . . . Fraction of inspired oxygen

HIV . . . . . Human immunodeficiency virus

ICU . . . . . Intensive care units

IDSA . . . . . Infectious Diseases Society of America

IQR . . . . . Interquartile range

IPW . . . . . Inverse probability weighting

ITT . . . . . Intention-to-treat

JAGS . . . . . Just Another Gibbs Sampler

MAP . . . . . Mean arterial pressure

MDRO . . . . Multi-drug resistant organisms

NHSN . . . . . National Healthcare Safety Network

OR . . . . . Odds ratio

OxTREC . . . Oxford Tropical Research Ethics Committee

PI . . . . . Principal investigator

PP . . . . . Per-protocol

PRISMA . . . Preferred Reporting Items for Systematic Reviews and Meta-Analysis

QALY . . . . . Quality-adjusted life year

RCT . . . . . Randomised controlled trials

REGARD-VAP REducinGAntibiotics tReatment Duration for Ventilator-Associated Pneumonia

SARS-CoV-2 . Severe acute respiratory syndrome coronavirus 2

SD . . . . . Standard deviation

SOC . . . . . Standard of care

SOFA . . . . . Sepsis-related Organ Failure Assessment score

SPIRIT . . . . Standard Protocol Items: Recommendations for Interventional Trials

SpO2 . . . . . Peripheral capillary oxygen saturation as measured by the pulse oximetry

TSC . . . . . Trial steering committee

VAP . . . . . Ventilator-associated pneumonia

WAIC . . . . . Widely Applicable Information Criterion

WHO . . . . . World Health Organization

1 Introduction

Contents

1.1 Background

Antimicrobial resistance (AMR) is a universally recognised public health crisis. The rise of AMR has threatened our ability to treat infections, increased risks in routine medical procedures, and severely impeded advancements in healthcare delivery. These dire consequences are expected to escalate as global antibiotic consumption continues to increase, mainly driven by low-and-middle income countries.[1] This is taking place against a backdrop of a progressively elderly and immunocompromised patient population that is highly susceptible to infections. Beyond the direct implications on health, the rippling effect of AMR challenges basic human rights targets including the United Nation's sustainable development goals such as health, responsible consumption and production, and poverty and inequality.[2]

Human consumption of antibiotics is a major driver of AMR.[3] Reducing unnecessary antibiotic use is thought to be the crux of mitigating AMR. Increasing calls to view antibiotics as a nonrenewable resource gave rise to systematic and evidence-based antibiotic stewardship programmes in the early 2000s.[4] The key antibiotic stewardship strategies can be broadly categorised as i) starting antibiotic treatments only when a bacterial infection is clearly indicated, ii) administering the optimal choice and dosing of antibiotic by considering the host, site of infection, pathogen, and pharmacokinetic/pharmacodynamic properties of the drug, and iii) limiting duration of treatment.[5–7] These programmes are strongly advocated by the World Health Organization (WHO) and international infectious disease authorities.

Reducing antibiotic treatment duration is one of the most commonly implemented antibiotic stewardship strategies reported in the literature,[8] and the one that is deemed to be safest and most acceptable by practicing clinicians.[9] This includes shortening courses for established bacterial infections, as well as rapid discontinuation of prescriptions after bacterial infections are ruled out. Numerous randomised trials, which have used either clinical criteria or inflammatory markers to aid in diagnosis and determine treatment response, have concluded that short treatment approaches for common bacterial infections are non-inferior to longer treatment courses in terms of clinical outcomes (Section 1.3).

This thesis focuses on optimising antibiotic treatment duration as an intervention to reduce AMR, one of the most practical and fundamental questions of antibiotic use. In this introductory chapter, I offer a comprehensive background on how antibiotic treatment durations for common bacterial infections have been defined, highlight the inherent challenges, and propose solutions. These will in turn form the rationale for the rest of the thesis. I also briefly cover the fundamental groundwork which enabled the research described in this thesis – the establishment of an inclusive multinational clinical research network, which comprises sites from AMR hotspots typically underrepresented in the literature but which bear the heaviest burden of AMR.

1.2 A historical perspective on antibiotic treatment duration

Prior to heightened concern about AMR, the conventional principle for duration of antibiotic therapy was to treat beyond clinical improvement in order to prevent both relapse of infection, and, counter to the tenet of this thesis, development of antibiotic resistance. The message of always completing antibiotic courses to prevent the development of resistance has remained widespread until recent years, promoted by the WHO, international health authorities, national health campaigns and school curricula.[9, 10]

Early knowledge about treatment failures with inadequate antibiotic duration emerged in the 1940s when penicillin was first used in clinical medicine. Abraham *et al.* reported in detail a case series involving patients with staphylococcal and streptococcal sepsis.[11] This was during a time when penicillin was produced in limited amounts locally in the laboratory and repeated doses of the drug were recovered from the patients' urine. Patients who initially responded to treatment eventually succumbed due to insufficient availability of drugs. When mass production of penicillin became available in mid-1940s, Keefer *et al.* found that the majority of 500 patients with community-acquired pneumonia recovered with only two to three days of penicillin.[12] In the following years, Dawson, Tillett, Meads *et al.* confirmed these findings,[13–15] but reported two instances of relapses.[15] They also identified the subgroups of patients who might require longer treatment duration, who were primarily those with pockets of pus such as empyema and local or systemic immunocompromised illnesses such as chronic obstructive pulmonary disease.[13] These potentially life-threatening treatment failures, coupled with the observation that antibiotic side effects were rare, drove the recommendations for community -acquired pneumonia to be ten to fourteen days until the early 2000s.[16–18]

Concerns about the development of antibiotic resistance with inadequate treatment were voiced initially by Fleming as he observed bacteria exposed to antibiotics *in vitro* become increasingly resistant.[19] His warning during his Nobel Prize acceptance speech that "underdosing" may promote resistance is highly cited. This is further supported by more recent *in vitro* models with *Streptococcus* spp., which reported that suboptimal antibiotic treatment may select for penicillin resistance.[20] Similar evidence from tuberculosis also showed that resistance seemed to arise within-host due to inappropriate treatment such as monotherapy or underdosing from patient non-adherence and pharmacokinetic variability.[21]

1.3 Systematic review of antibiotic duration randomised trials

The spotlight on excessive antibiotic use driving AMR in the late 1990s sparked numerous randomised controlled trials to shorten treatment duration. The majority of these trials used a non-inferiority design to challenge the standard-of-care treatment durations. Aside from directly comparing treatment durations, inflammatory markers such as C-reactive protein and procalcitonin, and individualised expert opinion such as infectious disease specialist consultations have also been used as trial interventions.[22]

To gain an in-depth understanding of the evidence base for antibiotic treatment duration, I performed a systematic review of all antibiotic duration randomised controlled trials performed to date. The specific objectives were to:

i) review antibiotic treatment durations for the various bacterial infections;

ii) identify gaps in evidence for antibiotic treatment durations in terms of settings, patient populations and infectious conditions;

iii) appraise trial methodologies and identify areas for improvement.

1.3.1 Methods

The systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.[23] Randomised controlled trials published up to 10 August 2021, and indexed in MEDLINE and EMBASE were sought.[24] Unpublished studies and pre-prints were excluded. The specific inclusion criteria were as follows:

i) **Participants**: Patients from both hospital and community settings who received antibiotics for the prevention or treatment of bacterial infections.

ii) **Intervention**: Varying duration of antibiotic treatment. The specific interventions which were used to guide treatment duration may include the use of inflammatory markers, expert opinion or electronic prescription systems.

iii) **Comparison**: Patients who received antibiotic(s) of different durations.

iv) **Primary or secondary outcome(s)**:

   a. **Clinical cure**: Number of patients who had clinical or microbiological cure, or infection relapses, or treatment failure.

   b. **Colonisation**: Number of patients who were colonised with antibiotic-resistant bacteria before and after antibiotic treatment. The sites of colonisation may include, but are not limited to, the digestive tract, respiratory tract, or urinary tract.

The exclusion criteria were:

i) **Participants**: Patients with non-infectious conditions; fungal or viral infections; tuberculosis.

ii) **Comparison**: Patients who received no antibiotic treatment, e.g. comparison between antibiotic treatment and a surgical procedure; patients who received a similar antibiotic in a slow-release formulation.

iii) **Outcome**: Feasibility for study intervention such as in pilot studies.

A Boolean search strategy with search terms pertaining to antibiotic treatment, duration, bacteria infections, human health, and randomised controlled trials was adopted (Table 1.1). Additional relevant titles found in other systematic reviews were also included.

After removing duplicates, the identified studies were screened according to the above inclusion and exclusion criteria. Full text for all studies included in the systematic review were retrieved for data extraction. Data extracted from all included studies were:

i) year of publication;

ii) type of bacterial infection including pathogen and site of infection;

iii) treatment durations compared;

iv) blinding of participants, prescribing physicians and investigators;

v) intervention to vary treatment duration;

vi) healthcare setting;

vii) age group of participants; and

viii) country/countries where patients were enrolled from.

Additional data were extracted from trials conducted in the last 15 years (2006 onwards) to evaluate the methodologies and quality of trial conduct. These additional data extracted included:

i) study hypothesis and sample size calculation;

ii) how the bacterial infection was identified as described in the inclusion criteria;

iii) choice of antibiotic treatment;

iv) randomisation process;

v) follow-up period;

vi) funding;

vii) non-adherence and how this was monitored;

viii) analysis methods;

ix) primary outcome;

x) whether antibiotic side effects and other health-economic outcomes were reported.

For randomised trials published in the past 15 years, studies which were adequately powered, i.e. reached the target enrolment number, were further assessed according to the key domains outlined in the revised tool for assessing risk of bias in randomised trials (RoB 2).[25] These domains included risk of bias arising from the randomisation process, deviations from the intended interventions, measurement and quality of outcome data, and selection of the reported result. In addition, quality of the statistical analysis was assessed by considering sample size calculations, choice of non-inferiority or equivalence margins and appropriateness in the choice of outcome.

Database	Search terms
MEDLINE	((antibiotic) AND (infection)) AND (weeks[Title]) **Filters**: Randomized Controlled Trial, Humans, 1920–2021 **Details**: (((((((("anti bacterial agents"[Pharmacological Action] OR "anti-bacterial agents"[MeSH Terms]) OR ("anti bacterial"[All Fields] AND "agents"[All Fields])) OR "anti bacterial agents"[All Fields]) OR "antibiotic"[All Fields]) OR "antibiotics"[All Fields]) OR "antibiotics"[All Fields]) OR "antibiotical"[All Fields]) AND ((((((((((((((((((((("infect" [All Fields] OR "infectability"[All Fields]) OR "infectable"[All Fields]) OR "infectant"[All Fields]) OR "infectants"[All Fields]) OR "infected"[All Fields]) OR "infecteds" [All Fields]) OR "infectibility"[All Fields]) OR "infectible"[All Fields]) OR "infecting"[All Fields]) OR "infections" [All Fields]) OR "infections"[MeSH Terms]) OR "infections"[All Fields]) OR "infection"[All Fields]) OR "infective"[All Fields]) OR "infectiveness" [All Fields]) OR "infectives"[All Fields]) OR "infectivities"[All Fields]) OR "infects"[All Fields]) OR "pathogenicity"[MeSH Subheading]) OR "pathogenicity"[All Fields]) OR "infectivity"[All Fields])) AND "weeks"[Title]
	((antibiotic) AND (infection)) AND (days[Title]) **Filters**: Randomized Controlled Trial, Humans, 1920–2021 **Details**: (((((((("anti bacterial agents"[Pharmacological Action] OR "anti-bacterial agents"[MeSH Terms]) OR ("anti bacterial"[All Fields] AND "agents"[All Fields])) OR "anti bacterial agents"[All Fields]) OR "antibiotic"[All Fields]) OR "antibiotics"[All Fields]) OR "antibiotics"[All Fields]) OR "antibiotical"[All Fields]) AND ((((((((((((((((((((("infect" [All Fields] OR "infectability"[All Fields]) OR "infectable"[All Fields]) OR "infectant"[All Fields]) OR "infectants"[All Fields]) OR "infected"[All Fields]) OR "infecteds" [All Fields]) OR "infectibility"[All Fields]) OR "infectible"[All Fields]) OR "infecting"[All Fields]) OR "infections" [All Fields]) OR "infections"[MeSH Terms]) OR "infections"[All Fields]) OR "infection"[All Fields]) OR "infective"[All Fields]) OR "infectiveness" [All Fields]) OR "infectives"[All Fields]) OR "infectivities"[All Fields]) OR "infects"[All Fields]) OR "pathogenicity"[MeSH Subheading]) OR "pathogenicity"[All Fields]) OR "infectivity"[All Fields])) AND "days"[Title]
	((antibiotic) AND (infection)) AND (duration) **Filters**: Randomized Controlled Trial, Humans, 1920–2021 **Details**: (((((((("anti bacterial agents"[Pharmacological Action] OR "anti-bacterial agents"[MeSH Terms]) OR ("anti bacterial"[All Fields] AND "agents"[All Fields])) OR "anti bacterial agents"[All Fields]) OR "antibiotic"[All Fields]) OR "antibiotics"[All Fields]) OR "antibiotic s"[All Fields]) OR "antibiotical"[All Fields]) AND ((((((((((((((((((((("infect"[All Fields] OR "infectability"[All Fields]) OR "infectable"[All Fields]) OR "infectant"[All Fields]) OR "infectants"[All Fields]) OR "infected"[All Fields]) OR "infecteds"[All Fields]) OR "infectibility"[All Fields]) OR "infectible"[All Fields]) OR "infecting"[All Fields]) OR "infections"[All Fields]) OR "infections"[MeSH Terms]) OR "infections"[All Fields]) OR "infection"[All Fields]) OR "infective"[All Fields]) OR "infectiveness"[All Fields]) OR "infectives"[All Fields]) OR "infectivities"[All Fields]) OR "infects"[All Fields]) OR "pathogenicity"[MeSH Subheading]) OR "pathogenicity"[All Fields]) OR "infectivity"[All Fields])) AND ("duration"[All Fields] OR "durations"[All Fields])
EMBASE	antibiotic AND infection AND week*:ti AND [randomized controlled trial]/lim AND [1920–2021]/py AND 'human'/de
	antibiotic AND infection AND day*:ti AND [randomized controlled trial]/lim AND [1920–2021]/py AND 'human'/de
	antibiotic AND infection AND duration AND [randomized controlled trial]/lim AND [1920–2021]/py AND 'human'/de

Table 1.1: Search terms used in the literature review.

1.3.2 Results

Overview of antibiotic treatment duration randomised trials

The initial electronic database search produced 3517 unique records published from 1969 to August, 2021 (Figure 1.1). Out of these, 283 fulfilled the inclusion criteria. Eighty-seven percent (247/283) of the trials concluded that there was either no statistical difference, equivalence or non-inferiority between the short and long treatment durations when clinical outcomes were compared. The median number of study participants randomised in the trials was 230 (IQR 100 to 448). Eighty-eight (31%) of the trials were double-blinded.

The most frequently studied bacterial infections were upper and lower respiratory tract infections (51 trials accounting for 18% and 61 trials accounting for 22% of the total number of trials respectively), genitourinary infections (49 trials, 17%), and post-surgical prophylaxis (46 trials, 16%) (Figure 1.2). Eighteen trials (6%) studied conditions with an infection focus that required surgical removal or debridement, and in 10 of these trials (10/18, 56%) the protocol mandated satisfactory source control in addition to antibiotic treatment.

In terms of patient characteristics, the most frequently represented groups consisted of adults from either the hospital general ward or the community in upper and upper-middle income countries (Figure 1.3). Only 23 (8%) of the trials were conducted in intensive care settings, and 35 trials (12%) enrolled patients from lower-middle or low income countries.

Favoured trial designs and methodologies have evolved with time (Figure 1.3). Prior to 2004, the antibiotic treatment duration trials exclusively tested various arbitrary durations. Starting in the late 2000s, biomarkers such as procalcitonin and C-reactive proteins were increasingly studied, but there has been a decrease in such studies since the mid-2010s. Another type of intervention that has emerged in recent years is the use of treatment protocols incorporating antibiotic stopping rules to terminate treatment according to individual clinical response.

Trial designs have shifted from superiority, which accounted for 74% (178/242) of trials prior to 2016, to non-inferiority trials, which accounted for 81% of all antibiotic duration trials conducted in the last five years (33/41 trials). The populations of study participants analysed for the trial hypothesis have also changed from only the intention-to-treat group (135/242 trials, 56%) prior to 2016 to also including the per-protocol participants (25/41 trials, 61%) in the last five years. The intention-to-treat analysis includes all participants who are randomised according to the group they were originally allocated to, regardless of the actual treatment they received. The per-protocol analysis includes only those patients who received the originally allocated treatment.

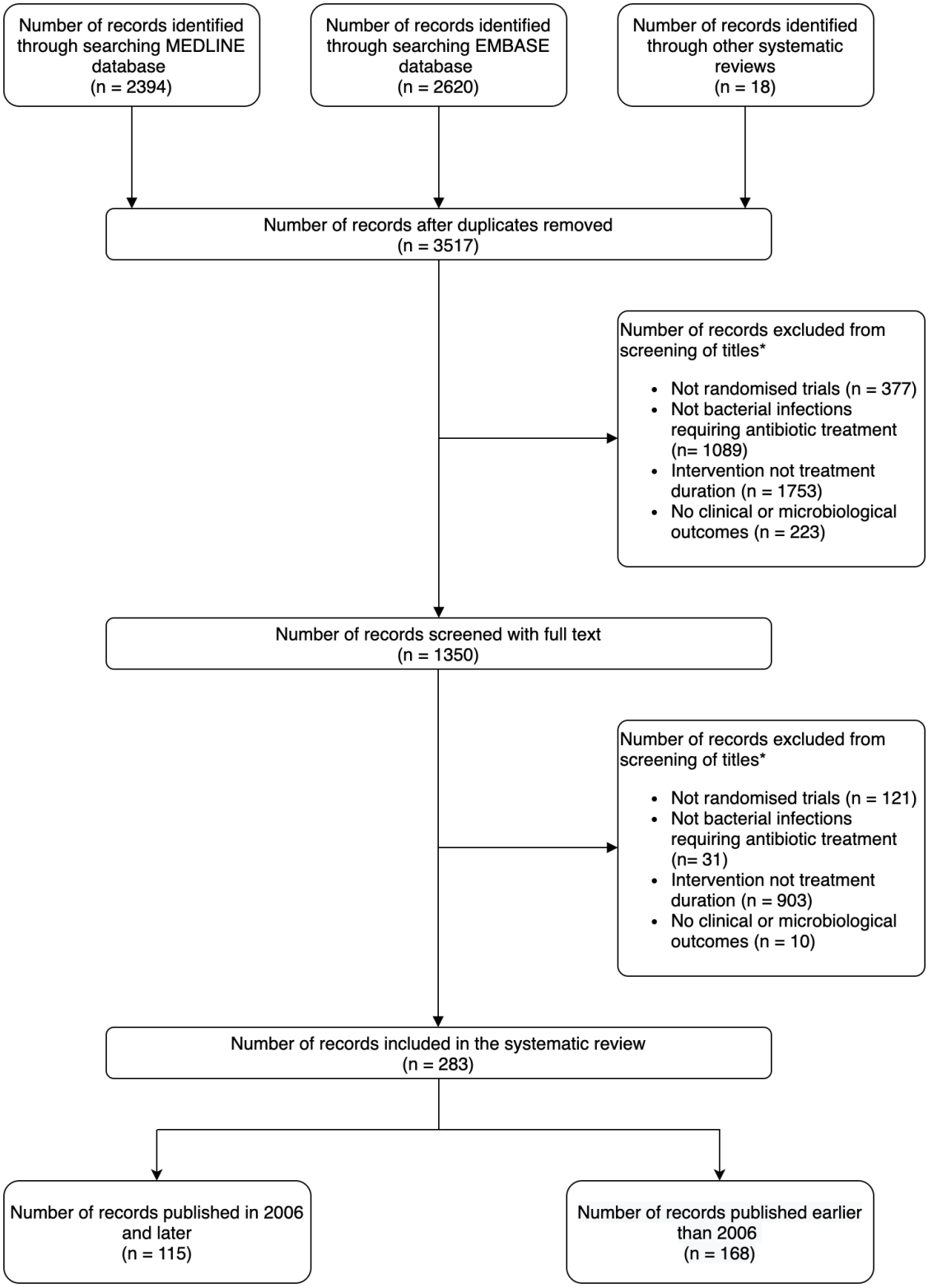


Figure 1.1: PRISMA flow diagram for systematic review on antibiotic duration randomised trials.

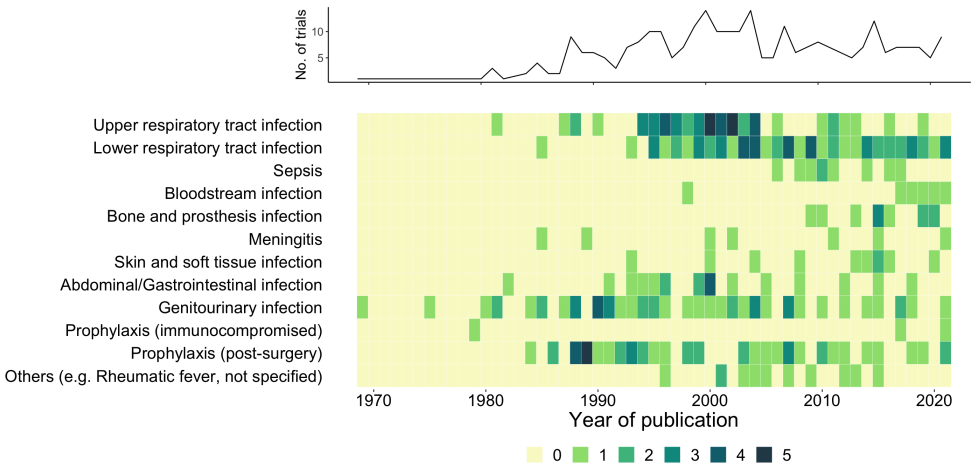


Figure 1.2: Types of bacterial infections studied by antibiotic duration randomised trials over time. The top panel shows the total number of published randomised trials per year from 1969 to 2021. The bottom panel shows the number of trials for each type of infection over time. The shading intensity of the green squares represents the number of trials per year.

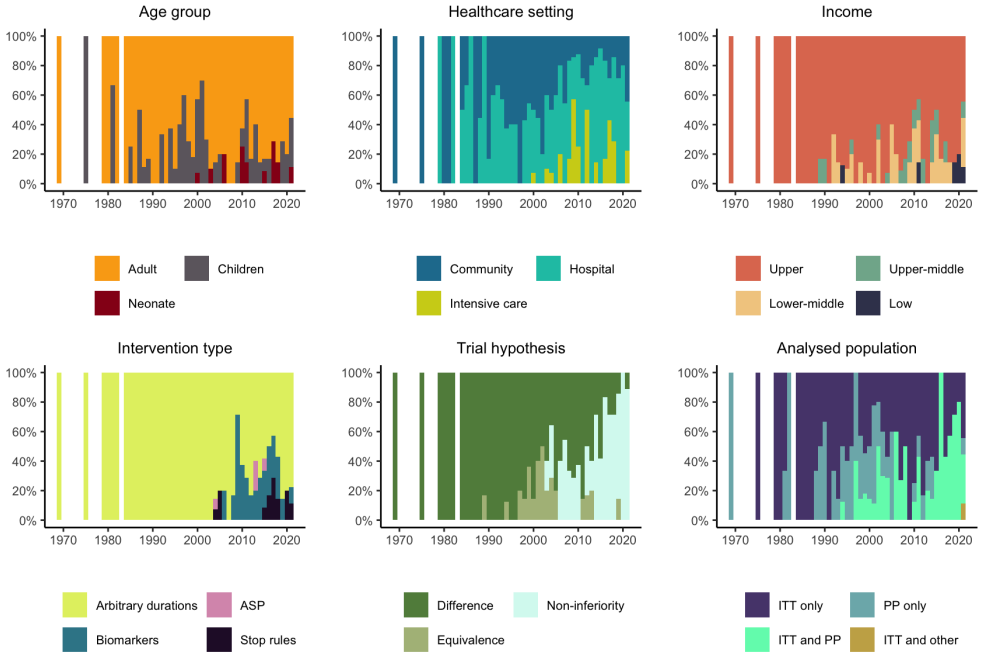


Figure 1.3: Characteristics of antibiotic duration trials over time. Each panel is labeled with a trial characteristic. These characteristics include i) age group of the trial participants, ii) the healthcare setting which the participants were enrolled from, iii) the minimum income level of the country/countries where the participants were enrolled from, iv) the type of intervention studied in the trials, v) the trial hypothesis design, and vi) the participant populations which the final conclusion of the trials were based on. ITT: intention-to-treat; PP: per protocol. Proportion of the trials published each year with a certain characteristic (y-axis) is plotted against the year of publication (x-axis) to illustrate the changes in these trial characteristics over time.

Quality of trial design, conduct and analysis

Intervention

Of 115 trials published in the last 15 years, 58 (50%) had included multiple antibiotic regimens within the randomisation arms. These included all of the trials using biomarkers and antimicrobial stewardship as the intervention. An additional 23 trials (20%) compared different antibiotics between the short and long arms, such as 5-day nitrofurantoin vs single-dose fosfomycin for the treatment of uncomplicated urinary tract infection.[26] In 52 trials (46%), the antibiotics prescribed were culture-directed, while the rest were empirical.

Twenty-eight trials (24%) were double blinded. These were trials that administered only one or two types of antibiotic treatment to the participants. Twenty-three trials (20%) were funded by a pharmaceutical or diagnostics company. Amongst these industry-funded trials, five studied a biomarker as the intervention while all others studied a particular type of antibiotic sponsored by the company.

Adherence to intervention and study population analysed

The sample sizes and analysed participant populations are shown in Figure 1.4. In 59 trials (59/115, 51%), randomised patients were left out of the final analysis (median 4% with IQR 2 to 9%). The commonest reasons for exclusion of these participants were protocol deviations, loss to follow-up, and consent withdrawal.

Eighty-three trials (72%) reported participants' adherence to randomised intervention. Adherence was reported in terms of either the proportion of participants who consumed a certain threshold of the antibiotic doses (frequently a criterion for per-protocol population) or the mean antibiotic duration actually observed in the respective randomisation arms. Forty-five out of those trials that reported a measure for adherence (45/83, 54%) reported participants from the short arm crossing over to the long arm (median 9%, IQR 0 to 15%). On the other hand, 44 trials (44/83, 53%) reported participants from the long arm crossing over to the short arm (median 7%, IQR 0 to 9%).

Ninety-four percent of the trials that reported the per-protocol number of participants had non-adherence (63/67). The mean proportion of non-adherent participants was 13% (IQR 7 to 21%). Forty-six trials reported the actual treatment durations observed in the respective randomisation arms. The difference between the actual and the intended treatment durations stated in the protocol ranged from -4 to 2 days (mean of 0 days). The commonest methods for adherence monitoring were medical chart review (46/83 trials, 55%) and self-reporting (20/83 trials, 24%).

Outcome assessment

One hundred and nine trials (95%) specified a follow-up period for the primary

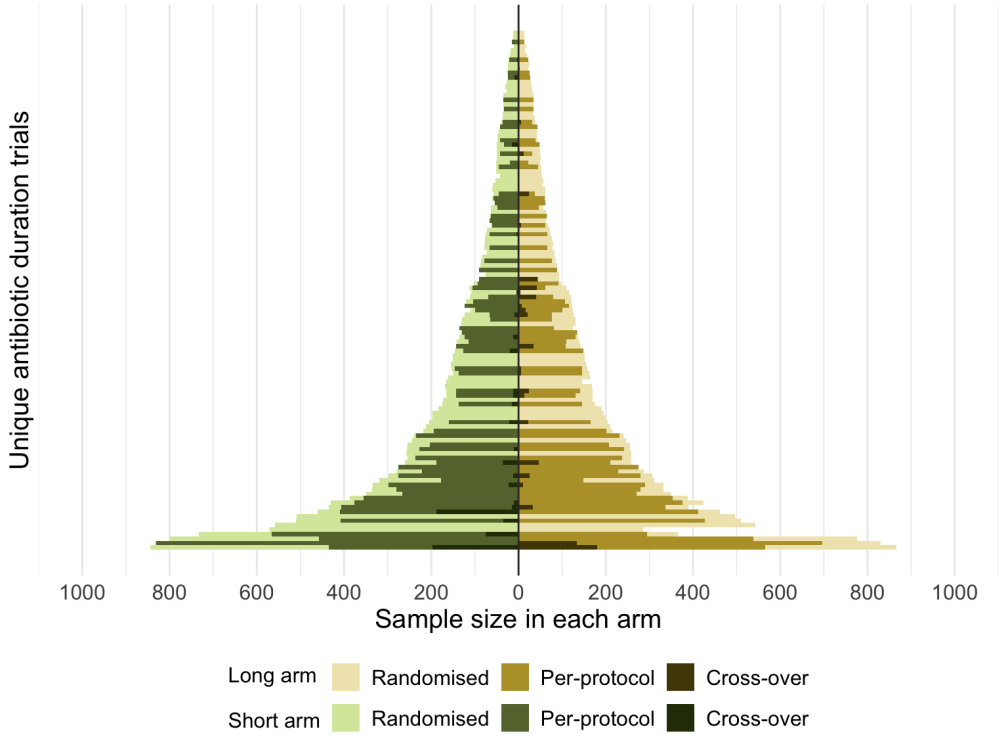


Figure 1.4: Comparison of randomised, analysed and per-protocol trial participants in antibiotic duration trials 2006–2021. One hundred and four antibiotic duration randomised trials are presented in the graph. Each horizontal bar represents a unique trial. One trial published within the period 2006–2021 with a sample size of more than 3000 participants is not shown on the graph to ensure visibility of the bars plotted from other trials. The green bars on the left represent trial participants who were randomised to the long duration arm; the gold bars on the right represent participants randomised to the short duration arm. The shade of the colours represents different populations of trial participants: the widest bars on each row represent the number of participants randomised in each trial (lightest colours); followed by the number of participants who were reported to be per-protocol; the darkest shades represent the number of participants who crossed over to the opposite arm.

outcome. In seven of these trials, follow-up periods were during hospitalisation or till the end of treatment. The median follow-up period was 28 days (IQR 15 to 40 days).

One hundred and ten trials (96%) specified a primary outcome. Most used clinical recovery, infection recurrence, microbiological eradication or mortality as the primary outcome (96/115, 83%), while the others used antibiotic duration or length of hospital stay. Ninety trials (78%) included at least one primary outcome which was subjective, and in only 26 (23%) were the assessors determining the outcome blinded to the randomisation arm. Ten trials (9%) used a composite outcome.

The overall mortality reported in the trials was low (median 0%, IQR 0 to 2%). Fifty-six trials (49%) reported the frequency of observed antibiotic side effects.

Similarly, fifty-six trials (49%) took follow-up samples from the participants to check for clearance or newly-acquired AMR bacteria at the sites of infection. Fifty trials (43%) reported a health-economic outcome, the most common being length of hospital stay (42/50, 84%). Only five trials (3%) took surveillance cultures during follow-up to assess for emergence of resistant bacteria colonisation.

Non-inferiority margin

About half of the trials used a non-inferiority design (61/115, 53%). Out of these, 55 reported a non-inferiority margin (90%) and all except one used absolute difference between the observed outcomes in the randomisation arms to calculate trial estimates. The mean non-inferiority margin was 10%, ranging from 0.35% to 25%. Figure 1.5 shows the observed proportion of trial participants who suffered negative clinical outcomes reported in these non-inferiority trials. Ninety-eight trials (85%) reported fewer than 10% of trial participants having a negative clinical outcome.

Bias assessment

Sixty-two trials (62/115, 54%) reported sample size calculations and completed enrolment. Assessment of the trials with the RoB2 tool identified 24 (39%), 34 (56%), and 4 (7%) trials to be at high risk, with some concerns, and low risk of bias respectively. The reasons which led to most studies being classified as having high risk of bias was because many of these trials were unblinded, suffered from a high degree of non-adherence or cross-overs, and that the analysis failed to account adequately for non-adherence (domain 2). These factors could potentially bias the trial estimates.

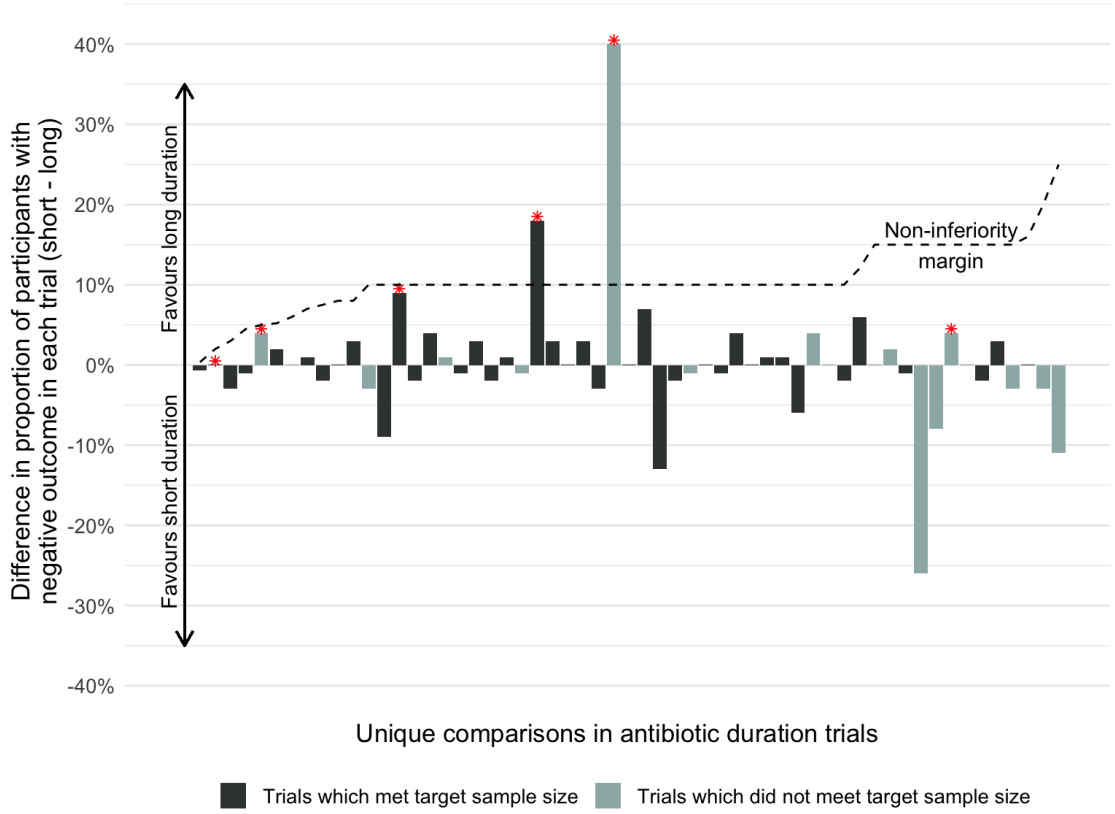


Figure 1.5: Trial estimates and respective non-inferiority margins in antibiotic duration non-inferiority trials. Each vertical bar represents a unique comparison between a long and a short duration. Fifty-eight comparisons are shown from 54 non-inferiority trials which reported clinical outcomes as the primary outcome. The bars represent the absolute difference in proportion of participants with negative clinical outcome in each comparison (point estimates calculated by the proportion of participants with negative outcome in the short arm minus that in the long arm). The light coloured bars represent trial comparisons that did not meet the target sample size. Trial comparisons that failed to conclude non-inferiority are marked with a red asterisk. These are the trials in which the upper bound of the confidence intervals crossed the non-inferiority margins. None of the trials concluded that a short duration was superior to long duration based on the primary outcomes. The dashed black line represents the non-inferiority margins.

Impact of antibiotic treatment duration trials on clinical guidelines

Well-conducted randomised controlled trials are considered the "gold standard" in empirical biomedical research, and help to establish clinical practice guidelines. Some successful examples for defining antibiotic treatment duration include community-acquired pneumonia, uncomplicated urinary tract infection, and post-surgery prophylaxis, which have been widely studied with randomised trials over the years (Figure 1.6). Antibiotic treatment duration for these infections has been reduced to three to five days for community-acquired pneumonia,[27] three days for uncomplicated urinary tract infection,[28] and a single dose for post-surgery prophylaxis.[29]

However, there are also conditions in which the practice guidelines have continued to support long treatment duration despite numerous randomised trials showing non-inferiority of a short course in terms of clinical outcomes. The reasons for this are varied and can be illustrated with the following examples:

- In acute group A streptococcal pharyngitis, though the clinical cure rate was similar between five and 10 days of penicillin V,[30] bacterial eradication was found to be consistently lower with a shorter treatment regimen.[31] Bacterial eradication is important in the treatment of group A streptococcal pharyngitis due to the potential immune sequelae including acute rheumatic fever and rheumatic heart disease, a leading cause of cardiovascular morbidity and mortality worldwide.[32]

- In the treatment for bacterial sinusitis, it has been shown in over a dozen randomised trials that there was no difference in response or relapse rates between fewer than six days and longer courses of antibiotics.[33] However, the heterogeneity in the trials in terms of symptom duration prior to enrolment and use of adjunctive medications led the guidelines to advise an arbitrary intermediate range of 5 to 7 days of treatment in adults.[34, 35] The recommendation for children remains at 10 to 14 days due to the lack of robust randomised trial data.

- Treatment duration for meningitis has mainly been studied in neonates and children. A previous meta-analysis published in 2009 found that four to seven days was similar to seven to 14 in terms of clinical cure rate.[36] Again, this failed to shift the practice guidelines, as important methodological issues were found in these trials including lack of blinding, small sample sizes, and relatively short follow-up. In addition, there was high variability in the causative pathogens in these trials. Despite one trial having over 1000 participants, it was deemed underpowered for interpreting organism-specific outcomes.[37, 38]

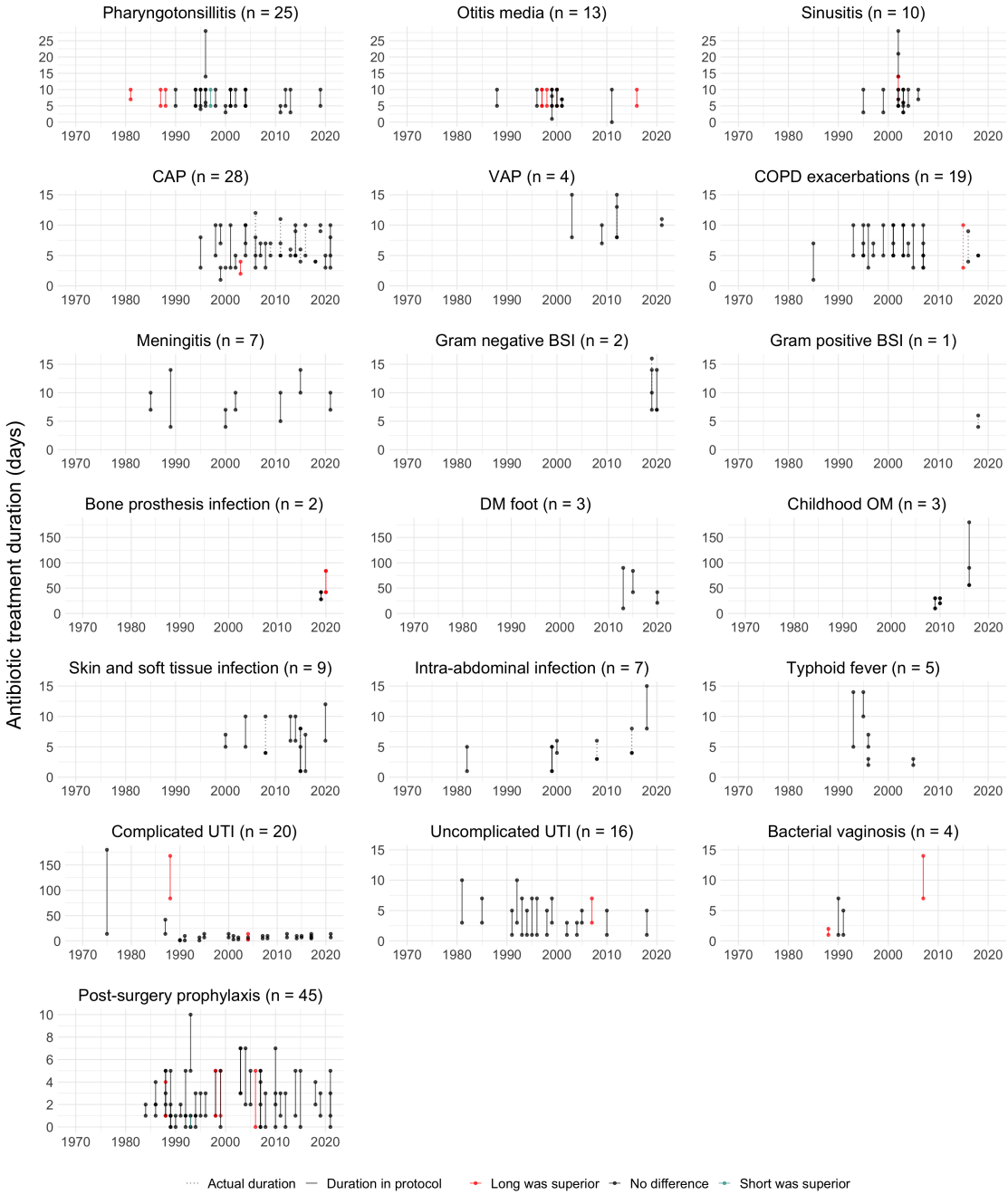


Figure 1.6: Antibiotic duration trials classified by bacterial infection syndromes. Each panel presents trial results for a type of bacterial infection. The number of trials included in each panel is shown in brackets. The vertical lines joining the points in each plot represent the durations compared in each trial. Red, black, and green vertical lines denote trials that concluded that long duration was superior, no difference or non-inferiority, and short duration superior, respectively, in terms of clinical outcomes. Solid vertical lines show the durations allocated to the study participants. Dotted vertical lines show the actual duration observed during the trial. Abbreviations: CAP - community-acquired pneumonia, VAP - ventilator-associated pneumonia, COPD - chronic obstructive lung disease, BSI - bloodstream infection, DM - diabetes mellitus, OM - osteomyelitis, UTI - urinary tract infections. The single trial shown in Gram positive BSI panel studied *Staphylococcus* BSI. The typhoid fever panel includes duration trials for both complicated and uncomplicated typhoid fever.

There remain gaps regarding relatively common conditions such as Gram-negative and *Staphylococcus aureus* bloodstream infections and ventilator-associated pneumonia caused by non-fermenting Gram-negative bacilli. The existing guidelines for these infections are based on weak or moderate evidence mainly from observational studies.[39, 40]

1.4 Considerations for defining antibiotic treatment duration

Increasing duration of antibiotic treatment conventionally decreases treatment failure, but this has to be balanced against more side effects, higher costs, and resistance selection. We can define optimal duration formally as one that minimises loss of quality-adjusted life years (QALY) (Figure 1.7). The ideal duration is hence the point where the combined QALY loss from treatment failure and side effects is minimised. Defining this optimal duration is challenging when treatment response and associated costs are context-specific and difficult to pinpoint.

The above systematic review has highlighted the key issues in the current antibiotic duration literature, which inadvertently introduced uncertainties in estimating both treatment response and costs. These issues threaten the trials' generalisability and internal validity in guiding decisions on treatment duration. Understanding these pitfalls and their impact on biasing trial estimates are critical first steps in designing future studies. In the following sections, I highlight these pitfalls using illustrative examples from the systematic review.

1.4.1 Evaluating treatment response

To precisely estimate treatment response, there need to be clear causal pathways between the disease, treatment and the monitored outcome. As with most diseases, there are substantial variations between patients' individual physiology and immunology systems and external environment, leading to different short- and long-term responses to the same treatment. Randomised trials address this by considering the average response from groups of patients who are allocated the same treatments within-group. However, there are features unique to bacterial infections that may reduce the strength of causal relationships between disease definition, treatment, and outcome.

i) Disease definition

- *Defining disease by anatomical site*: Different bacteria may lead to disease at the same anatomical site, e.g. *Staphylococcus* bloodstream infection may be caused by *Staphylococcus aureus*, which is associated with relatively high mortality and morbidity, or coagulase-negative *Staphylococcus*, which is usually considered a contaminant and does not require antibiotic treatment. Considering both organisms in the same

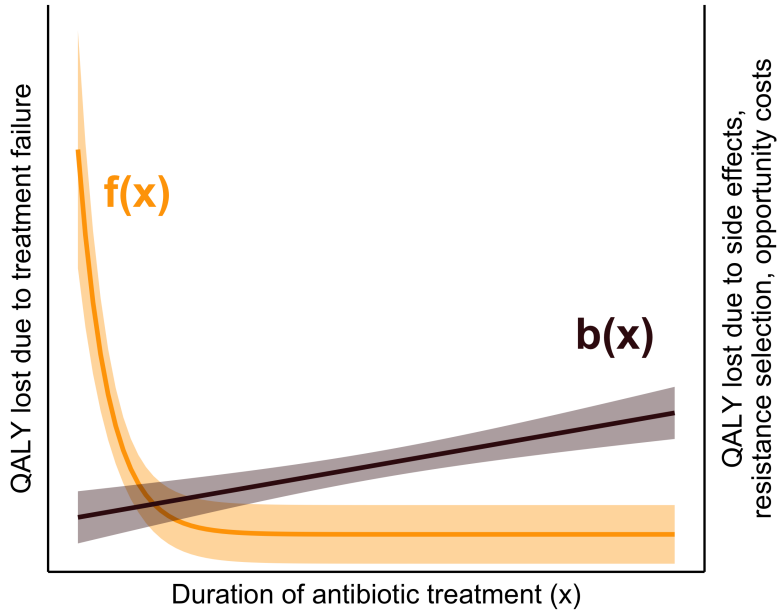


Figure 1.7: Considerations in defining antibiotic treatment duration. During antibiotic treatment, a patient is expected to recover, and quality-adjusted life years (QALY) lost due to treatment failure decreases with time (indicated by the yellow line labelled $f(x)$); QALY lost due to side effects, costs of antibiotic treatment, and resistance selection increase with treatment duration (indicated by the brown line labelled $b(x)$). The optimal antibiotic treatment duration is the x value that minimises $f(x) + b(x)$. The translucent areas represent the uncertainties in quantifying these variables. Uncertainties in quantifying $f(x)$ and $b(x)$ can be introduced by study design, and by poor trial conduct.

randomised trial will introduce large variations in treatment response among participants allocated the same antibiotic duration.[41]

- *Defining disease by bacteria*: The same bacterial infection may lead to diseases of varying severity or disseminated disease involving different anatomical sites, e.g. staphylococcal skin and soft tissue infection may remain localised but may also be complicated by invasive bloodstream infection leading to systematically disseminated infection. Treating both localised and disseminated disease with the same antibiotic duration is expected to lead to vastly different treatment responses.[42]
- *Lack of diagnostics to differentiate bacterial infection from viral or non-infectious conditions*: Many bacterial infections involve anatomical sites which are not normally sterile, such as the respiratory or urinary tracts. 'By-stander' bacteria, which form the healthy microbiome, cannot be easily differentiated from true pathogens. From the above systematic review, only 15% (43/283) of the antibiotic duration trials studied infections involving sterile sites. In addition, many studies enrolled patients based on clinical symptoms that are non-specific to bacterial infections. Examples include bacterial exacerbation of chronic obstructive pulmonary disease,[43] fast-breathing pneumonia in children,[44] and

ventilator-associated pneumonia.[45] This may result in a substantial proportion of patients who may not have any bacterial infection being enrolled into the trial.

ii) Treatment

- *Multiple antibiotic regimens*: Frequently there are multiple antibiotic regimens that can be used for the same infection. Seventy-five out of 115 trials (65%) in the systematic review had either various antibiotic regimens within the same randomisation arm or between the arms. This is because clinically, the treatment choice is based on the susceptibility profile of the bacteria, pharmacokinetic/pharmacodynamic properties, site of the infection, side effects profile, cost, and availability of the drug. Further variations are introduced when the initial antibiotic choice is often empirical, based on historical antibiogram data, and then later adjusted according to the microbiology susceptibility report.
- *Source control as a key treatment*: When infections progress to cause tissue damage and form foci of pus and bacteria, the more effective immediate treatment is source control (if the foci are accessibly surgically). Antibiotics act as adjunctive treatments. When applicable, the protocol should require adequate source control as part of the treatment intervention in all randomisation arms.
- *Non-adherence to allocated duration*: Non-adherence is ubiquitous in randomised trials. Unique to duration trials, non-adherence to allocated intervention often results in the participants crossing over from the short to long duration arm and vice versa. This direct cross-over was reported in over half of the 85 trials which reported a measure for adherence. An illustration of the impact of non-adherence in duration trials is the study by Arguedas *et al.*,[46] which compared single-dose extended-release azithromycin versus a 10-day regimen of amoxicillin/clavulanate for the treatment of children with acute otitis media. Adherence was found to be poor in the 10-day arm (88%) while adherence in the single-dose arm was ensured by direct observation in the clinics (100%). The trial concluded that the two duration arms are equivalent, with 81% of children in the single-dose and 85% of children in the 10-day group (95% confidence interval -10 to 3) responding clinically, even though more in the 10-day group had bacterial eradication (83% in the single-dose group versus 92% in the 10-day group). The authors considered the poor adherence in the long duration arm to be evidence justifying a shorter duration. However, if adherence to the 10-day regimen was improved there could have been a wider difference between the two regimens, favouring the longer treatment. The superiority of longer treatment in acute otitis media in children was subsequently shown in a later trial.[47]
- *Difficulties in blinding duration*: The majority of the trials were unblinded because the study participants required different antibiotic treatments

that catered to individuals' comorbidities and pathogens' antibiotic susceptibilities. The prescribing physicians' personal perception and practices on treatment duration will influence their, and potentially the patients', adherence to the allocated duration.

iii) Outcome

- *Choice of outcome*: In bacterial infections, the relevant trial outcomes could include immediate cure or long-term recurrence and relapse, microbiological eradication, and antibiotic consumption. The choice of outcome will influence the sample size calculation. Hence, meeting the sample size criterion based on one outcome may not produce adequate power to make meaningful conclusions for another. For example, in a trial in which the intervention is using procalcitonin,[48, 49] the sample size required to demonstrate reduced antibiotic use is usually not adequate to show a difference in clinical outcomes that tend to be more rare.
- *Follow-up period*: Many duration trials assess recurrence or relapse as an outcome.[45, 50] The differential time period during which these episodes were assessed and the potential for erroneous classification of persistent colonisation as recurrent infection may bias the short arm to appear to suffer more relapses or recurrences compared to the long arm. This was cited as a potential reason behind the reported association between Gram-negative non-fermenting bacilli and recurrences in ventilator-associated pneumonia antibiotic duration trials.[40]
- *Blinding of outcome assessment to allocation*: In the systematic review, 85% of trials used a subjective outcome, e.g. clinical symptom resolution or recovery. Because duration trials are often unblinded, these subjectivities in outcome determination may be biased when the assessors are aware of the allocation arm.

1.4.2 Determining non-inferiority

The majority of antibiotic duration trials in recent years have used a non-inferiority design. This is because the existing standards of care for many bacterial infections have been established and demonstrated to be highly efficacious, making demonstration of superiority against standard of care implausible, and placebo–control trials without any active comparators unethical to perform.[51, 52]

Non-inferiority trials compare the short with the standard treatment duration against a margin which the investigators are willing to sacrifice in terms of treatment effects, to reap benefits such as cost savings or reduced side effects. A short treatment can be shown to be non-inferior if its effect is 'no worse than' the standard duration by this margin of tolerance when compared to the standard duration. All of the issues highlighted in the earlier section (Section 1.4.1) can potentially lead to the effects of the two duration arms appearing more similar, and thereby increase the probability of demonstrating non-inferiority of the short arm even when the short

duration is actually inferior in clinical efficacy.

Another important consideration in the design of an antibiotic duration trial is deciding whether the primary interest is in efficacy, defined as the performance of an intervention under ideal and controlled circumstances, or effectiveness, which refers to its performance under 'real-world' conditions.[53] To maximise the chance of estimating efficacy, the included patients will have to be highly selected and relatively homogeneous. On the other hand, a focus on effectiveness may allow the inclusion criteria and intervention to be less stringent. Hence in effectiveness trials, external factors from the patients, prescribers and the wider healthcare system are incorporated, which might moderate the intervention's efficacy. The limitation of this latter approach is that the findings may not be easily generalisable in another setting when the systemic and adherence patterns are different.

The choice of non-inferiority margin is a subject of debate across various fields of medicine and is widely discussed in guidelines.[51, 54, 55] The non-inferiority margin is usually defined with prior knowledge of the standard-of-care treatment's efficacy compared with a placebo, and consensus from subject matter experts.[52] With antibiotic duration trials, placebo–control trials are frequently lacking, and the efficacy of standard-of-care treatments increases with time due to improvements in general healthcare delivery.[56] In addition, the actual event rate in the trial should also be monitored and compared against that used in the power calculations. The systematic review showed that 31% of the antibiotic duration non-inferiority trials reported a very low event rate of $< 10\%$. A lower than expected event rate can lead to insufficient power, favouring non-inferiority.[57]

1.4.3 Quantifying individual and population cost of antibiotic use

The penalties of prolonged antibiotic courses are the side effects and associated costs for both treated individuals and the population. Direct side effects of antibiotics for the treated individuals include allergies, kidney and liver injuries, and opportunistic infections such as *Clostridium difficile* colitis. Though mostly mild, these side effects are not infrequent and have been reported in up to 20% of hospitalised patients who received antibiotics.[58, 59] At a population level, the hypothesis that limiting antibiotic duration leads to a reduction in AMR remains unproven. Previous large population-level observational studies have shown that trends between specific antibiotic exposure and AMR are not always predictable,[60, 61] likely due to the heterogeneity of the interventions and outcome measures.

As noted in the systematic review, the majority of antibiotic duration trials investigated the short-term clinical outcomes of reduced antibiotic duration on individual patients, but few followed up on treated patients to determine subsequent AMR colonisation or infection. This is echoed by earlier reviews of studies attempting to quantify the impact of stewardship strategies, which found that microbiology

outcomes are seldom reported, especially those from longer term follow-ups.[8]

1.5 Thesis outline

The above appraisal of the antibiotic duration literature yields several key findings. Firstly, antibiotic duration trials have been successful in shortening treatment duration for some bacterial infections such as community-acquired pneumonia and urinary tract infections. However, gaps remain, especially for conditions associated with higher mortality, critically ill patients and patients from low-middle income countries.

Secondly, non-inferiority is the most commonly adopted design for antibiotic duration trials. When designing antibiotic duration non-inferiority trials, investigators should be cognizant of the features concerning disease definition, treatment, and outcome assessment that may result in outcomes in the randomisation groups appearing more similar, as these will increase the probability of demonstrating non-inferiority. Focusing on effectiveness may be more informative than efficacy in many instances, e.g. using a set of inclusion criteria that is more sensitive than specific may be more reflective of real-world practice. However, lapses in the conduct of the trial, such as non-adherence, are potentially not generalisable, and should be minimised. Non-adherence in non-inferiority trials, though commonly reported, is not adequately addressed in terms of analysis methods and reporting from international guidelines.[51, 54]

Lastly, there is limited evidence to show the effectiveness of shortening antibiotic duration in reducing AMR. The effect of treatment duration on the development and spread of AMR is expected to vary due to multiple pathogen, host, and environmental factors. A clear theoretical rationale for the approach is lacking.

This thesis investigates the above issues with a multifaceted approach, using skill sets from statistics, causal inference, mathematical modelling, and the design and implementation of randomised controlled trials.

- **Chapter 2: Statistical considerations in non-inferiority trials with non-adherence**
  In this chapter, I dissect the various patterns of non-adherence and assess their effect on non-inferiority trials. I describe a simulation study of a hypothetical non-inferiority trial with binary outcomes to: i) explore the impact of various patterns of non-adherence and analysis methods on trial treatment effect estimates; ii) quantify the probability of claiming non-inferiority when treatment efficacy is actually inferior; iii) compare and evaluate alternative analysis methods such as inverse probability weighting and instrumental variable estimation; and iv) provide an accessible tool for investigators to design non-inferiority trials that anticipate non-adherence. I also suggest

measures to improve study design, statistical analysis and reporting to address this issue.

- **Chapter 3: REducinG Antibiotics tReatment Duration for Ventilator-Associated Pneumonia (REGARD-VAP)**
  The REGARD-VAP trial is a multi-centre randomised controlled hierarchical noninferiority-superiority trial to assess the clinical effect of a short versus standard-of-care duration in adults with VAP. The trial intervention is a shift from the conventional 'one-size-fits-all' treatment duration, and allows individualisation of treatment duration for patients responding to treatment. In this chapter, I share the study protocol and findings from the second interim analysis.

- **Chapter 4: Effect of antibiotic treatment duration on AMR in the hospital setting**
  I modelled within- and between-host dynamics of susceptible and resistant gut bacteria in response to systemic antibiotic treatment to gain a mechanistic understanding of the intermediary steps relating antibiotic duration to the emergence and maintenance of resistant bacteria colonisation. To confirm the model findings, I performed a systematic review and meta-analysis of randomised controlled trials that studied the effect of antibiotic treatment duration on colonisation with resistant bacteria.

1.5.1 Formation of an inclusive antimicrobial resistance trial network

The disproportionate burden of AMR in low-middle income countries, as well as the lack of representation of these countries in academic research, are symptoms of inequality. Weak governance, lack of infrastructure, and limited regulated access to appropriate antimicrobials against common infections are some of the factors fueling antibiotic consumption in these regions.[62] They also act as deterrents for quality research to be conducted, especially randomised controlled trials.

A research network that includes both high and low-middle income countries is potentially beneficial on multiple fronts when an ethically sound mutual benefit framework is upheld. Firstly, an 'emerging' resistant bacteria in a high-income country is often already prevalent in low-middle income countries. Potential funders from high-income countries keen to investigate emerging AMR pathogens may ensure enrolment from partnering sites in low-middle income countries, during which new drugs or treatment protocols will become available in areas where the AMR burden is high, with the potential for these treatments to become cheaper and accessible once efficacy is shown. Secondly, on a micro-level, exchange of expertise in conducting research activities promotes capacity building in a sustainable way. Local ownership of data collected through research may raise awareness of important health issues and further empower sites to initiate new quality improvement or

research projects.[63]

The network in which REGARD-VAP was implemented initially consisted of the Sunpasitthiprasong Hospital, Singapore National University Hospital, and Tan Tock Seng Hospital. To ensure participation and adherence to the study protocol, I briefed the doctors, nurses, and pharmacists from all intensive care units (ICUs) in these centres. While it was relatively straightforward to engage with the Singaporean sites, Sunpasitthiprasong Hospital proved to be challenging as there were no dedicated teams in the ICUs. I had to hold individual meetings with doctors and nurses to address their concerns. The other sites in Thailand and Nepal were established subsequently through local collaborators.

The central values of the REGARD-VAP network are sustainable capacity building, transparency, and equity. With the expansion of the network, there were ample opportunities for the research teams to exchange technical expertise and cultural practices. The teams have travelled to various other sites during monitoring visits and attend educational courses. Many have leveraged their experience with REGARD-VAP to pursue further studies. Local investigators were also encouraged to utilise the data collected from the trial and propose new collaborations.

Figure 1.8: The REGARD-VAP trial network.

1.5.2 Published work

The following chapters contain work that has been published in peer-reviewed journals. I confirm that I authored and produced the following publications with no more input or advice than would be considered appropriate or acceptable for a thesis. The relevant sections which are re-used from these publications have been referenced at the appropriate parts.

- **Chapter 2: Statistical considerations in non-inferiority trials with non-adherence**

  i) Yin Mo, Cherry Lim, Mavuto Mukaka, Ben S Cooper. Statistical considerations in the design and analysis of non-inferiority trials with binary endpoints in the presence of non-adherence: A simulation study. Wellcome Open Res. 2020;4:207. doi:10.12688/wellcomeopenres.15636.2

  ii) Yin Mo, Cherry Lim, James A Watson, Nicholas J White, Ben S Cooper. Non-adherence in non-inferiority trials: Pitfalls and recommendations. BMJ. 2020;370:m2692. doi:10.1136/bmj.m2692

- **Chapter 3: REducinG Antibiotics tReatment Duration for Ventilator-Associated Pneumonia (REGARD-VAP)**

  i) Yin Mo, West TE, MacLaren G, *et al.*. Reducing antibiotic treatment duration for ventilator-associated pneumonia (REGARD-VAP): A trial protocol for a randomised clinical trial. BMJ Open. 2021;11:e050105. doi: 10.1136/bmjopen-2021-050105

2

Statistical considerations for non-inferiority trials with non-adherence

Contents

2.1 Introduction

Randomised controlled trials that test for non-inferiority of the experimental arm are performed when a new treatment is compared with an established standard of care. Instead of being required to have superior clinical efficacy, the new treatment might be preferred for its improved safety, convenience, or reduced cost. These trials are increasingly prevalent because highly efficacious standard of care treatments have been established for many diseases, making demonstration of superiority against standard of care implausible and placebo-controlled trials without any active comparators unethical to perform.[51, 52]

A basic weakness of non-inferiority trials compared with superiority trials is that poor conduct of the trial or deviations from the protocol may result in false rejection of the null hypothesis that the experimental treatment is inferior. Most trials report that some participants do not adhere to their allocated treatment. Non-adherence, unlike treatment assignment, will often not be an independent event but driven by confounding and/or non-confounding factors (Figure 2.1). Non-confounding factors affect the probability of adhering to the intervention but do not directly affect the study outcome. An example may be intolerance to study medications due to mild side effects such as nausea or rash, which does not directly affect disease outcome, but may cause enough discomfort to affect adherence.[64, 65] This contrasts with non-adherence driven by factors that directly influence the study outcome, such as disease severity.[66, 67] For example, consider an open label study where more severely ill patients are more likely not to adhere to the experimental treatment and these patients cross over to the standard of care control arm. On the other hand, patients with less severe disease are more willing to adhere to the experimental treatment. In this case, disease severity is a confounder because it affects both adherence and disease outcome. Comparing the groups of patients according to the actual treatments received will be biased, as they have different disease severities.

It is important to assess the patterns of non-adherence that might occur in a non-inferiority trial during the planning stage both to inform power calculations and to enable development of an appropriate analysis plan.[68, 69] However, no easily accessible tools are currently available to guide investigators in non-inferiority trial design while accounting for these considerations.

In this chapter, I go beyond the topic of antibiotic duration and consider all late-phase non-inferiority randomised controlled trials with non-adherence. I first describe findings from a systematic review of recent non-inferiority trials and guideline recommendations. This is followed by a simulation study of a hypothetical non-inferiority trial with binary outcomes. The simulation aimed to: i) explore the impact of various patterns of non-adherence and analysis methods on trial treatment effect estimates; ii) quantify the probability of claiming non-inferiority when treatment efficacy is actually inferior; iii) compare and evaluate alternative analysis methods such as inverse probability weighting and instrumental variable estimation; and iv) provide a tool for investigators to design non-inferiority trials

that anticipate non-adherence. Lastly, I propose measures to improve study design, statistical analysis, and reporting to address this issue.

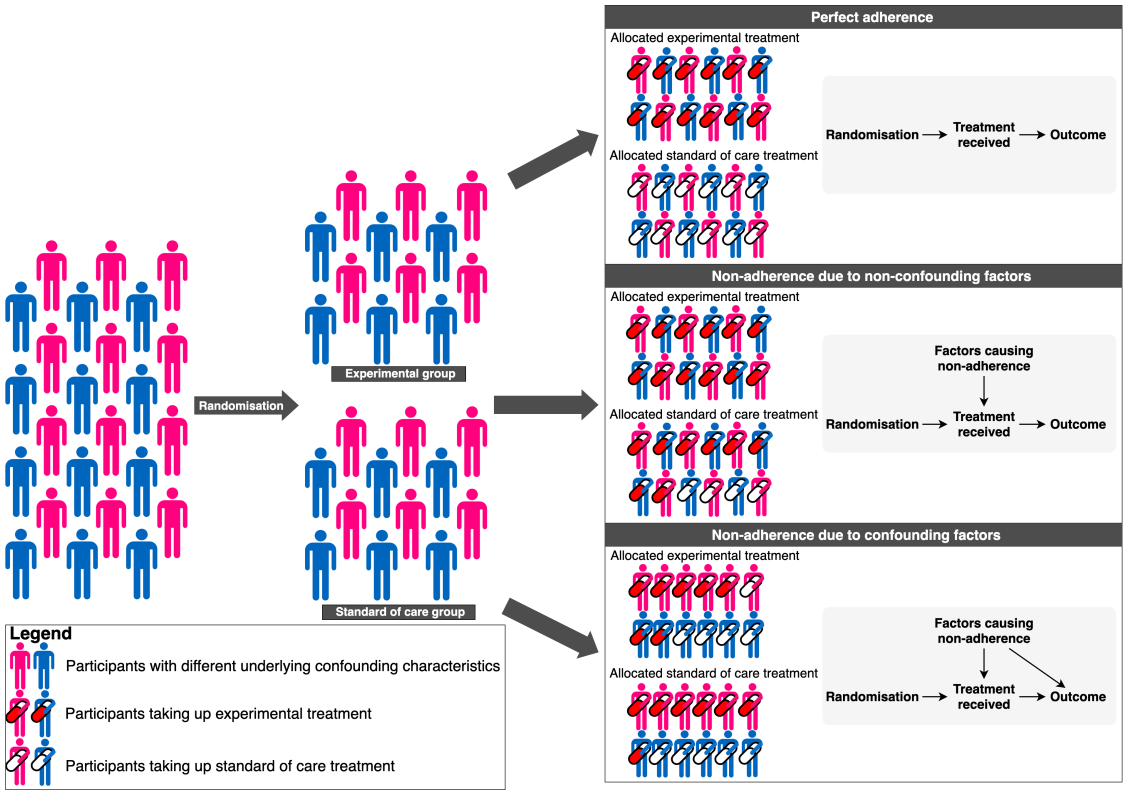


Figure 2.1: Common patterns of non-adherence in clinical trials. Blue and pink figures represent participants with different underlying confounding characteristics. Red and white pills represent experimental and control treatments respectively. The directed acyclic graphs in the right-side panels illustrate the causal pathways from treatment allocation to outcome, highlighting the mechanisms causing non-adherence to allocated treatments.

2.1.1 Conventional analysis methods for non-inferiority trials

Intention-to-treat (ITT) analysis estimates the treatment effect accounting for 'real-world' adherence patterns by comparing outcomes between groups of participants defined by their allocated treatment. Often called an effectiveness trial, it measures the effect of allocating a treatment on participant outcomes, instead of the actual effect of treatment. If the primary research interest is the causal effect of assigning treatments, then this is likely to be the most relevant estimate. In other situations, the question of primary interest is the causal effect of the treatment itself. Because many patterns of non-adherence result in reduced observed differences between the comparison arms, there is a risk that relying on the ITT analysis to conclude non-inferiority will lead to the adoption of treatments which, when applied, lead to

worse outcomes.[70–72]

Many trials also report the per-protocol (PP) analysis, which estimates the effect of the treatment only from participants who received the treatment according to randomisation assignment. However, because adherent participants may differ systematically in underlying prognostic factors from non-adherent patients, and the PP participants in each allocation group may differ in terms of prognostic characteristics, the PP analysis can give biased treatment effect estimates. In non-inferiority trials, this may lead to false conclusions of non-inferiority when the treatment effect is actually inferior.

Most non-inferiority trials continue to rely on ITT and PP analyses even in the presence of high degrees of non-adherence. These issues have been highlighted in international guidelines and simulation studies, but consensus on the best way forward has not been reached.[51, 54, 72, 73] Of note, Kim has previously shown that the standard approaches can lead to erroneous conclusions about treatment efficacy in non-inferiority trials with non-adherence, and so proposed using an instrumental variable estimator as an alternative statistical method.[74] Sanchez and Chen reached a similar conclusion: depending on the pattern of protocol deviation, both PP and ITT populations may show non-inferiority when the treatment effect is actually inferior.[68] Practical guidance is needed to inform trial designs on how incremental levels of non-adherence affect the chance of reaching different trial conclusions.

2.1.2 Challenges of non-inferiority trials

In non-inferiority trials, we pose the question: "is the new treatment no worse than the standard of care treatment?" in contrast to "is the new treatment better than the standard of care treatment?" in a typical superiority trial. This shift in the focus of the comparison complicates non-inferiority trials for two main reasons.

The first complication is in deciding what we mean by "no worse than". If the new treatment does lead to worse outcomes, but only by a small amount, we might reasonably conclude that it is non-inferior. The largest such "small amount" that is compatible with a conclusion of non-inferiority is known as the non-inferiority margin. It is a practically acceptable compromise in treatment efficacy that we are willing to sacrifice in exchange for the secondary benefits offered by the new treatment. The subjective nature of this measure arises from the debate around what margin is 'clinically acceptable' and how the advantages of the new treatment are weighed against the potential loss in treatment efficacy; it also needs to be decided whether the non-inferiority margin should be on a relative or absolute scale.[54] Non-inferiority margins are often chosen for practical reasons such as reduction of sample size while maintaining adequate power to conclude non-inferiority. Despite the development of numerous objective methods to justify the non-inferiority margin, its determination remains contentious and highly context-specific.[52, 55]

The second complication is that poorly designed and conducted non-inferiority trials will often have an increased chance of concluding non-inferiority.[54, 68] Some examples include:

i) Non-specific endpoint measures; e.g. using 30-day mortality as the primary outcome in a trial comparing drugs for treating cardiac arrhythmia in the intensive care unit. Even if one treatment is more effective than the other, the measurable difference will be diluted by mortality due to other reasons such as sepsis or hypovolemic shock.[75]

ii) Inappropriate participant cohort; e.g. a trial comparing conservative medical treatment (control treatment) versus percutaneous coronary intervention (experimental treatment) in patients with stable angina but relatively good exercise tolerance, with a primary endpoint defined as exercise increment at 6 weeks. The majority of participants, regardless of allocation, would not be expected to achieve this endpoint even though there may be a clinically significant difference between the two arms in patients with lower baseline exercise tolerance.[76] Conversely, choosing patient groups with a high chance of spontaneous cure may also give misleading results; e.g. in malaria-endemic areas adults commonly self-cure, and treatment responses with ineffective medicines may produce excellent outcomes, but the same treatments in children can lead to high failure rates.[77]

iii) Markedly different pharmacokinetic properties between the treatments with insufficient follow-up; e.g. when the outcome of interest is recurrence in the treatment of malaria, recurrence may be delayed by slowly eliminated drugs. Terminating follow-up before all recurrences have occurred will favour the more slowly eliminated drug.[77, 78]

To counteract such problems, all major guidelines for non-inferiority trials emphasise choosing appropriate control treatments that have been previously demonstrated to be superior to placebo, and ensuring consistency in study design with the historical placebo-controlled studies that established the standard of care.[51, 54, 78] While these are useful regulatory approaches for licensing new drugs, in situations where placebo-controlled trials were never performed, the appropriate choice of participants and outcomes becomes a more contentious issue.

Another challenge prevalent in clinical trials is non-adherence to allocated treatment: when non-adherence leads to a lower average treatment effect measured in the control group, or similar treatment effects measured in both groups, the experimental group will be more likely to appear non-inferior. Both types of non-adherence are frequently observed in non-inferiority trials. Because the control is usually a clinically available standard of care treatment, non-adherence often leads to study participants taking up a treatment from the opposite arm or taking an alternative treatment with similar efficacy to the control.

2.2 Systematic review of recent non-inferiority trials and guideline recommendations

2.2.1 Recent non-inferiority trials

I searched five medical journals publishing clinical trials (New England Journal of Medicine, the Lancet, Journal of American Medical Association, Annals of Internal Medicine and the British Medical Journal) from 1 January 2017 to 31 May 2019 for phase III and phase IV randomised clinical trials. These articles were retrieved from MEDLINE, Web of Science, Google Scholar and EMBASE.

Out of a total of 425 unique phase III and phase IV randomised controlled trials, 100 aimed to demonstrate non-inferiority in their primary outcomes (24%). The main specialisation domains of the studies were infectious disease (27%, 27/100), cardiology or cardiothoracic surgery (23%, 23/100), and oncology (12%, 12/100).

A majority (86%, 86/100) concluded non-inferiority. The primary analysis population used to determine non-inferiority was ITT in 82 (82%) and PP in 15 (15%); 3 (3%) used both ITT and PP populations. Eighty-three (83%) trials reported the number of participants in PP populations, out of which 82 (99%, 82/83) reported a proportion of non-adherent participants. The median difference between the number of patients included in the ITT and PP populations was 9% (IQR 5–16%). Out of 70 trials that analysed the PP population (excluding three with protocols not published online), 44 (63%, 44/70) predefined PP in the protocols.

PP definitions were varied, and included adherence to protocol-specified interventions, measurement of outcomes, and/or follow-up schedules. Out of 95 trials that reported the pattern of non-adherence to the allocated intervention, crossing over to the opposite arm took place in 55 trials (58%, 55/95). Fifty-five out of 74 trials (74%, 55/74) with prolonged time-varying interventions had some measures for adherence to trial interventions, of which the most common method was patient-reported compliance via pill count or diaries (33%, 18/55). Seventy-five trials (75%, 75/100) reported the degree of adherence to allocated interventions.

Thirty-eight trials (38%) reported non-adherence in more than 10% of the study participants. In these trials, 15 calculated unadjusted ITT and/or PP estimates as primary and supplementary analysis (39%, 15/38). Seventeen trials (45%, 17/38) adopted PP estimates adjusted with prognostic factors as either primary or secondary analysis.

Case study demonstrating the effect of non-adherence on intention-to-treat and per-protocol analyses in a non-inferiority trial

A study compared disease activity–guided dose reduction versus continuous prescription (control treatment) of biologic disease-modifying antirheumatic drugs in patients with rheumatic arthritis.[79] The primary outcome was the proportion of

participants who experienced a major flare by day 180 of follow-up. In the continuous treatment arm, 15% (9/59) of patients had dose reduction, as the participants either had low disease activity or developed side effects and could not tolerate continuous treatment. In the dose reduction arm, 37% (45/121) had continuous treatment due to poorly controlled disease. The study concluded non-inferiority based on a PP analysis (absolute risk difference 2%, 95%CI −12% to 12%) given a non-inferiority margin of 20%. Supplementary analysis with ITT concurred with the PP analysis.

However, crossing over of participants may have resulted in the PP patients in the dose reduction arm including more patients with mild disease, and the PP patients in the continuous dosing arm including more patients with severe disease. Were there such a difference in baseline disease severity in the two per-protocol groups, the dose reduction group would be likely to have fewer patients with major flares than the continuous group. The PP estimate may therefore be biased in favour of the dose reduction group. Crossing over of participants resulted in a proportion of participants receiving the treatment of the opposite arm. In an ITT analysis, this dilutes the treatment effect difference measured between the two arms. In this example, both ITT and PP estimates have a heightened risk of claiming non-inferiority.

2.2.2 International guidelines on non-inferiority trials

International guideline recommendations for study analysis and reporting are heterogeneous.[80] On the issue of addressing non-adherence, most guidelines caution investigators to design and conduct non-inferiority trials carefully to reduce non-adherence as much as possible, as adjustment for poor adherence may not be possible.[51, 73] The FDA guidelines in 2016 warned against using ITT analysis but did not recommend any alternative methods of analysis.[54] The most recent guidelines from EMA in 2000 stated that ITT and PP are equally important and that similar results should be demonstrated for the full analysis set and PP analysis set.[81] The latest CONSORT 2012 guideline advised using a non-ITT analysis method as a supplementary analysis or a hybrid ITT/PP method, without specifying its methodology.[51, 68]

In addition, there is no requirement for reporting the definitions of the analysis populations. Varied definitions of PP and the frequently used modified ITT population often lead to different estimates, and therefore impact the determination of non-inferiority.[80, 82, 83]

2.3 Impact of various non-adherence patterns and analysis methods on non-inferiority trial treatment effect estimates in the presence of non-adherence: A simulation study

2.3.1 Methods

Simulation models and analysis methods

I simulated a two-arm non-inferiority randomised controlled trial, where the randomised treatment, Z, and outcome, Y, are binary and time fixed. Randomisation, Z, is done in a 1:1 ratio. Actual treatment received is denoted A. An example of such a trial is the study on optimising antibiotic treatment duration for community–acquired pneumonia.[84] The experimental treatment is five days of antibiotic treatment ($A = 1$), while the control treatment is a duration as decided by the physicians ($A = 0$). Outcome is treatment failure as defined by a set of questionnaire scores on day 30 ($Y = 1$ represents treatment failure, $Y = 0$ represents treatment success). With this single end-point, I consider adherence as a binary variable where non-adherent patients in the short arm would receive more than five days of treatment, and non-adherent patients in the long arm would receive fewer than five days of treatment. The effect estimate is the absolute risk difference, calculated as the difference in the proportion of participants with treatment failure between treatment arms.

I calculated the sample size based on the assumption that 40% of patients in both experimental and control arms experience treatment failure, with a non-inferiority margin of 10% and tolerable type 1 error of 0.025. This required 505 participants per arm for 90% power.[85] Each simulation was done with 1000 iterations. All simulations and analyses were carried out with R.[86] The simulation code is available on GitHub (https://github.com/moyinNUHS/NItrialsimulation).

Notation

I broadly adopt the notation of Hernán and Robins.[87] In the subsequent paragraphs, $Y^{a=0}$ represents the potential outcome if the control treatment were to be administered ($A = 0$); $Y^{a=1}$ represents the outcome that would occur if the experimental treatment were to be administered ($A = 1$); and $Y^{a=2}$ represents the potential outcome if an alternative inferior treatment compared to both the control and experimental treatments were to be administered ($A = 2$). For an individual i, $Y_i^{a=0}$, $Y_i^{a=1}$ and $Y_i^{a=2}$ are therefore counterfactual outcomes. I use the shorthand Y_i and Y_i^a to refer to the actual and potential outcomes Y and Y^a respectively for a given individual i. Because only one of the outcomes is observed in the real world, the actual observed outcome, Y_i, is either equal to $Y_i^{a=0}$, $Y_i^{a=1}$ or $Y_i^{a=2}$ depending on the treatment received, i.e. $Y_i = Y^{a=A_i}$, where $A_i = 0$ if the individual received the control treatment, $A_i = 1$ if the individual received the experimental

treatment, and $A_i = 2$ if the individual received an alternative treatment inferior to both the control and experimental treatments.[88] Similarly, the observed outcomes depending on randomisation (Z) are represented by $Y_i^{z=1}$ and $Y_i^{z=0}$ respectively, and the actual treatments depending on randomisation are represented by $A_i^{z=0}$ and $A_i^{z=1}$. C refers to the confounding factors that may increase or decrease the probabilities of adhering to the allocated treatment and outcome.

Simulation models

I generated individual level data which included the following variables: treatment allocation, participant characteristics that may affect adherence and outcome, actual treatment received, counterfactual outcomes and observed outcomes. Allocation is a binary variable with each individual having a 50% probability of being allocated to the experimental treatment. Participant characteristics were represented by a single continuous variable on the interval [0, 1] drawn from a beta distribution. This can be thought of as a disease risk score.[89]

I considered two common reasons for non-adherence. The first is when non-adherence is due to factors that affect the probability of taking up the allocated treatment but do not affect the study outcome through any other pathway (Figure 2.2, panel A). The second is driven by confounders, defined as the study participants' prognostic factors that affect both the probability for taking up the allocated treatment and the outcome (Figure 2.2, panel B).

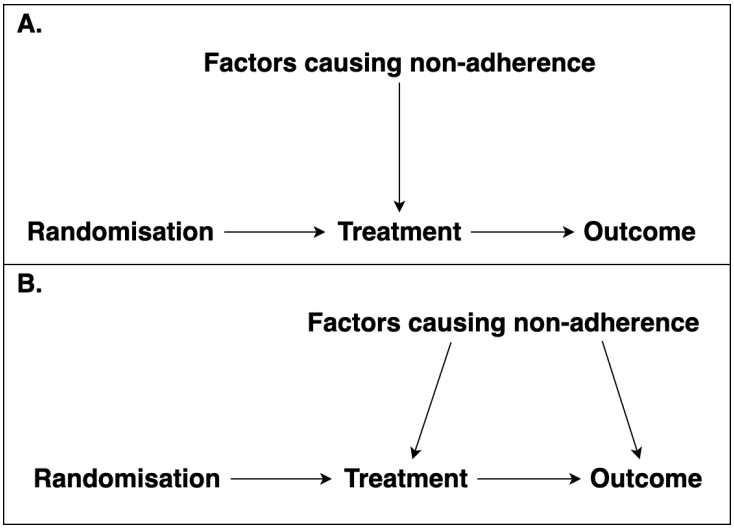


Figure 2.2: Directed acyclic graphs demonstrating the causal relationships of the variables generated for each study participant in the simulation study. In scenario A, factors that cause non-adherence affect the probability of the participant taking up the allocated treatment but do not directly affect the outcome; e.g. minor side effects of the treatment drug. In scenario B, factors causing non-adherence affect both the probability of the participant taking up the allocated treatment as well as the outcome; e.g. disease severity.

The actual treatment received by an individual differs from the allocated treatment when there is non-adherence. I considered scenarios where non-adherent participants cross over to the opposite treatment arm, or receive alternative treatments that are inferior to both the control and experimental treatments. In the case where the participant characteristics encourage an individual to switch to an experimental treatment, the individual's probability for crossing over to the experimental treatment when randomised to the control arm is increased. An example is a trial studying an experimental treatment for a terminal disease which has few effective treatment options. An individual with more severe disease may be more likely to switch to the experimental treatment even when the person is randomised to the control treatment. In another case where the factor causing non-adherence discourages an individual from taking up an experimental treatment, the individual's probability for adhering to the experimental treatment after being randomised to the experimental arm is decreased. The individual might take up the control treatment or refuse treatment altogether. An example is a trial comparing an experimental exercise regime to nicotine patches for smoking cessation. An individual with chronic obstructive lung disease may be more likely to be non-adherent to the experimental exercise regime and take up the nicotine patch or decline all treatments.

I generated counterfactual outcomes for each individual, one for experimental treatment, one for control treatment, and one for alternative treatments inferior to both the control and experimental treatments. The overall averaged difference between the counterfactual outcomes for experimental and control treatments for all study participants is the pre-defined true treatment effect assumed in the simulations. The participant characteristics may cause an increase or decrease in the probability of having the outcome, depending on the direction of influence the confounder has on the outcome. The observed outcome is then chosen from one of the counterfactuals depending on the actual treatment that the individual received. Detailed descriptions of the simulations are included below in Box 2.1.

Box 2.1 Details of the simulation model

The simulation model generates four key random variables for each study participant, namely:

i) treatment assignment (i.e. randomisation, Z)

ii) participant characteristics score which may affect the probability of adhering to treatment assignment and treatment failure, C

iii) actual treatment taken up, A

iv) outcome, Y

Subscripts i and j are used to indicate the values of these variables for participant i in iteration j. In each iteration, an independent set of simulated trial data is

generated and analysed to give an estimate of the treatment effect.

The strength of association between the participant characteristics score, probability of adhering to treatment assignment, and treatment failure directly affects the degree of bias in commonly used treatment effect estimates. I define these relationships by generating these variables independently of each other and then, in each iteration, using the participant characteristics scores to rank the participants in terms of their probability of adhering to the treatment assignment and treatment failure. The ranking approach is adopted to ensure that the model used to generate the data is different from the models used to analyse the data.

In each iteration j, data for n participants per arm were generated. Half of the participants were randomised to the experimental treatment arm ($Z = 1$), the other half to the control treatment arm ($Z = 0$).

Step 1: Generating distributions for participant characteristics scores, probabilities of adhering to treatment assignment, and treatment failure.

i) Let C_{ij} represent the participant characteristics for participant i in iteration j that affect the participant's probability of treatment failure,

$$C_{ij} \sim \text{Beta}(\alpha = 2, \beta = 2)$$

Without loss of generality, I assume that the patient indices i are ordered such that $C_{1j} < C_{2j} < \cdots < C_{nj}$.

ii) Let p_j^z represent the probability a participant in iteration j adheres to the allocated treatment, $\Pr[A^z = z]$,

$$p_j^z \sim \text{Beta}(\alpha_j, \beta_j),$$

where

$$\alpha_j \sim \text{Uniform}(2, 10);$$

$$\beta_j = \begin{cases} \frac{\alpha_j(1-k_1)}{k_1} & \text{when } \ z = 1 \\ \frac{\alpha_j(1-k_0)}{k_0} & \text{when } \ z = 0 \end{cases}$$

where k_1 and k_0 are the means of the corresponding beta distributions i.e. k_1 is the expected proportion of adhering participants in the experimental treatment arm, and k_0 is the expected proportion of adhering participants in the control treatment arm. $z = 0$ and $z = 1$ represent randomisation to the control and experimental arms respectively. α_j is drawn from a uniform distribution from 2 to 10, which was chosen to prevent the generation of very small or large probabilities for p_j^z. This ensured positivity in our data generation mechanism,

meaning that there will be study participants in both the experimental treatment and control treatment arms across the various participant characteristic values.

iii) Let q_j^a represent the probability a participant in iteration j has treatment failure, $\Pr[Y^a = 1]$,

$$q_j^a \sim \text{Beta}(\alpha_j, \beta_j),$$

where

$$\alpha_j \sim \text{Uniform}(x_{\min}, x_{\max});$$

$$\beta_j = \begin{cases} \frac{\alpha_j(1-l_0)}{l_0} & \text{when } a = 0 \\ \frac{\alpha_j(1-l_1)}{l_1} & \text{when } a = 1 \\ \frac{\alpha_j(1-l_2)}{l_2} & \text{when } a = 2 \end{cases}$$

where $a = 0$, $a = 1$ and $a = 2$ represent control, experimental and alternative treatments respectively. $x_{\min}$ and $x_{\max}$ determine the distribution of the probabilities of treatment failure, and therefore the influence of the confounder on treatment failure. l_1 is the expected proportion of treatment failure in the experimental treatment arm, l_0 is the expected proportion of treatment failure in the control treatment arm, and l_2 is the expected proportion of treatment failure in the non-adhering participants receiving an alternative inferior treatment compared to both the control and experimental treatments.

Step 2: Ranking of participants according to participant characteristics and probability for adherence.

i) In the scenario where non-adherence is driven by non-confounding factors, i.e. patient characteristics do not affect adherence to assigned treatment, n values from p_j^z are drawn and randomly indexed $p_{ij}^z : \ i \in (1, \ldots, n)$.

ii) In the scenario where non-adherence is driven by confounding factors, i.e. patient characteristics do affect adherence, n values are drawn from p_j^z and indexed according to the patient characteristics C_{ij}.

There are two scenarios:

a. In the case when patient characteristics increase the probability of taking up the experimental treatment in both arms, the probabilities are indexed such that the values are strictly increasing $p_{1j}^z < p_{2j}^z < \cdots < p_{nj}^z$. This ensures that patients with larger values of C_{ij} have larger values of p_{ij}.

b. Conversely, when patient characteristics decrease the probability of taking up the experimental treatment in both arms, I index the probabilities such that the values are strictly decreasing $p_{1j}^z > p_{2j}^z > \cdots > p_{nj}^z$. This ensures that patients with larger values of C_{ij} have smaller values of p_{ij}.

Step 3: Ranking of participants according to participant characteristics and probability of treatment failure. n values are drawn from q_j^a and indexed according to the patient characteristics C_{ij}.

There are two scenarios:

i) In the case when patient characteristics increase the probability of treatment failure, the probabilities are indexed such that the values are strictly increasing $q_{1j}^a < q_{2j}^a < \cdots < q_{nj}^a$. This ensures that patients with larger values of C_{ij} have larger values of q_{ij} .

ii) Conversely when patient characteristics decrease the probability of treatment failure, the probabilities are indexed such that the values are strictly decreasing $q_{1j}^a > q_{2j}^a > \cdots > q_{nj}^a$. This ensures that patients with larger values of C_{ij} have smaller values of q_{ij} .

Step 4: Actual treatments received and treatment failure.

i) Let A_{ij} represent the actual treatment received by each individual i,

$$A_{ij} \sim \text{Bernoulli}\left(p_{ij}^Z\right)$$

ii) Each individual has three counterfactual outcomes. Let $Y_{ij}^{a=0}$ be the outcome if an individual takes up the control treatment, $Y_{ij}^{a=1}$ if an individual takes up the experimental treatment, and $Y_{ij}^{a=2}$ if an individual takes up the alternative inferior treatment compared to the control and experimental treatments.

$$Y_{ij}^{a=0} \sim \text{Bernoulli}\left(q_{ij}^{a=0}\right),$$

where $q_{ij}^{a=0}$ is the probability given participant i takes up the control treatment.

$$Y_{ij}^{a=1} \sim \text{Bernoulli}\left(q_{ij}^{a=1}\right),$$

where $q_{ij}^{a=1}$ is the probability given participant i takes up the experimental treatment.

$$Y_{ij}^{a=2} \sim \text{Bernoulli}\left(q_{ij}^{a=2}\right),$$

where $q_{ij}^{a=2}$ is the probability given participant i takes up the alternative inferior treatment compared to both the control and experimental treatments.

The observed outcome, Y_{ij}, is one of the counterfactual outcomes depending on the actual treatment received.

Analysis methods

The ITT analysis considers all randomised participants according to their assigned arms, regardless of whether the participants had the intended treatment. It estimates

the effect of Z on Y, i.e. $\Pr[Y = 1|Z = 1] - \Pr[Y = 1|Z = 0]$. The PP analysis only considers participants who received treatment according to their allocation stated in the study protocol, i.e. $\Pr[Y = 1|A = 1, Z = 1] - \Pr[Y = 1|A = 0, Z = 0]$.

In addition, I used an inverse probability weighting approach on the PP population to estimate the causal effect of treatment on the outcome. This approach applies a logistic regression model incorporating the confounder as an explanatory variable to estimate an individual's probability of adhering to a particular allocation arm. The inverses of these predicted probabilities are used as weights to inflate or deflate the individual's influence on the overall treatment effect in the arm.[90] The inverse probability weighting estimate, μ_{ipw}, is given by (for simplicity iteration subscripts j are omitted):

$$\mu_{ipw} = \mathbb{E}\left[Y^{a=1} - Y^{a=0}\right] = \frac{1}{n}\sum_{i}^{n}\frac{A_i Y_i}{\Pr[A = 1 \mid C]} - \frac{1}{n}\sum_{i}^{n}\frac{(1 - A_i)\, Y_i}{1 - \Pr[A = 1 \mid C]},$$

where C represents measured confounders.

The denominators are estimated probabilities of taking up the experimental treatment conditional on confounders, C, for those who were randomised to the experimental arm; and estimated probabilities of taking up the experimental treatment conditional on C for those who were randomised to the control arm.

These probabilities are estimated using a logistic regression with the following link equation:

$$\text{logit}(\Pr[A = 1 \mid C]) = \beta_0 + \beta_1 C$$

I further stabilised the individual-specific inverse probability weights by using the estimated probability of receiving experimental treatment $\Pr[A = 1]$ in the experimental treatment arm, and the estimated probability of receiving control treatment $\Pr[A = 0]$ in the control treatment arm.[91] Stabilising the weights addresses the issue of potentially very large weights when the propensity scores (denominators) are very close to 0 (i.e. those very unlikely to be treated).

For the experimental treatment arm the weights $W^{A=1}$ are defined,

$$W^{A=1} = \frac{\Pr[A = 1]}{\Pr[A = 1 \mid C]}$$

For the control treatment arm the weights $W^{A=0}$ are defined,

$$W^{A=0} = \frac{1 - \Pr[A = 1]}{1 - \Pr[A = 1 \mid C]}$$

Lastly, I used instrumental variable estimation in scenarios where non-adhering participants receive the treatment of the opposite arm. This approach analyses all participants by quantifying first, the degree to which allocated treatment predicts actual treatment; and second, the degree to which treatment predicts outcome.[92] I

adopted the structural mean model, first proposed by Robins *et al.* for estimation of the received treatment effect on a binary outcome in randomised trials.[93] The main assumptions in using instrumental variable estimation are that: i) the instrument Z is associated with the actual treatment received, A; ii) Z does not affect the outcome, Y, except through its potential effect on A; and iii) Z and Y do not share causes.[94] Out of these conditions, only the first is verifiable. In the context of a randomised controlled trial, randomisation is an appropriate instrument. When done correctly, randomisation satisfies the first and third conditions as it randomly allocates treatment to the participants, independent of the final outcomes. The second condition is satisfied in a successfully double blinded study. When the non-adherence pattern involves switching of treatment to an alternative other than the experimental or control treatment, preference-based analyses using a framework involving 'compliers', 'preferers' and 'insisters' allows for comparison of treatment effects if two active treatments are available.[95] However, this involves additional assumptions on the treatment effects in the various arms of participants which are often not verifiable.

Instrumental variable estimation is typically used to deal with continuous outcomes in a linear model. Because the simulated trial has a dichotomous outcome, I used structural mean models. These have no explicit constraints other than the mean model for the causal effect of interest.[94] I used the semiparametric two-step generalised method of moments from the R package `gmm`. Further details can be found in the package description available at: https://cran.r-project.org/web/packages/gmm/gmm.pdf.

Non-inferiority hypothesis testing

The null hypothesis is tested by comparing the upper bound of the two-sided 95% confidence interval of the effect estimate with the non-inferiority margin. Non-inferiority is concluded if the upper bound of the 95% confidence interval for the absolute risk difference between the experimental and control treatments is less than the non-inferiority margin.

I simulated 18 different patterns of non-adherence. The conditions of these non-adherent patterns are shown in Figure 2.3.

Comparing the analysis methods

To examine type 1 error, i.e. concluding non-inferiority when the experimental treatment is actually inferior, I assumed a difference of 0.1 in the probability of treatment failure between the control and experimental arms (i.e. the experimental treatment is inferior and the true treatment effect is 0.1 on an absolute scale). In the case of non-adherent participants receiving an alternative inferior treatment compared to the experimental and control arms, I assumed the difference in the probability of treatment failure between the control and alternative treatment, and experimental and alternative treatment to be 0.1 and 0.2 respectively. Since the

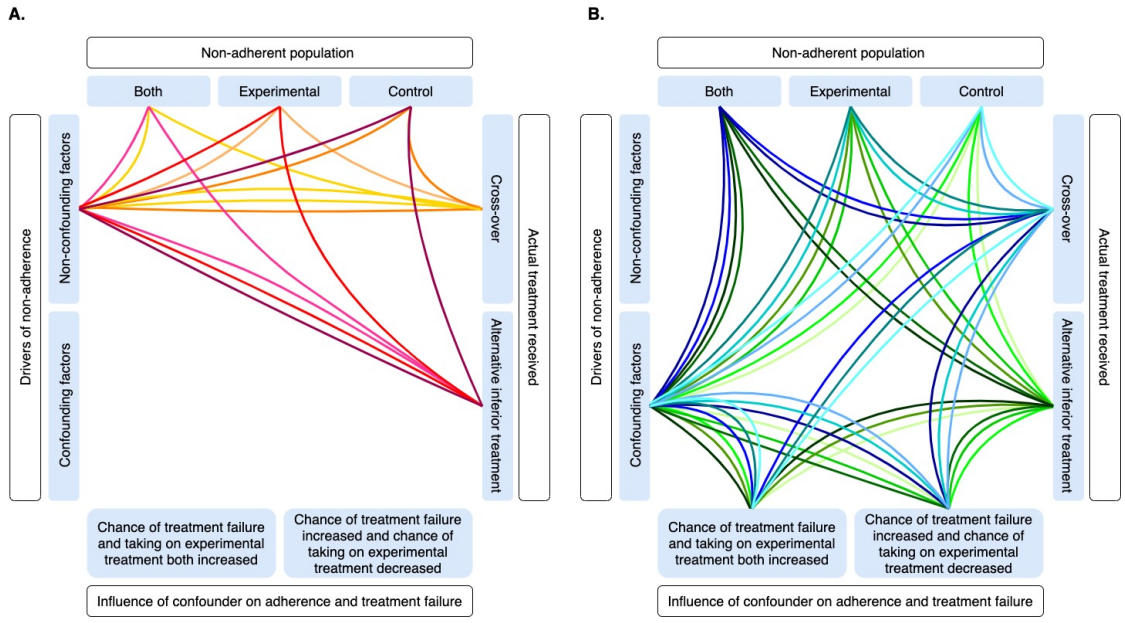


Figure 2.3: Simulation scenarios. Simulation scenarios explored permutations of four factors: i) non-adherent population (both arms, experimental arm, control arm); ii) actual treatment received by the non-adherent population (crossing over to the opposite arm, another treatment inferior to both the experimental and control treatment); iii) reason for non-adherence (due to confounding factors or non-confounding factors); iv) if non-adherence is due to non-confounding factors, direction of influence of the confounders on the probability of taking up the experimental treatment and outcome (both probabilities may increase or decrease, or the two probabilities are in opposite directions). The left panel shows the six possible scenarios when non-adherence is due to non-confounding factors. The right panel shows the 12 possible scenarios when non-adherence is due to confounding factors. Each coloured line represents one scenario.

non-inferiority margin is assumed to be 10%, simulation iterations which concluded non-inferiority were considered to have committed type 1 error (Figure 2.4).

Power, given by one minus the type 2 error, is the proportion of non-inferiority trials which conclude non-inferiority correctly. Here, I assumed the true treatment effect to be zero. Thus, the experimental treatment arm has the same probability of having treatment failure, i.e. non-inferior to the control treatment. Simulation iterations which concluded inferiority were considered to have committed a type 2 error (Figure 2.4).

The above assumptions on treatment effects used in calculating type 1 error and power for the scenarios below are arbitrary and intended for illustrative purposes. Other assumptions can be explored with the Shiny app (https://moru.shinyapps.io/samplesize_nonadherence/).

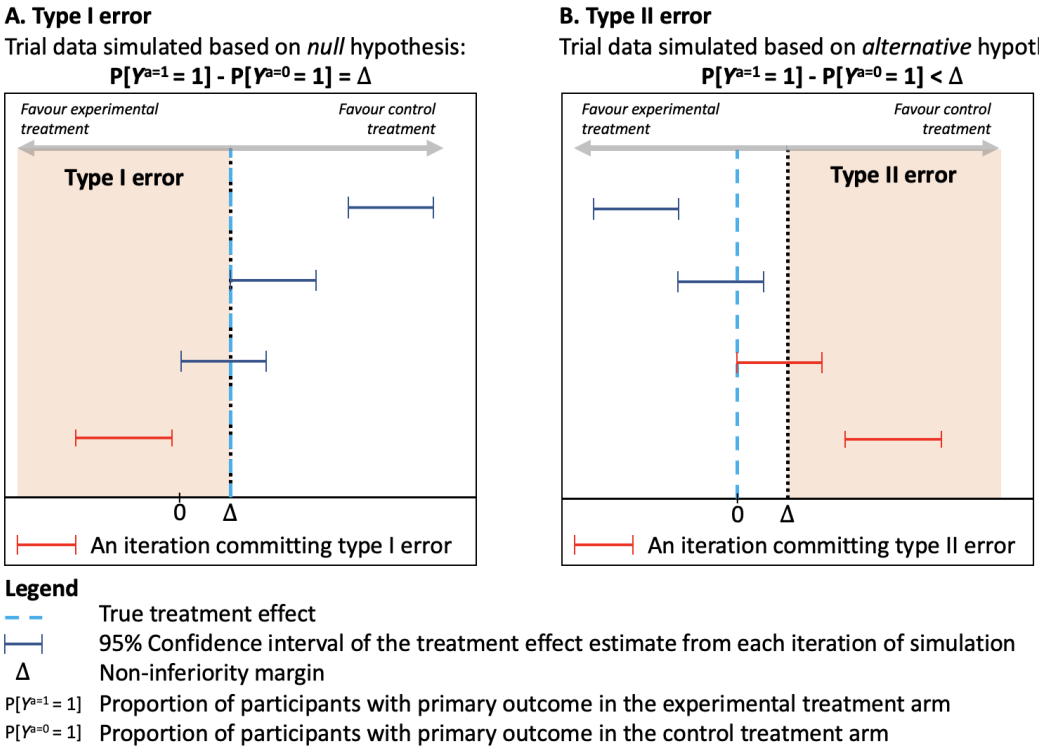


Figure 2.4: Derivation of type 1 and 2 errors in the simulation. Non-inferiority margins are represented by black dotted lines. True treatment effects are highlighted in blue broken lines. Trial estimates falling to the left of 0 favour the experimental treatment, while trial estimates falling to the right favour the control treatment. Panel A: The type 1 error rate is estimated by simulating trial data based on the null hypothesis with the true trial effect equivalent to the non-inferiority margin (10%). The upper bound of the treatment effect estimate's 95% confidence interval is compared with the non-inferiority margin. Iterations with the upper 95% confidence interval boundary less than the non-inferiority margin (marked in red) are considered to have committed type 1 error. I report the proportion of iterations that commit type 1 error in the results. Panel B: The type 2 error rate is estimated by simulating trial data based on the alternative hypothesis with the true treatment effect less than the non-inferiority margin (10%). Iterations where the upper 95% confidence interval boundary is greater than the non-inferiority margin (marked in red) are considered to have committed a type 2 error. I report 1 minus proportion of iterations that commit type 2 error as power in the results.

2.3.2 Results

Non-adherent participants receive treatment from the opposite arm

Non-adherence due to non-confounding factors

In most patterns of non-adherence, ITT estimates tend to shift towards a zero difference between the control and experimental arms. The only exceptions are when study participants allocated to the experimental arm actually received no treatment or a treatment inferior to both treatments offered in the trial. Compared to treatment efficacy estimates, ITT analysis has a higher tendency to claim non-inferiority when the experimental treatment is actually inferior and when there is

non-adherence. Figure 2.5 illustrates the case where non-adherent study participants cross over to the opposite arm. Even at a relatively high adherence of 90%, the type 1 error of the ITT estimate can be as high as 10%. All other analysis methods are unbiased in this case where non-adherence is due only to non-confounding factors.

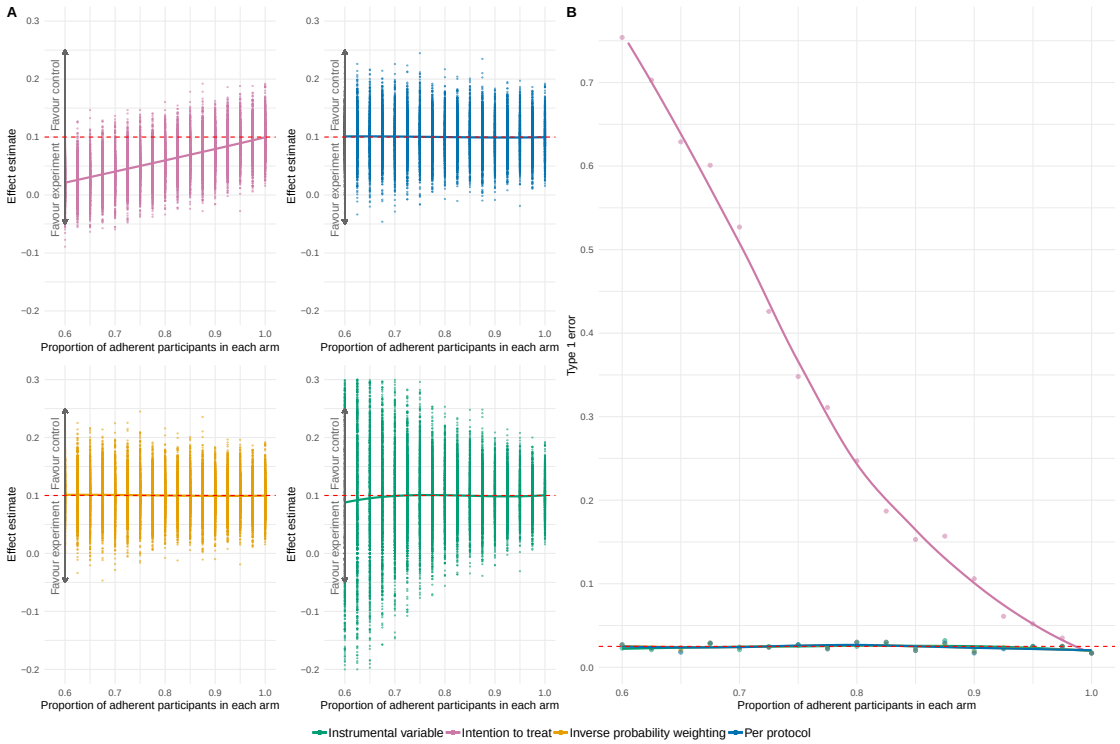


Figure 2.5: Non-adherence caused by non-confounding factors. A: Dots represent trial estimates calculated from each iteration. Coloured lines are the locally estimated scatterplot smoothing (LOESS) lines through mean trial estimates from all iterations. Because the outcome in the simulated trial refers to treatment failure, higher effect estimate values favour the control treatment. The red dotted line is the true effect size estimate assumed in the simulations. B: Dots represent type 1 error calculated from all iterations at various degrees of adherence. The tolerable type I error is set at 0.025 at full adherence.

Non-adherence due to confounders and no unobserved confounding
In the case where confounders influence non-adherence behavior, PP analysis is biased in estimating the causal effect of treatment. Figure 2.6 illustrates an example where increasing confounder value decreases the probability of taking up the experimental treatment (with a corresponding increase in the probability of taking up the control treatment) and increases the probability of treatment failure. Participants with the highest confounder values in the experimental arm cross over to the control arm and participants with the lowest confounder values in the control arm cross over to the experimental arm. This will lead to an inflated type 1 error rate. In this case, inverse probability weighting and instrumental variable estimation give unbiased estimates with conservative type I error rates.

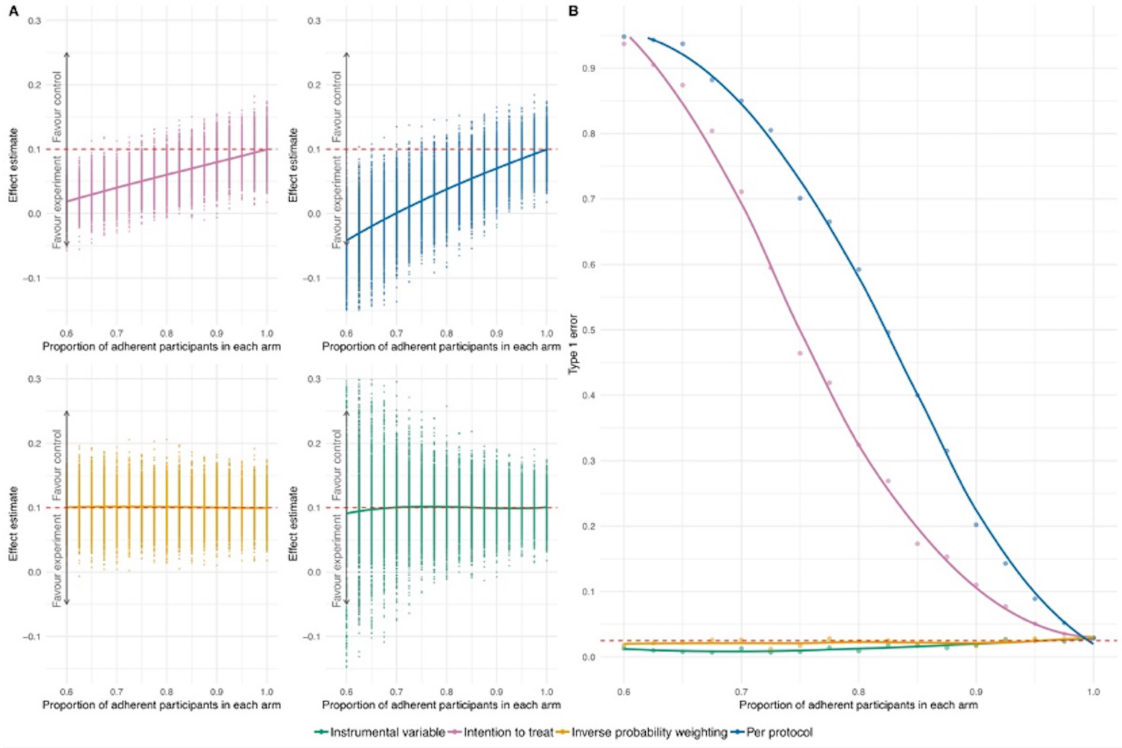


Figure 2.6: Non-adherence caused by confounding factors I. Non-adherence caused by confounding factors where participants with higher confounder values have lower probability of taking up the allocated treatment regardless of the allocation, and the probability of treatment failure is increased.

The more influence the confounder has on treatment failure, the more biased PP estimates will be, leading to higher type 1 error rates (Figure 2.7). When the confounder increases both the probability of taking up the experimental treatment and of treatment failure, the treatment effect estimated with the PP analysis will be higher than the true value (Figure 2.8).

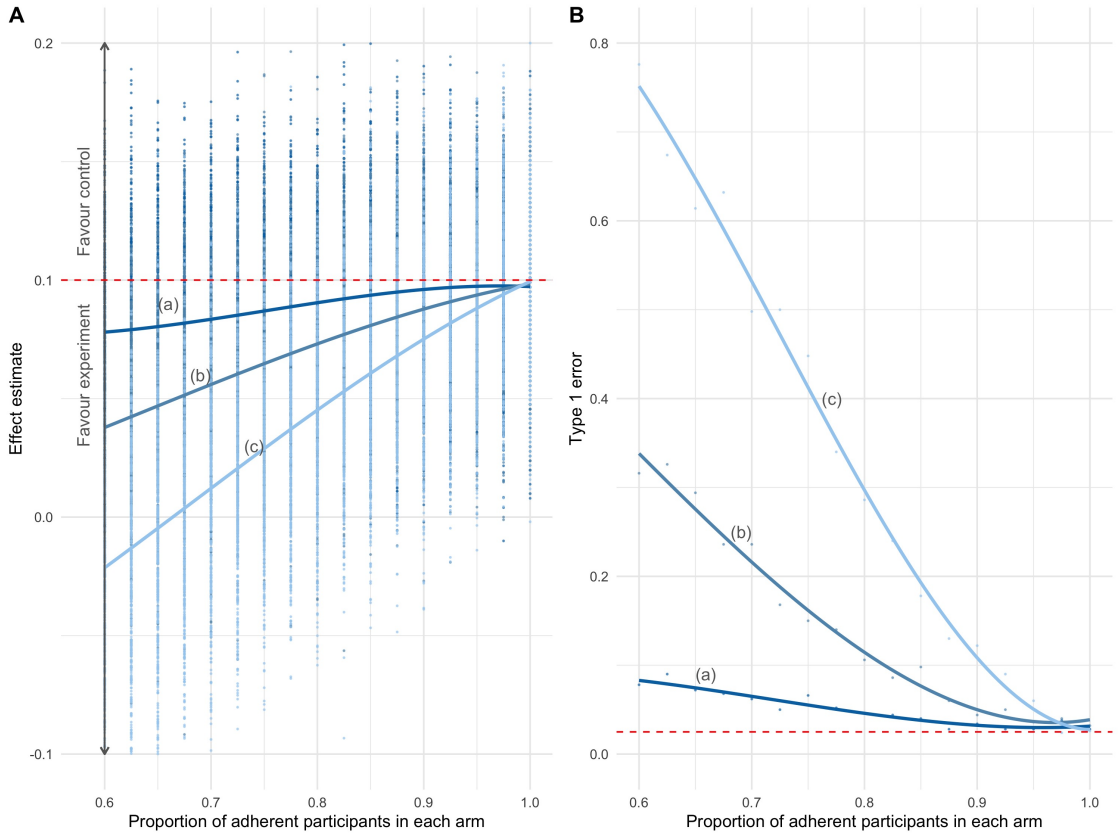


Figure 2.7: Non-adherence caused by confounding factors II. Non-adherence caused by confounding factors where participants with higher confounder values have lower probability of taking up the experimental treatment regardless of allocation, and the probability of treatment failure is increased. Per-protocol analysis is shown (in various shades of blue) to illustrate the impact of increasing direct confounder effect on treatment failure, in terms of the treatment estimates and associated type 1 errors. The magnitude of direct confounder effect on treatment failure is calculated with treatment failure as the dependent variable and confounder as the independent variable in a regression. (a) magnitude of direct confounder effect on treatment failure = 1; (b) magnitude of direct confounder effect on treatment failure = 5; (c) magnitude of direct confounder effect on treatment failure = 9.

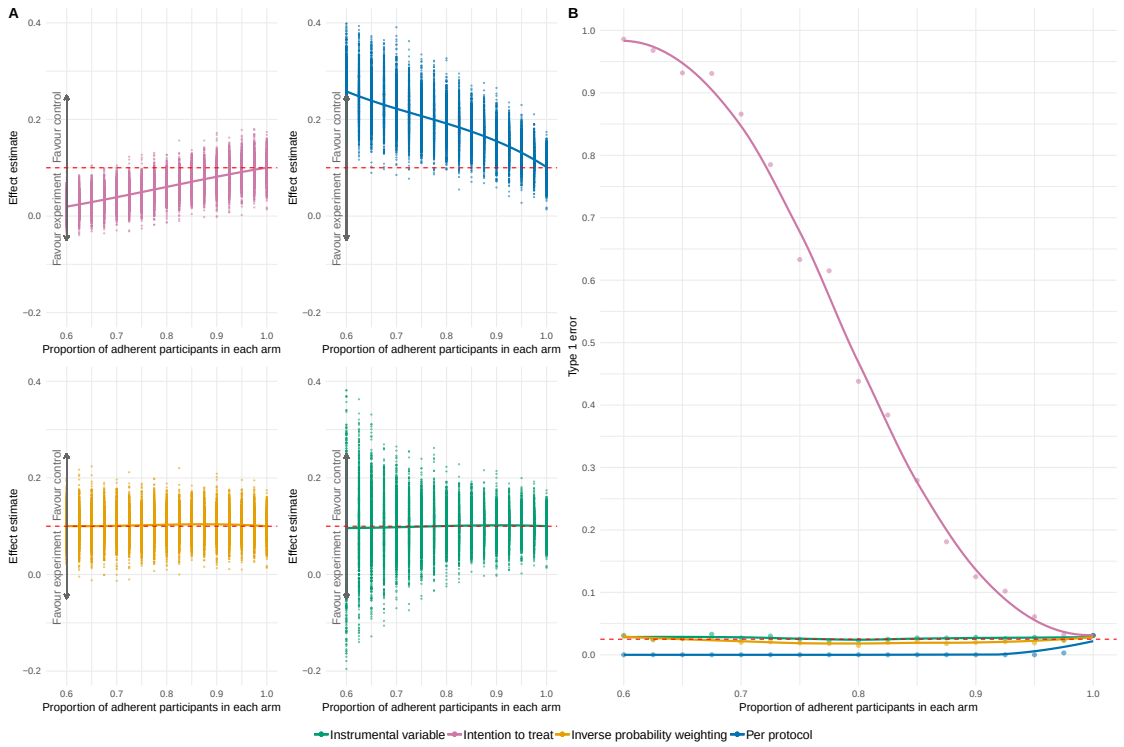


Figure 2.8: Non-adherence caused by confounding factors III. Non-adherence caused by confounding factors where participants with higher confounder values have higher probabilities of taking up the experimental treatment regardless of the allocation, and treatment failure.

Non-adherence due to confounders with unobserved confounding

In practice, not all confounders will be observed, and those which can be observed may not be measured perfectly, so that it will only be possible to partially adjust for confounding. In such cases, inverse probability weighting can become biased (Figure 2.9). Adjusting for more confounders can reduce but not eliminate bias in treatment estimates. Instrumental variable estimation, on the other hand, remains unbiased even with unobserved confounders, as it does not depend on knowledge of the confounders to compute treatment effect estimates when all the above-mentioned assumptions are met.

Non-adherent participants receive alternative treatment that is inferior to both the experimental and control treatments

If non-adherent participants do not cross over to the opposite arm, they may receive an alternative inferior treatment or default care. The effect of this on the ITT treatment estimates depends on the allocation arm that is predominantly non-adherent. When most of the non-adherent participants are from the control arm, the control treatment will appear worse compared to the experimental treatment using the ITT analysis, thereby favouring the experimental treatment (Appendix A.1 Figures A.6, A.12, A.18). However, when most of the non-adherent participants are from the experimental arm, the experimental treatment will appear worse compared to the control treatment using the ITT analysis, thereby favouring the

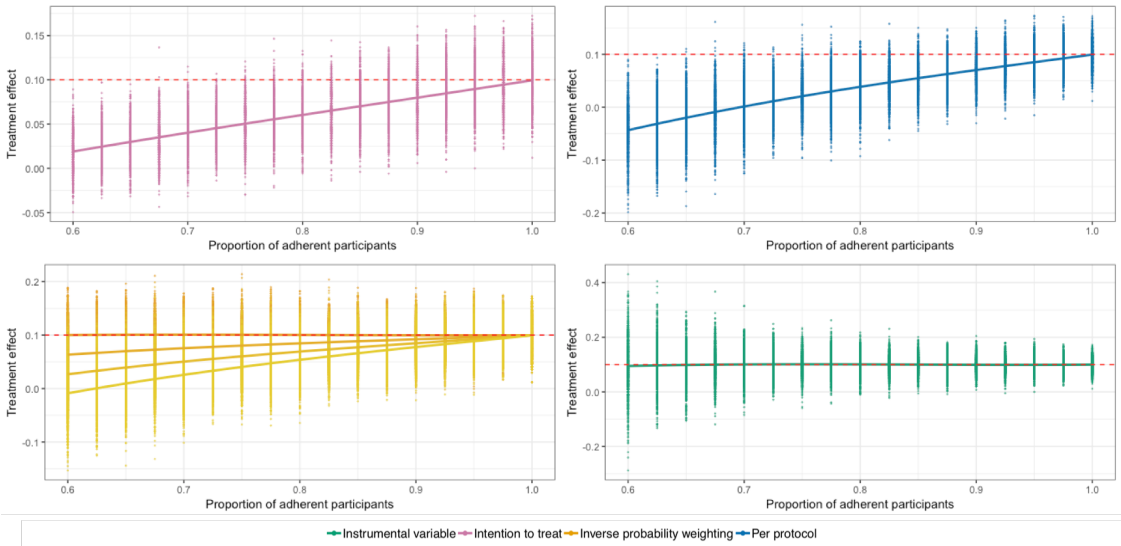


Figure 2.9: Non-adherence caused by both known and unknown confounders. Four confounders were added in the simulation. ITT, PP and instrumental variable analyses did not adjust for any confounders. For the inverse probability weighting analysis, the four lines represent situations when one, two, three, and all confounders were adjusted for. With more confounders accounted for, treatment estimates become less biased.

control treatment (Appendix A.1 Figures A.4, A.10, A.16).

Where non-adherence is caused by confounding factors, PP estimates become biased. The direction of bias is determined by the difference in the underlying prognostic characteristics of the non-adherent participants, similar to the cases where non-adherent participants cross over to the opposite arms.

Graphs illustrating effect estimates and associated type 1 errors for all simulated scenarios are included in Appendix section A.1.

Effect of non-adherence on power

In addition to affecting treatment estimates, non-adherence decreases the power to detect truly non-inferior experimental treatments. I consider the effect of non-adherence on inverse probability weighting and instrumental variable effect estimates, as these methods can potentially give unbiased treatment efficacy estimates despite non-adherence.

To maintain power, the sample size required for instrumental variable estimation increases drastically (approximately quadratically) when adherence falls below 90%. In contrast, required sample size for inverse probability weighting increases only linearly with the decrease in adherence (Figure 2.10). In the presence of non-adherence, the more influence the confounder has on treatment failure, the lower the power.

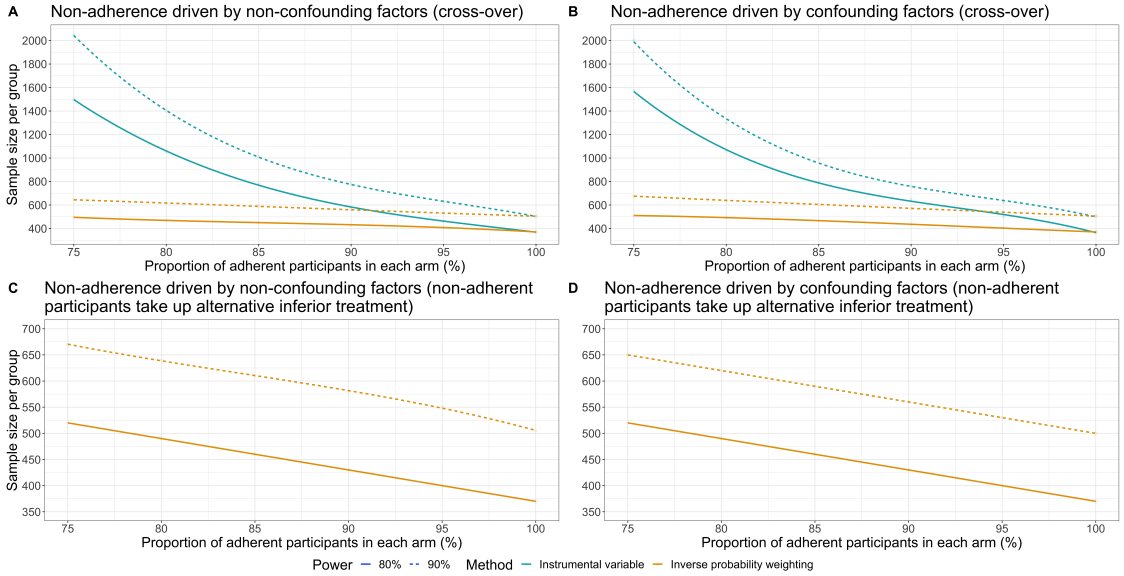


Figure 2.10: Decrease in adherence requires inflated sample sizes to maintain power. Panels A and C show the output from the scenario where non-adherence is driven by non-confounding factors. Panels B and D show the output from simulating the scenario where non-adherence is driven by confounding factors. Panels A and B show the impact of cross-over type of non-adherence on power. The assumed true proportion of treatment failure in the experimental treatment was set to be the same as the control treatment at 40%. Panels C and D show the impact on power when non-adherent participants receive an alternative treatment inferior to the control and experimental treatments by 10%. In all four scenarios, the non-inferiority margin was set at 10%.

Different patterns of non-adherence and choice of analysis methods affect power to differing degrees. To aid investigators in planning for clinical trials anticipating non-adherence, I created an online power calculator based on the simulation mechanisms shown here (https://moru.shinyapps.io/samplesize_nonadherence/). Using the same simulation mechanism as above, the calculator caters for a two-arm non-inferiority trial with binary outcomes and time-fixed treatment. The application is an interactive platform that calculates power using user inputs for the following: whether non-adherence is mainly caused by non-confounding and confounding factors; number of participants who are anticipated to be non-adherent; the expected influence the confounder is likely to have on treatment failure; and the various directions of influence the confounders have on adherence and probability of outcomes.

2.3.3 Discussion

These simulations illustrate the complexities in interpreting non-inferiority trials with non-adherence, taking both qualitative and quantitative perspectives. ITT effect estimates, due to 'dilution' from participants who received treatments different from those allocated, tend to be lower than true treatment effects at low adherence

under most non-adherence patterns. As non-adherence increases, the chance that ITT analysis will conclude non-inferiority increases. The probability of concluding non-inferiority when the treatment is actually inferior can be increased to as high as 0.1 from the acceptable 0.05 when non-adherence is 90%. The direction of bias in PP analysis is dependent on whether the confounders increase or decrease the probability of taking up the allocated treatment and the probability of the outcome occurring. This bias is increased when the confounder is more influential on the outcome.

Inverse probability weighting accounts for the difference in confounders between the allocation arms to ensure that the reweighted arms are comparable. It eliminates bias if all confounders can be appropriately adjusted for, but in general this will not be possible. In addition, the PP population should be carefully defined, with considerations for both treatments actually received and how representative this selected population is of the overall randomised population. Sensitivity analysis methods are available to address unobserved confounding and covariate measurement errors.[96, 97] In contrast, instrumental variable estimation can account for unknown confounders but requires the "exclusion restriction" to be fulfilled (i.e. treatment allocation only influences the outcome through the treatment and not through any other pathways). This assumption is unverifiable and one is only likely to be confident that it holds in a double blinded study. The other drawback of using an instrumental variable is the need for large sample sizes when adherence is low, as the method relies heavily on the strength of the instrument (i.e. randomisation) in predicting the treatment. Recent methods using doubly robust procedures have been developed to boost power when using instrumental variable estimation.[98]

Though the simulation mainly illustrates the analysis of time-fixed treatments and outcomes, time-varying treatments and outcomes can be analysed with inverse probability weighting [99] and g-estimation methods.[93] These methods are also used to address missing data and censoring.[100, 101] Another limitation of the study is that non-adherence is either due to cross-over or switching to a treatment that is inferior to both the control and experimental treatments. In practice, both types of non-adherence may occur within the same trial. However, the simulations use these extreme examples to clarify the impacts of non-adherence on trial analyses and outcomes.

Some degree of non-adherence is near ubiquitous in clinical trials. Though ITT will, under some circumstances, represent the 'real-world' effectiveness of treatment allocation, the effects of treatment itself are relevant estimates generalisable to other situations with different adherence patterns. They are also likely to be of particular interest for those with agency in their adherence. When the interest is in the actual treatment effects, as I have shown, the conservative nature of ITT in a conventional superiority trial (i.e. lower probability of concluding superiority in the presence of non-adherence) is compromised under many patterns of non-adherence in a non-inferiority trial.

2.4 Addressing non-adherence in non-inferiority trials

As shown in the earlier section with ITT analysis, if just 10% of participants cross over to the opposite arm, the probability of claiming non-inferiority can be inflated to as high as 10% from the nominal value of 5% when the experimental treatment is actually inferior in efficacy. This may lead to ineffective treatments being adopted as the standard of care and lowering the bar for subsequent clinical trials, enabling successively worse treatments to be accepted into clinical practice.[102] Such a ratchet of ever-worsening care has been termed "biocreep" (Figure 2.11).

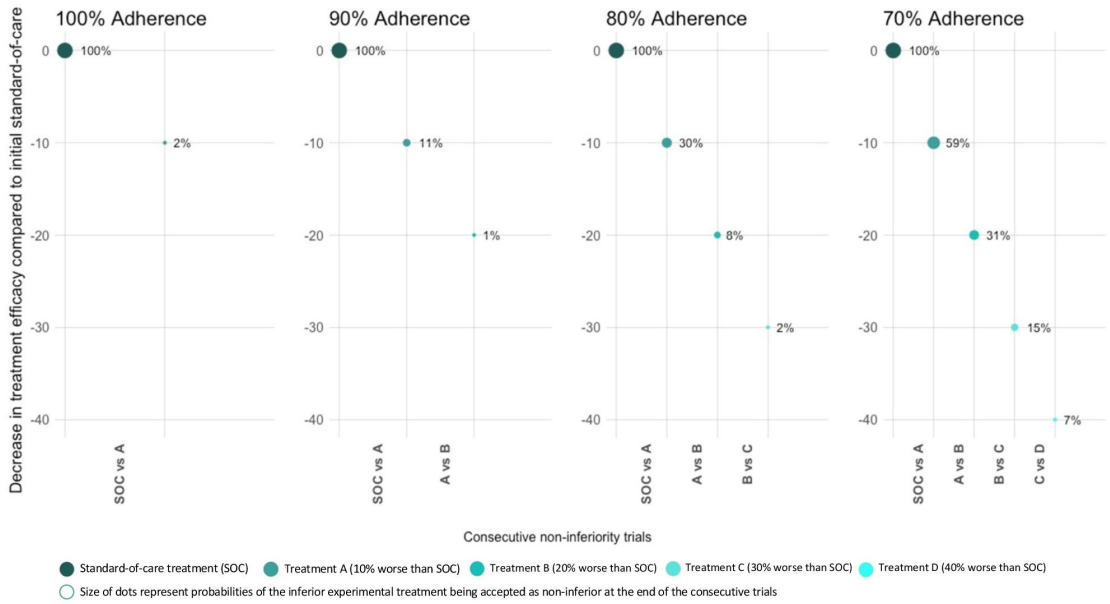


Figure 2.11: Effect of non-adherence on biocreep. The figure panels show four scenarios, if consecutive non-inferiority trials comparing standard of care (SOC) vs treatment A; treatment A vs treatment B; treatment B vs treatment C; treatment C vs treatment D were to be carried out with 100%, 90%, 80% and 70% adherence. The x-axis represents the consecutive non-inferiority trials. The y-axis represents the decrease in true treatment efficacies of A, B, C and D compared with the initial standard of care treatment. Treatments A, B, C and D are 10%, 20%, 30% and 40% less effective than the standard of care respectively. Sizes of the dots are the probabilities (specified by the percentages beside the dots) for the new and inferior experimental treatment to be accepted as non-inferior at the end of each trial. For example, if 100% adherence is maintained in the trials (first panel), the probability of treatment A being accepted as the new standard of care is 2%. This is in contrast to when the consecutive trials are conducted with 70% adherence (last panel). In this case, there is a 7% chance that treatment D will be accepted as the new standard of care, when its true efficacy is 40% less than the current standard of care. The pattern of non-adherence here is crossover, i.e. in the 70% adherence scenario, 30% of participants from each arm cross over to the opposite arm.

The Consolidated Standards of Reporting Trials (CONSORT) group, the Food and Drug Administration (FDA), and the European Medicines Agency (EMA) guidelines for conducting non-inferiority trials have uniformly emphasised the importance of quality control in the study design but do not make specific recommendations on the appropriate analysis methods to account for non-adherence.

2.4.1 Recommendations

Though there are no straightforward solutions to the above problems, there is room for improvement in the design and reporting of non-inferiority trials with non-adherence in order to minimise bias in treatment efficacy estimates. When investigators anticipate substantial non-adherence (i.e. 5% or more) in a non-inferiority trial, an adjusted PP analysis should be planned for and performed either as the primary or supplementary analysis, depending on the primary questions of interest. I summarise my recommendations below, which are complementary to the existing guidelines.

Measurement of adherence

Processes to promote adherence to allocated treatment and to minimise loss to follow-up should be included in the protocol and the final report. Processes to assess adherence should also be described. This is particularly important in non-inferiority trials because poor adherence usually leads to higher probability of declaring non-inferiority. Adherence should be assessed in an objective and transparent manner and pre-defined in the protocol, especially in trials where interventions are not binary and definitions of adherence may be arbitrary.

For example, in treatment comparisons where drug administration is not observed, efforts should be made to assess the completeness of treatment from questionnaires, pill counting, or telemedicine (e.g. smart containers which record the time of their opening). Drug level measurement (particularly if the parent compound or a metabolite is slowly eliminated) may be informative and provide a quantitative estimate of non-adherence compared to pharmacokinetic profiles derived from observed treatments.

Collection of data on confounders of non-adherence and outcome

Identify potential confounders and consider the direction in which they may affect the probability of adhering to the allocated intervention and how they may affect the primary outcome. Describe how these confounders are observed and recorded. Data on confounders should be collected for both adherent and non-adherent study participants with similar rigour. A pilot study may be helpful to observe the types of behavior driving non-adherence and feasibility of data collection.[103, 104]

Study intervention, participants and outcome measure(s)

Enrolment criteria should be specific, and exclude individuals who are unlikely to benefit from the intervention. For example, a study comparing two antibiotic treatment regimens for *Staphylococcus* bloodstream infection should exclude simple positive coagulase-negative *Staphylococcus* bacteraemias that are usually contaminants, which do not require treatment.[41] When there is only a weak causal relationship between the study intervention and outcome, the risk of falsely claiming non-inferiority increases, as the participants in both groups can be expected to improve with or without either the control or experimental intervention.

Pre-define study populations

Define all analysis populations clearly and explain the reasons when excluding randomised participants. The PP population should refer to participants who meet pre-defined adherence definitions for their allocated intervention. In trials where interventions are administered in multiple doses over time, therapeutic relevance of non-adherence needs to be considered when classifying study participants as PP or not; e.g. in a seven-day course of twice daily doxycycline for infection it may not matter if one or two doses are missed – but for a patient with metallic heart valve replacement, missing a few doses of anticoagulant may contribute to detrimental outcomes. Participants who do not adhere to the protocol in other study procedures not related to the primary outcome should not be excluded in the primary analysis population meant for the test of non-inferiority.

Power calculation

Consider the expected degree of non-adherence, the causal relationships between the confounders and the primary outcome, and the primary analysis method when performing power calculations. These factors can be incorporated into simulations to estimate the sample size required for a pre-defined power and type I error.

Reporting of results

Data on non-adherent participants should be reported with similar rigour as for the adherent participants, including their prognostic characteristics and primary outcomes. In addition, the types of interventions taken up by non-adherent participants should also be reported.

Analysis methods

Specify the primary analysis population and method in the study protocol, and justify the choice, given the potential confounders and patterns of non-adherence. Consideration should be given to the context and audience of the trial when deciding whether ITT is an appropriate primary analysis population. An adjusted PP analysis may be done as a primary analysis when treatment efficacy estimates are more pertinent, while it may be more appropriate as a supplementary analysis in other

contexts where the ITT estimate is more relevant.

Inverse probability weighting is preferred over standardisation as an adjustment method, as the latter becomes highly inefficient with an increasing number of confounders that need to be adjusted for.[105] In addition, inverse probability weighting has the advantage of being able to handle confounders measured after randomisation.[106] Confounder selection should be based on subject matter knowledge using a causal framework rather than statistical associations.[107] An example of such a framework is to select factors or proxies causing non-adherence or outcome or both, without including factors that are instrumental variables.[108] Sensitivity analysis for unmeasured confounding can be done via g-estimation methods.[109]

An instrumental variable approach is also possible if resources are sufficient for a large trial, only a crossover pattern of non-adherence is observed, and the assumptions are satisfied.[98]

2.4.2 A hypothetical worked example

In an open-label non-inferiority clinical trial, such as in the case of the REducinG Antibiotics tReatment Duration for Ventilator-Associated Pneumonia (REGARD-VAP trial described in Chapter 3), a short duration treatment (experimental arm) is compared with a long duration treatment (control arm) for ventilator-associated pneumonia. The primary outcome is death or pneumonia recurrence by 60 days. The treatment effect estimate is given by the combined outcome of mortality and pneumonia recurrence (as a proportion) in the experimental arm minus the same combined outcome in the control arm at this time point.[110]

Physician preference may affect adherence, as some are accustomed to prescribing a long duration of treatment. Physicians may also affect patient outcomes. Disease severity is another potential confounder as physicians usually prescribe short durations for mild disease and long durations for severe disease (Figure 2.2).

A sample size of 800 is required for 80% power and 2.5% one-sided type 1 error if we consider the efficacy of both short and long duration treatments to be 60% and choose a non-inferiority margin of 10%. During the trial, data on the potential confounders and the primary outcome are collected from all participants. The overall adherence was 75%, comparable to observed adherence in pragmatic trials studying duration of treatment.[111] Figure 2.12 summarises the simulated data according to the allocation and per-protocol groups of participants.

It was observed from the results that participants who actually received the experimental treatment were those with milder disease, and they tended to be treated by physician B. An adjusted PP analysis with inverse probability weighting was then used as the primary analysis, and instrumental variable estimation as the sensitivity analysis. PP analysis with inverse probability weighting showed

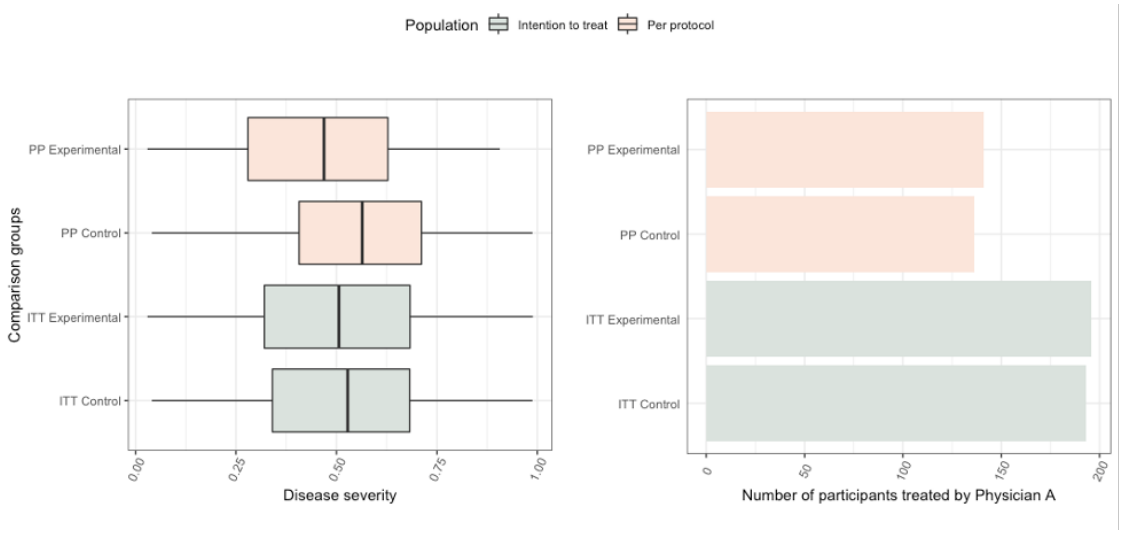
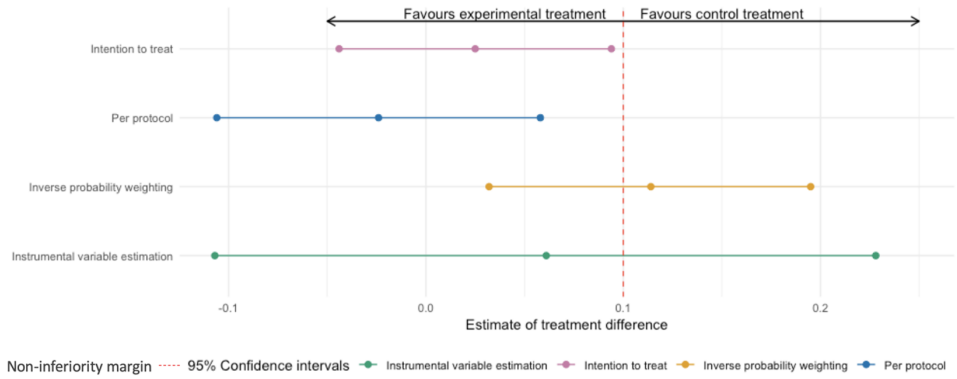


Figure 2.12: Summary of data from the hypothetical example trial. Panel A shows the distribution of disease severity, which ranges from 0 to 1. Panel B shows the number of participants in each comparison group treated by physician A. There were 2 physicians in the simulation.

that the experimental treatment is worse than the control with an estimate of 0.114 (95%CI 0.032–0.195). As expected, the instrumental variable approach gave a wide confidence interval and crossed the non-inferiority margin. Both of these methods agreed that the experimental treatment is not non-inferior to the control treatment. ITT and PP analyses, however, would have concluded non-inferiority, possibly committing type 1 error. A comparison of the analysis methods is given in Figure 2.13.



	Outcome	
	Estimate of treatment difference	**95% confidence interval**
Intention to treat	0.025	-0.044 to 0.094
Per protocol	-0.024	-0.106 to 0.058
Inverse probability weighting	0.114	0.032 to 0.195
Instrumental variable estimation	0.061	-0.107 to 0.228

Figure 2.13: Comparison of analysis methods for the hypothetical example trial. Red dashed line is the non-inferiority margin. Horizontal coloured lines show the 95% confidence intervals of the estimates calculated by the respective analysis methods. The intention-to-treat and per-protocol analyses reject the null hypothesis and agree that the experimental treatment is not inferior to the control treatment. Inverse probability weighting and instrumental variable estimation, on the other hand, would not have concluded non-inferiority.

2.5 Conclusion

The effects of allocation and treatment differ with various patterns of non-adherence and analysis methods, and hence affect the determination of non-inferiority. In accounting for non-adherence in non-inferiority trials, investigators should consider the context of the trial (i.e. if the non-adherence pattern is generalisable to other settings) and the perspective of the user to decide on the appropriate effect measure to determine non-inferiority. When the interest is in treatment efficacy, it is important to determine potential confounders during the trial design and to collect appropriate data from both adherent and non-adherent participants to adequately adjust for these factors. While it may not be possible to account for all confounders that fully explain non-adherence, the proposed measures serve as a guide to reducing bias in estimates of treatment efficacy from non-inferiority trials and improve transparency in their reporting. These conclusions and recommendations were incorporated into the REGARD-VAP study protocol and are elaborated in Chapter 3.

3

REducinG Antibiotics tReatment Duration for Ventilator-Associated Pneumonia (REGARD-VAP)

A regional multi-centre randomised controlled trial

Contents

3.1 REGARD-VAP study protocol

3.1.1 Introduction

Ventilator-associated pneumonia (VAP) is the most common hospital-acquired infection in patients admitted to the intensive care unit (ICU).[112] Estimates of all-cause mortality in patients with VAP range from 20 to 50%,[40, 113] and can be as high as 94% in low- and middle-income countries.[114] Given its high prevalence and frequent association with multi-drug resistant organisms, treatment of VAP is likely to be a key driver of antimicrobial resistance in ICUs.

We continue to rely on clinical, radiographic, and microbiological criteria with low sensitivity and specificity (~70 and ~75% respectively) to diagnose VAP.[115] Identification of causative organisms can be difficult as the upper respiratory tract is non-sterile and can contaminate specimen collection from the lower respiratory tract. Concordance between non-quantitative tracheal cultures and cultures of lung tissue from open lung biopsy has been found to be as low as 40%.[116] These factors result in over-diagnosis and over-treatment of organisms thought to be causing VAP with empirical combinations of broad-spectrum antibiotics.

For those patients who are prescribed culture-directed definitive antibiotics, duration of treatment remains controversial. There are two notable French clinical trials that have suggested that a short course of seven to eight days has comparable clinical efficacy to a long duration of 15 days.[45, 50] However, these studies could not confidently conclude that the finding can be applied to VAP caused by Gram-negative non-fermenting bacilli, due to increased recurrence in such patients (odds ratio 2.18; 95% CI 1.14 to 4.16).[117] Important potential biases exist in these studies, for example the differential time period during which recurrence was assessed and the potential for erroneous classification of persistent colonisation as recurrent infection. Moreover, unplanned subgroup analyses are known to be unreliable and the higher rate of recurrence in pneumonias caused by Gram-negative non-fermenting bacilli could simply be a chance association.[45, 50]

Furthermore, the chosen empirical duration of seven days in the above trials did not make use of individual patients' clinical responses to guide antibiotic duration. Some studies have suggested that treatment duration less than seven days would suffice for VAP. One conducted by Singh *et al.* evaluated three days of empirical ciprofloxacin monotherapy for patients who satisfy a set of clinical criteria signifying low likelihood of active VAP at day three of treatment.[117] Compared to those who received a longer duration of antibiotics, there was no difference in mortality or length of ICU stay. Another randomised study by Micek *et al.* adopted an antibiotic discontinuation policy to shorten VAP treatment.[118] Similarly, there was no difference between the short (6.0 ± 4.9 days) and long duration treatment groups in terms of mortality and VAP recurrences. The above evidence supports an individualised duration of antibiotic treatment for VAP depending on disease severity and clinical response.

Another gap in the evidence is that there is no randomised study defining antibiotic treatment duration for culture-negative VAPs. There are two observational studies that compared outcomes between those whose antibiotics were withheld on the basis of negative respiratory cultures and those whose antibiotics were continued.[119, 120] Patients whose antibiotics were discontinued did not have a higher mortality or rate of new respiratory infection compared to patients whose antibiotics were continued.

The above limits the applicability of the current recommendation for short-course antibiotics for VAP, especially in Asia where most of these infections are caused by Gram-negative non-fermenting bacilli such as *Pseudomonas aeruginosa* and *Acinetobacter baumannii*.[121] The current median number of days of antibiotic treatment remains at 12-13 days in the academic centers in Thailand.[122, 123] I present a VAP trial protocol addressing the above issues, comparing the clinical outcomes of a short-course antibiotic treatment strategy to standard-of-care duration in adult patients with VAP, with the aim of reducing unnecessary antibiotic use and antimicrobial resistance selective pressure in the ICUs.

3.1.2 Study design

The REGARD-VAP trial is a multicentre randomised controlled hierarchical noninferiority-superiority trial to assess the clinical effect of a short versus standard-of-care duration in adults with VAP. The short-course treatment strategy considers the participants' clinical response, defined by defervescence for 48 hours and stabilising blood pressure, and discontinues antibiotics within seven days of treatment.

Eligibility criteria

The US Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN) VAP diagnostic criteria on patients who have been mechanically ventilated for $\geq$ 48 hours are adopted as the study subject inclusion criteria.[124] While there are no 'gold standard' diagnostic criteria for VAP, and that clinical criteria correlate poorly with autopsy findings (previously determined to be 69% sensitivity and 75% specificity), the CDC NHSN diagnostic criteria are sensitive and practical for use in ICUs of various resource levels.[115] Only one episode of suspected VAP per participant is included.

The 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America (IDSA) and the American Thoracic Society (ATS) concluded that the additional use of infection scores and other biomarkers including procalcitonin, c-reactive protein, and soluble triggering receptor expressed on myeloid cells-1 lack sensitivity and specificity.[40] For these reasons they are not included in the inclusion criteria. Microbiological culture results are not part of the inclusion criteria so that patients who have suspected VAP but negative respiratory cultures can also be enrolled.

The specific inclusion and exclusion criteria are as follows.

Inclusion criteria:

i) Patients 18 years and older

ii) Invasive mechanical ventilation $\geq$ 48 hours

iii) At least one of the following:

1. temperature $>$ 38°C
2. white blood cell count $\geq$ 12,000 cells/mm^3 or $\leq$ 4,000 cells/mm^3
3. altered mental status with no other causes in >70 year-olds; AND

iv) Two or more chest imaging tests demonstrating at least one of the following:

1. new and progressive OR progressive and persistent infiltrate
2. new and persistent OR progressive and persistent consolidation
3. new and persistent OR progressive and persistent cavitation, AND

v) At least two of the following:

1. new onset of purulent sputum, or change in character of sputum, or increased respiratory secretions, or an increase in suctioning requirements
2. new onset or worsening tachypnea or dyspnea
3. rales or bronchial breath sounds
4. worsening gas exchange defined by oxygen desaturations (e.g. PaO2/FiO2<240), increased oxygen requirements or increased ventilation demand

Exclusion Criteria:

i) Poor likelihood of survival as defined by a Sepsis-related Organ Failure Assessment score (SOFA score) of >11 points[125]

ii) Immunocompromised patients (HIV with CD4 <200 cells/mm^3, corticosteroids $>$ 0.5 mg/kg per day for $>$ 30 days, received chemotherapy in the past 3 months, solid organ or hematopoietic cell transplant)

iii) Patients receiving antibiotic therapy for any other defined extra-pulmonary infections that warrant a duration of antibiotics longer than seven days, or complications of pneumonia such as lung abscess or empyema that warrant a duration of antibiotics longer than seven days (excluding anti-tuberculosis treatment, antifungal medications, antibiotics meant for chronic suppression of chronic infections or chronic obstructive lung disease)

iv) Patients who have been treated for VAP for more than seven days from screening

v) Vulnerable patients including prisoners and refugees

Recruitment and Participating sites

The trial is conducted in ICUs across Singapore, Thailand and Nepal. These hospitals include university academic and provincial-level centers from various resource settings to ensure generalisability of the trial findings. Enrolment to the study commenced in June 2018. As of July 2021, there are 36 participating ICUs from seven hospitals. These hospitals are:

1. National University Hospital, Singapore
2. Tan Tock Seng Hospital, Singapore
3. Sunpasitthiprasong Hospital, Ubon Ratchathani, Thailand
4. Srinagarind Hospital, Khon Kaen, Thailand
5. Khon Kaen Hospital, Khon Kaen, Thailand
6. Patan Hospital, Patan Academy of Health Science, Nepal
7. Civil Service Hospital of Nepal, Nepal

Dedicated research teams in each study site screen all admitted patients who have been intubated for more than 48 hours for eligibility. Potential participants are then recruited by the local investigators. Written informed consent is obtained from every participant, or the participant's legal representative or next of kin if the participant is sedated and does not have decision-making capacity. When the participant is deemed to have decision-making capacity, he/she is re-consented.

Treatment protocols and intervention

Antibiotic treatment for VAP is tailored to the susceptibilities of the respiratory pathogens in accordance with the 2016 IDSA/ATS VAP guideline.[40] The number of days of antibiotics are calculated from the first day of appropriate coverage according to the susceptibility of at least one of the pathogens recovered from respiratory cultures taken within 48h of screening or VAP symptom onset. Primary physicians are encouraged by the research team to convert the initial empirical regimen to narrow-spectrum therapy based on culture results during the screening and enrolment process. In culture-negative cases, the empirical antibiotic choice is made depending on local hospital antibiograms reported by the respective microbiology laboratories (Figure 3.1).

Following enrolment, participants are reviewed daily according to the fitness criteria for stopping antibiotics (Figure 3.2). These criteria include:

(a) body temperature $\leq$ 38.3 °C (core body temperature measured orally or rectally) or 38.0 °C (axillary) for 48 hours, and

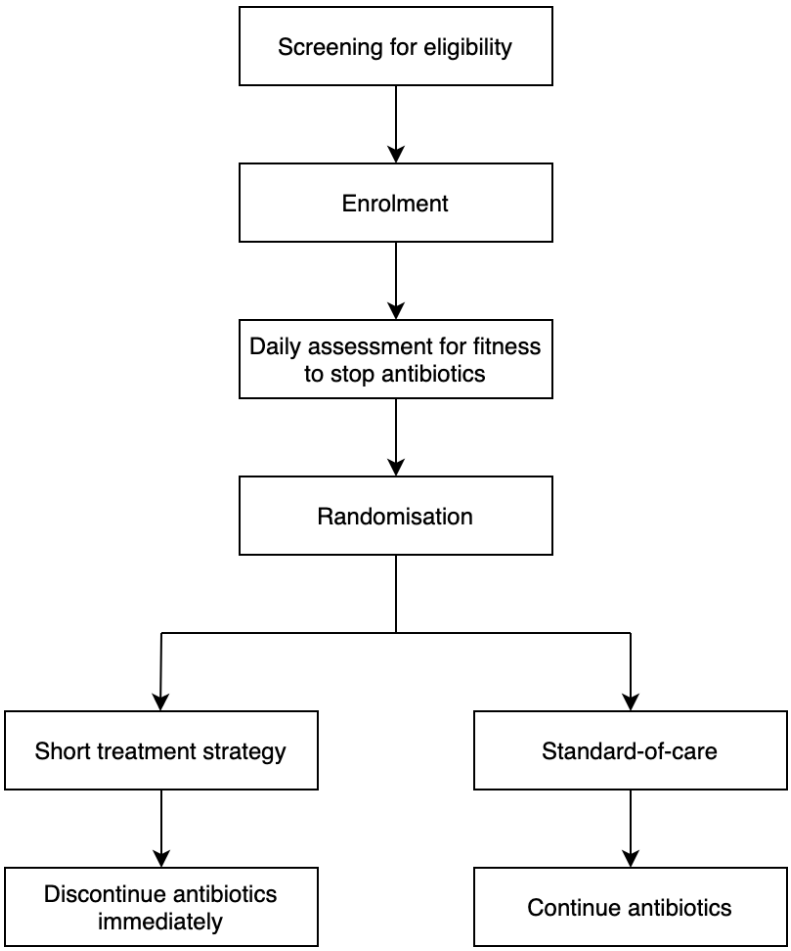


Figure 3.1: Study treatment protocol flow diagram. Main study interventions are illustrated in the flow diagram. Enrolled patients are assessed daily for clinical response to stop antibiotics. When the fitness criteria are met, all antibiotics for participants randomised to the short-duration treatment strategy arm are stopped as early as day three if the respiratory culture is negative, or day five if the respiratory culture is positive. Participants in the standard-of-care arm receive antibiotic treatment for at least eight days, with the exact duration decided by the primary physician.

(b) hemodynamic stability (systolic blood pressure $\geq$ 90 mm Hg without inotropic support or no requirement of inotropic support to maintain systolic blood pressure above 90 mm Hg).

When the above criteria are met, all antibiotics for participants randomised to the short-duration treatment strategy arm are stopped as early as day three if the respiratory culture is negative, or day five if the respiratory culture is positive. Antibiotics administered via all routes i.e. intravenous, oral and nebulisation, should be stopped within seven days. Participants in the standard-of-care arm receive antibiotic treatment for at least eight days with the exact duration decided by the primary physician.

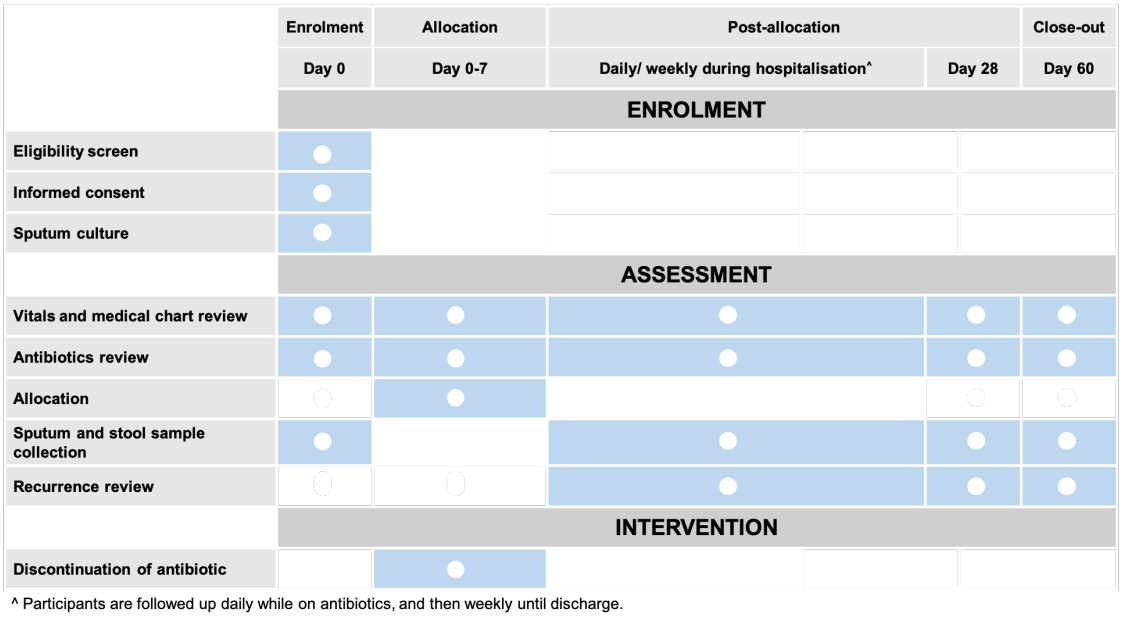


Figure 3.2: SPIRIT diagram of the schedule of enrolment, interventions and assessments.

Randomisation and blinding

Randomisation is done in a 1:1 ratio, via permuted blocks and stratified by study sites. The randomisation sequence is generated with a computer program using a seed to allow reproducibility. Allocation is performed using sequentially numbered opaque envelopes. Fitness criteria to discontinue antibiotics must be met prior to randomisation.

Patients would be blinded, as they are not informed of the treatment duration and likely to be sedated and unaware of the treatment regimens. To minimise observer bias by the primary physicians and study investigators, randomisation takes place after study participants have met the fitness criteria, such that study participants do not receive differential treatments during the episode of VAP except for antibiotic duration. After randomisation, investigators would contact the primary physicians to stop antibiotics for those participants randomised to the short treatment arm. Independent experts, who are assigned to determine pneumonia recurrences, would be blinded to the randomisation arms.

Outcomes and measures

The primary outcome is the composite endpoint of death and pneumonia recurrence within 60 days of enrolment. Recurrent pneumonia is defined as an additional episode of pneumonia as determined by two independent infectious disease or respiratory medicine experts blinded to the randomisation. Day 60 is chosen for the primary outcome in preference to day 30 to reduce bias that may occur with participants in the short-treatment strategy arm having more antibiotic-free days,

thereby leading to differential detection of recurrences between the arms. Previous observational studies suggest that mortality attributable to VAP persists to day 60.[126]

The secondary outcomes are ventilator-associated events, duration of mechanical ventilation, duration of hospitalisation, duration of exposure to antibiotics during hospitalisation, acquisition of multi-drug resistant infection or colonisation during hospitalisation, and the number and types of extra-pulmonary infections identified from sterile sites during hospitalisation.

Detailed definitions of the above outcomes are provided in the Appendix B.2 (Table B.1). For all these outcomes, the absolute risk difference between the proportion of participants with the outcomes in the standard-of-care arm and the short treatment strategy arm will be calculated.

3.1.3 Trial procedures

Baseline procedures

Upon enrolment, a set of respiratory cultures is performed if this has not been done for the index episode of VAP. This may be collected either via the endotracheal tube or bronchoalveolar lavage as ordered by the primary physicians. Brocheoalveolar lavage is not mandated as this has been shown not to improve mortality and clinical outcomes and may be difficult to perform in low-resource settings.[127, 128] Microbiology cultures are processed and reported in the local laboratories, which are expected to have standing quality control measures. Susceptibility studies are reported using European Committee on Antimicrobial Susceptibility Testing (EUCAST) or Clinical and Laboratory Standards Institute (CLSI) agar method and breakpoints.[129, 130]

Relevant clinical and laboratory related information including demographics, medical history, antibiotics administration records, chest X-ray or other imaging findings, biochemical, microbiological and haematological results and clinical parameters are collected using the case record form (CRF).

Follow-up

Participants are followed up daily while on antibiotics, and subsequently weekly when remaining hospitalised. Following discharge, two further follow-ups are scheduled at day 28 and 60 (Figure 3.2). During follow-up visits, patients are interviewed to identify possible episodes of pneumonia recurrences.

Laboratory studies

Respiratory and stool samples are collected from the study participants weekly during hospitalisation, then at day 28 and 60. These samples will undergo DNA extraction and shotgun metagenomics analysis. The characteristics of the microbiota

will be compared between the groups of patients to explore the short- and long-term impact of the various regimes and durations of antibiotics (Section 5.2.4).

3.1.4 Statistical analysis plan

The primary and secondary outcomes of the study populations will be analysed using unadjusted and adjusted methods in both the per-protocol and intention-to-treat populations. The per-protocol population includes all study participants who fulfill the eligibility criteria, undergo randomisation, meet fitness criteria for antibiotic discontinuation, and who receive seven days or less of appropriate antibiotics in if the short treatment strategy arm, or eight days or more if they are in the standard-of-care arm. The intention-to-treat population includes all study participants who have been randomised during the conduct of the study. Adjustment will be done with inverse probability weighting, using baseline patient characteristics (study site, age, gender, comorbidities, residence prior to admission, type of ICU admitted to, SOFA score, VAP infection with carbapenem-resistant organisms, maximum heart rate and minimum mean arterial blood pressure on randomisation day, duration of intubation prior to developing VAP, reason for intubation, number of days from first respiratory symptom onset to first day of appropriate antibiotics) as independent variables.[90] These are potential confounders that are not on the direct causal pathway between duration of antibiotic treatment and the outcomes.[87]

Subgroup analysis will be performed using the above primary outcome amongst patients with VAP caused by Gram-negative non-fermenting and carbapenem-resistant bacilli. These groups are of interest because VAPs caused by Gram-negative non-fermenting bacilli have previously been shown to be associated with increased recurrence;[45, 50] and VAPs caused by carbapenem-resistant bacilli have no standardised treatment. We will compare the primary outcome in these subgroups in both the per-protocol and intention-to-treat populations, and if the sample sizes allow, the same set of baseline patient characteristics stated above will be used in an inverse probability weighting model. We also will adopt the same non-inferiority hypothesis as the main analysis in these subgroups.

This is a non-inferiority trial with a hierarchical noninferiority–superiority hypothesis. The first analysis to be conducted will be for determination of non-inferiority. Only if non-inferiority is established by this primary analysis will a second analysis for superiority be conducted using closed testing methods without requiring adjustment of the significance level for multiple comparisons.[54] The trial estimates for the primary outcome will be calculated with the absolute risk difference (proportion of participants with the primary outcome in the short arm minus that in the long arm). Hence, non-inferiority will be concluded if the upper boundaries of the one-sided 95% confidence intervals from both unadjusted and adjusted analyses are below the non-inferiority margin. The purpose of using both adjusted analyses on the intention-to-treat and per-protocol populations to determine non-inferiority is to minimise the inflation of type 1 error associated with non-adherence in non-

inferiority trials.[51] Superiority will be declared if the 95% confidence intervals for all the trial estimates are entirely above zero.

Sample size calculation and non-inferiority margin determination

Mortality after sustaining an episode of VAP has been reported to be 14–43% globally.[50, 131–138] VAP recurrence rates range from 14 to 40%, with higher incidence in those caused by Gram-negative non-fermenting bacilli.[45, 133, 139] Mortality observed in these recurrence episodes were 17–50%.[131, 133] Based on these, we expect the primary outcome (a composite binary outcome of mortality and VAP recurrence) to occur in 55% of the patients in the standard of care arm. We derived an absolute non-inferiority margin of 12% with the fixed-margin method, preserving at least 50% of the efficacy of standard treatment in VAP.[140, 141] Using a group sequential design adopting the boundaries proposed by Fleming, Harrington and O'Brien, a maximum of 412 patients will be required to achieve a power of 80% to conclude non-inferiority between the two groups with a one-sided α risk of 5%.[142] As we anticipate a loss to follow-up and/or non-adherence of up to 10%, we plan to enrol a maximum of 460 patients.[143]

Data collection and management

Paper CRFs are completed at the study sites and entered onto an electronic database, MACRO Electronic Data Capture.[144] A dedicated data manager and study monitor at the Mahidol-Oxford Research Unit Clinical Trial Support Team supervise the overall quality of the data collection. The data manager reviews data entered on a monthly basis and any unexpected values are clarified with the site study teams. The final dataset will be archived in the data repository after the publication.

The study monitor and a project coordinator conduct regular site visits for quality control. All study sites are assessed prior to initiation of the study for capacity to conduct the randomised controlled trial, during the study and upon completion to ensure data quality. Monitoring reports are made available to the study sites and investigators after each visit.

3.1.5 Safety monitoring plan

Serious adverse events, including all mortality and pneumonia recurrences, are reported to the local ethics committees and the study sponsor according to the respective requirements and timeline. The data safety and monitoring committee (DSMC), consisting of an infectious disease physician, a respiratory/intensive care physician and a statistician, is responsible for the monitoring and evaluation of the clinical data generated, with a focus on safety in an independent and objective manner. The DSMC reviews all serious adverse events on a monthly basis, as well as interim analysis reports, to make recommendations on study conduct such as continuation, modification, suspension and termination. A trial steering committee has been constituted and will decide on the continuation of the trial and report to

Interim analysis	Z- value upper bound	α-error spending	Power
1	3.47	0.00026	0.014
2	2.45	0.00697	0.241
3	2.00	0.01795	0.329
4	1.73	0.02482	0.216
Total		**0.05**	**0.80**

Table 3.1: α-error spending and power at each interim analysis for determination of non-inferiority.

the central ethics committee.

Four interim analyses will be performed on the primary endpoint each time a further 25% of the estimated sample size has been randomised and the patients have completed 60 days of follow-up. The group sequential design adopting the boundaries proposed by Fleming, Harrington and O'Brien will be used to terminate the trial prematurely if the Z value derived exceeds the defined boundaries for non-inferiority (Table 3.1).[142, 145, 146]

3.1.6 Trial implementation

Strategies to ensure adherence and assessment of adherence to protocol

Non-adherence, especially in non-inferiority trials, is challenging to account for in the analysis and complicates interpretation of results.[54, 143, 147] To maintain engagement with the local investigators and healthcare providers, the study team carries out regular meetings with the stakeholders to elicit feedback on study procedures. Prior to enrolment and randomisation, the study team contacts the primary physicians to confirm their adherence to allocated interventions. Post-randomisation, close follow-ups are done to ensure antibiotics are stopped or continued according to allocation.

Adherence is assessed by duration of culture-directed antibiotics. A participant is considered to meet the definition of per-protocol if he/she fulfilled the eligibility criteria for enrollment, fitness criteria for discontinuation of antibiotics, and received seven or fewer days of appropriate antibiotics in the short treatment strategy arm, or eight or more days of appropriate antibiotics in the standard-of-care arm.

Qualitative component and implementation of intervention

The study team sought opinions from the local investigators and ICU healthcare providers during the trial design stage and subsequently while conducting the study to gather feedback on acceptance of the study intervention and to anticipate practical and operational issues. An important debate was on the indicators of fitness to stop criteria. There were suggestions to include inflammatory markers such as procalcitonin as criteria to discontinue antibiotics. This approach has

previously been shown to reduce antibiotic use in patients without increase in adverse events.[128, 148] However, as such tests are not routinely available in Thailand and Nepal, and performing these tests off-site would be likely to cause prolonged delay in discontinuation of antibiotics, a decision was made not to include these as part of our intervention.

Early discontinuation of antibiotics was also deemed to be a challenge to implement for some intensive care physicians due to variations in practices. Champions for the trial were identified locally, who advocated for the trial intervention prior to initiation. External pilots were conducted when required, for the sites to improve adherence in the main trial.[149]

A qualitative study is being carried out concurrently at the study sites to understand the interactions of the ICU staff with the trial procedures. This study aims to improve adherence to the study protocol and find strategies to sustain the trial intervention as part of local antibiotic stewardship programs if the trial demonstrates that the short treatment strategy is non-inferior.

3.1.7 Institutional review board approvals

The overall sponsor of the study is the University of Oxford. Approval by the Oxford Tropical Research Ethics Committee (OxTREC) was obtained prior to applications to the respective local ethics committees. This trial protocol was developed in accordance with the SPIRIT 2013 Statement and CONSORT statement extension for 'Non-inferiority and Equivalence Trials'.[51, 150] The full protocol is provided in Appendix B.1.

The REGARD-VAP trial has been approved by the following ethics committees: Oxford Tropical Research Ethics Committee (40-17), Srinagarind Hospital Center for Ethics in Human Research (4.6.03: 59/2563), Khon Kaen Hospital Institute Review Board in Human Research (KEF64003), Sunpasitthiprasong Hospital Center for Ethics in Human Research (085/2561), Singapore National Health Group Domain Specific Review Board (2018/00250) and Nepal Health Research Council (630-2018).

3.2 Pilot study

A pilot study was carried out from 19 February 2018 to 23 May 2018 at the Sunpasitthiprasong Hospital, Thailand. Sunpasitthiprasong Hospital is a large provincial hospital with over 1300 beds and 300 ventilators, and reports a high prevalence of VAPs caused by multi-drug resistant Gram-negative bacteria. It is expected to enrol the greatest number of patients for the REGARD-VAP trial. As such, local investigators were engaged early during the study development phase and their feedback on the acceptability of trial design and intervention, specifically the antibiotic stopping criteria, was taken into consideration. In spite of initial resistance, all medical and surgical physicians unanimously supported that the trial should proceed by an internal vote.

The pilot study included 25 patients with VAP. Out of these, 18 met the fitness criteria and were randomised. Eight participants were randomised to the short arm and 10 to the long arm. Only two out of eight study participants in the short arm received the allocated intervention, with an adherence of 25% (Figure 3.3).

3.2.1 Assessing feasibility and acceptability of the trial design, intervention and procedures

The pilot study highlighted numerous challenges and potential sources of protocol deviations. These issues and eventual solutions adopted by the study teams are discussed below. Though some of the more specific solutions described mainly pertained to the Sunpasitthiprasong Hospital, variations of these steps were implemented in the other sites and the broad principles were adopted in either the study protocol or manual standardised across the study sites.

Screening

Because REGARD-VAP included culture-negative VAP episodes, screening criteria were based solely on the ICU patients' signs and symptoms. Hence, screening had to be performed at the bedside for all intubated patients beyond 48 hours of ventilation. This was an onerous task as the signs and symptoms had to be retrieved from multiple document sources including ventilator settings, medicine administration records, vital signs records and medical notes. Frequently, these variables were missing or inconsistently documented by the treating teams.

In addition, the research nurses reported that the various dates of symptom onset, respiratory cultures, and empirical versus definitive antibiotics initiation were confusing and could lead to screening errors. For example, a patient develops fever and has increased sputum production (day 0). This triggers ceftazidime/avibactam and amikacin to be started empirically. Subsequently, the patient fulfills all other inclusion criteria on day 2. The endotracheal respiratory culture reports third-generation cephalosporin resistant *Klebsiella* spp. on day 3 and the patient's antibiotic regimen is streamlined to meropenem. The patient is then screened

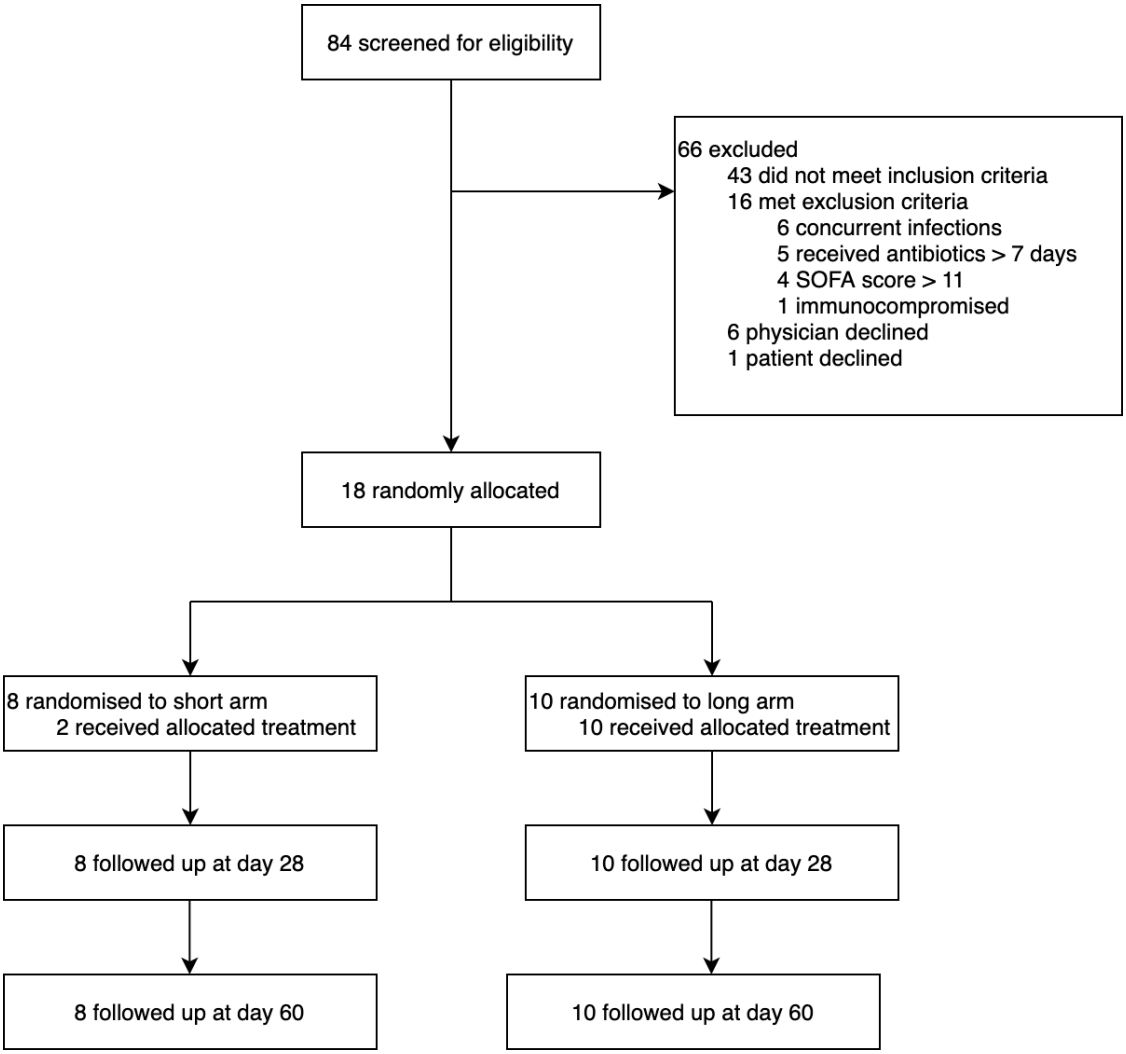


Figure 3.3: CONSORT diagram for the pilot study.

and enrolled on day 5. In this example, the initial date for culture-appropriate antibiotic should be day 0 but could be erroneously taken as day 3 or 5 due to the time lag between symptom onset, culture report, initiation of empirical and definitive antibiotics, screening and enrolment. This would affect when the patient was randomised and the total treatment duration.

To address this issue, I developed a standardised screening sheet. This sheet specifies the source documents where screening criteria should be sought, and automatically generates the time windows for signs and symptoms screening, sputum cultures, antibiotic initiation and treatment duration. Initiation dates for culture-appropriate antibiotics were also standardised. The first respiratory symptom onset date or culture-appropriate antibiotic initiation date, whichever is later, is taken as day zero for the purpose of counting treatment duration. To confirm the eligibility criteria for potential enrolments, I developed a centralised platform for the research teams to confirm all enrolments with either the local investigators

or myself in a timely manner.

Enrolment

Most of the potential study participants were sedated and unable to provide consent. The research staff had to contact family members, who sometimes lived in a different province, for face-to-face consent. This was challenging because the time window for randomisation is narrow. Also, many family members were understandably anxious as their loved ones were in critical condition and suffering from a hospital-acquired infection. This was often aggravated by hearing that the study intervention early cessation of antibiotics, a medicine widely perceived to be life-saving.

To deal with these issues, the research team was trained in consent taking using standardised scripts. On top of the standard text in accordance with the Good Clinical Practice guideline, these scripts focused on providing reassurance to the families that antibiotics would only be stopped when there was response to treatment, and could be restarted as required. In addition, consent taking was planned in advance especially during the weekends and public holidays. Communications were improved within the team, and with the co-investigators and treating physicians.

Randomisation and intervention

Randomisation errors were due to new infections that arose between enrolment and randomisation; e.g. multi-drug resistant catheter-related bloodstream infections, which required antibiotic treatment for more than seven days and precluded the patient from randomisation. Also, errors occurred due to miscalculations in the treatment days and fitness criteria. Non-adherence to allocated intervention was due to physicians who preferred longer treatment durations and declined to stop antibiotics when the allocation was short-arm.

We conducted individual meetings with the physicians to re-explain the protocol and sought their feedback on the trial intervention. The most common reasons for reluctance to stop antibiotics were perceived poor infection prevention and control practices putting ICU patients at high risk of contracting hospital-acquired infections including VAP, distrust in the quality of locally-produced antibiotic, fear of missing untreated infections in a vulnerable patient population, and resistance to change from an established practice passed on from peers and seniors. We used these opportunities to engage with these physicians, offer reassurance and highlight the potential repercussions of excessive antibiotic use.

To ensure adherence to allocated intervention, an additional step was put in place that required the research staff to discuss the case with the treating physicians immediately prior to randomisation. This served as a reminder for the physician on the study protocol, and for the research team to pick up new issues which may prevent the patient from receiving either short or long antibiotic durations.

The high non-adherence amongst the physicians presented an opportunity for a more in-depth investigation into the socio-behavoral aspects of antibiotic prescribing. A qualitative component was designed and conducted in parallel with the REGARD-VAP trial, which aimed to understand i) healthcare workers' interactions with the study intervention, ii) perception and attitude towards antibiotic stewardship in the ICU, and iii) to identify and design stewardship interventions (Section 5.2.1).

Follow-ups

Detecting pneumonia recurrences was demanding during follow-up. I found that searching medical charts for documented VAP episodes or signs and symptoms compatible with VAP was not adequately sensitive. This was further complicated by patient transfers between hospitals, and medical charts from other hospitals could not be easily retrieved.

Using the pilot patients as case studies, I found that the most sensitive method of detecting pneumonia recurrences was screening for signs and symptoms whenever new respiratory cultures were performed or new antibiotics were started. If the suspected recurrence episode fulfilled the US CDC criteria, two independent physicians would then be asked to review the medical chart. The suspected recurrences that fulfilled the screening criteria but were not considered to be true recurrences by the two independent physicians would be reported as 'ventilator-associated events' as a secondary outcome.

To prevent loss to follow-up when patients transfer between hospitals, research staff called the clinical team from the new hospital weekly to seek relevant information on VAP symptoms, respiratory cultures and antibiotics prescriptions.

Outcome

During determination of pneumonia recurrences, it is crucial to minimise bias that may be caused by knowledge of the randomisation arms. During the pilot phase, we attempted various methods of blinding the independent physicians. Some examples included only presenting the relevant pages of medical notes for a specific window during which recurrences were suspected, and documenting the allocation arm separately from the main set of medical notes during randomisation.

3.2.2 Facilitating determination of effect size for sample size calculation

Amongst the randomised patients, three out of eight patients in the short arm and two out of ten in the long arm died (3/8, 38% vs 2/10, 20%). None suffered recurrences. Amongst those who were not randomised, four out of seven suffered the primary outcome (4/7, 57%) which consisted of three cases of mortality and one recurrence. These were within the expected proportion of patients with primary

outcome proposed in the original protocol (55% of the patients in the standard of care arm). Hence, the sample size calculation remained unchanged after the pilot study.

3.3 Interim analysis results

From 25 May 2018 to 9 July 2021, 340 patients were enrolled and randomised (Figure 3.4). Prior to the COVID-19 pandemic, the mean enrolment rate was 12 participants per month. Since January 2020, enrolment decreased to a mean of 5 participants per month. This section includes the first two hundred thirty-one patients considered in the second interim analysis, which concluded in September 2020.

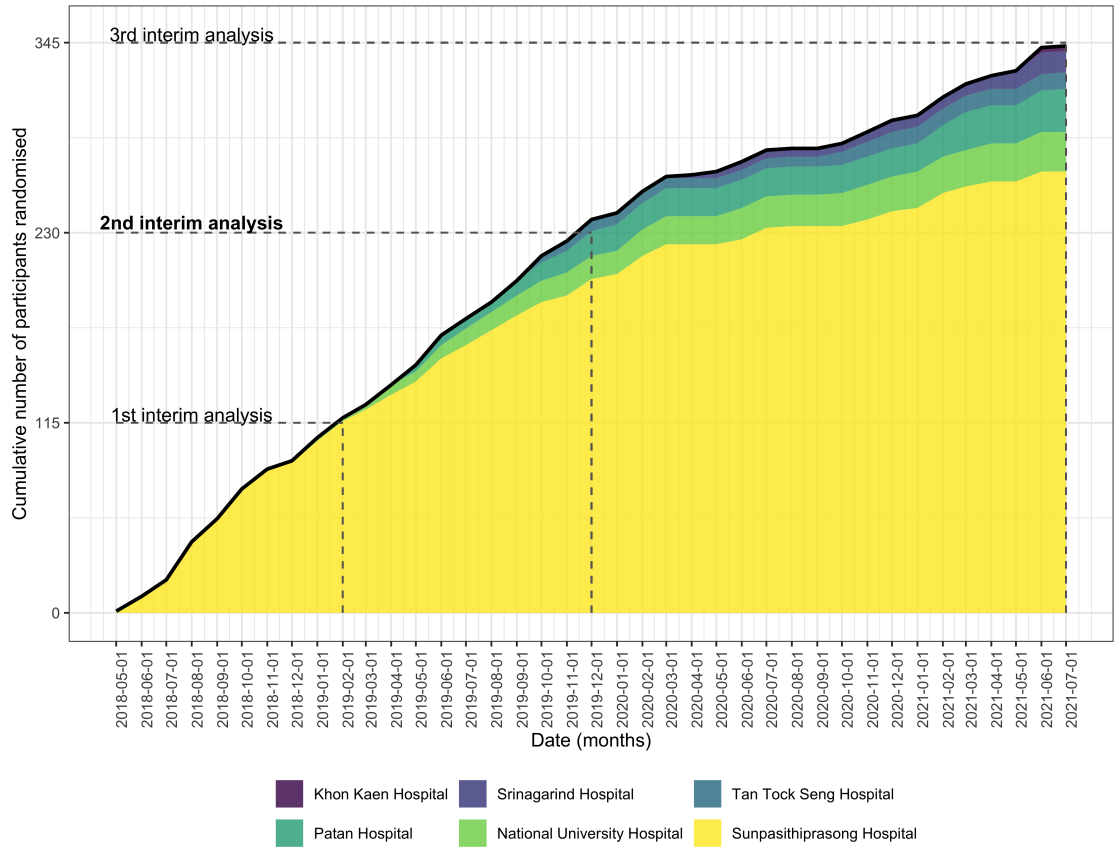


Figure 3.4: Enrolment from the REGARD-VAP participating sites.

3.3.1 Demographic and clinical characteristics of the patients

From 25 May 2018 to 27 December 2019, 4198 patients were screened and of these 231 patients were randomised. A large number of screened patients were not randomised because the trial included culture-negative VAP episodes, which required all patients intubated for $\geq$ 48 hours to be screened regardless of pneumonia symptoms. The intention-to-treat analysis consisted of 230 participants, which

excluded one participant who withdrew and requested that his data not be used in the analysis (Figure 3.5). The baseline characteristics of these patients were similar, except that the percentage of patients intubated due to heart failure was slightly higher in the long duration arm (32/115, 28% compared to 21/115, 18% in the short arm in Table 3.2).

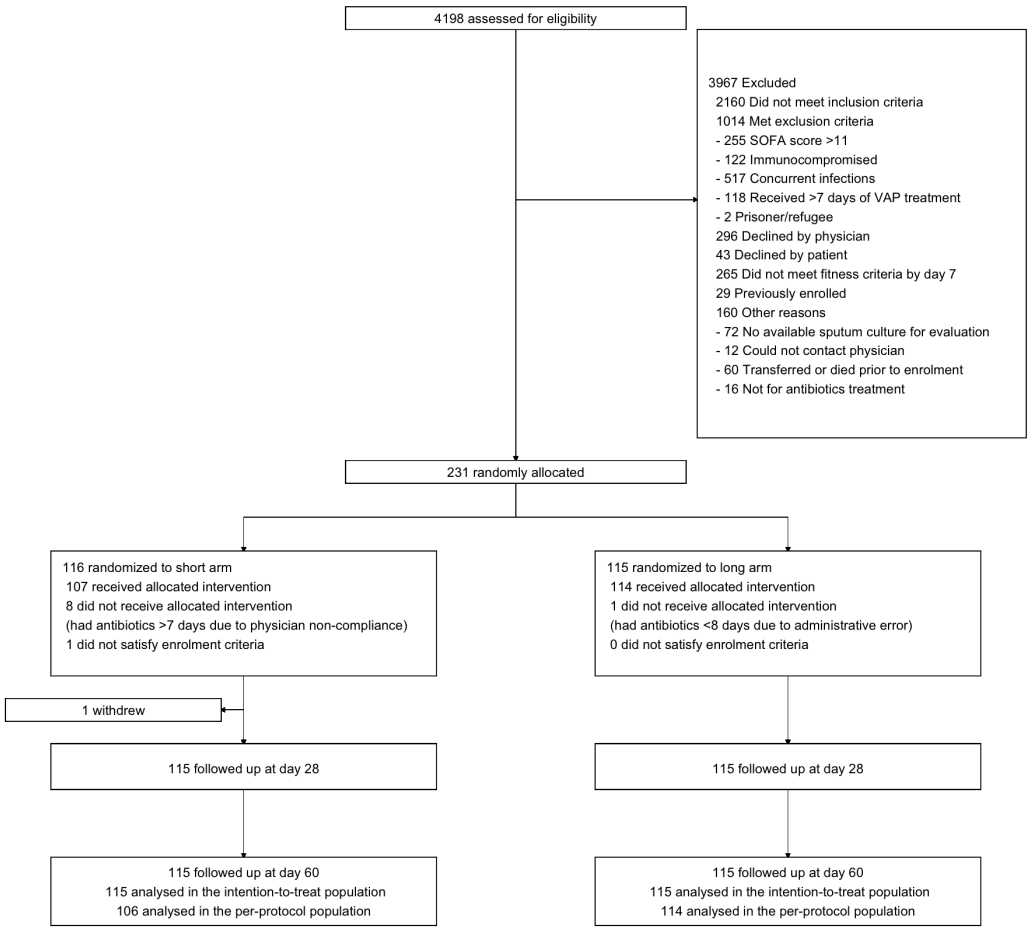


Figure 3.5: CONSORT diagram for the second interim analysis.

	Long arm (n = 115)	**Short arm** (n = 115)	**Overall** (n = 230)
Age			
Age (mean, SD)	64.9 (16.6)	60.8 (17.5)	62.9 (17.2)
Age (median, IQR)	66.0 [54.3,77.2]	63.5 [48.7,74.8]	65.1 [50.8,76.0]
Gender			
Female	46 (40%)	49 (43%)	95 (41%)
Male	69 (60%)	66 (57%)	135 (59%)
Study site			
Patan Hospital	8 (7%)	7 (6%)	15 (7%)
National University Hospital	7 (6%)	7 (6%)	14 (6%)
Tan Tock Seng Hospital	2 (2%)	4 (3%)	6 (3%)
Sunpasitthiprasong Hospital	97 (84%)	97 (84%)	194 (84%)
Srinagarind Hospital	1 (1%)	0 (0%)	1 (0%)
Residence prior to ICU admission			
Own home	29 (25%)	37 (32%)	66 (29%)
Transfer from another hospital	86 (75%)	78 (68%)	164 (71%)
Charlson comorbidity score			
Score (mean, SD)	3.2 (2.2)	3.1 (2.0)	3.2 (2.1)
Score (median, IQR)	3.0 [2.0,4.0]	3.0 [2.0,5.0]	3.0 [2.0,4.0]
Comorbidities			
Congestive heart failure	6 (5%)	17 (15%)	23 (10%)
Coronary heart disease	10 (9%)	8 (7%)	18 (8%)
COPD	9 (8%)	11 (10%)	20 (9%)
Liver cirrhosis	1 (1%)	1 (1%)	2 (1%)
Chronic kidney disease	13 (11%)	15 (13%)	28 (12%)
Cancer	4 (3%)	4 (3%)	8 (3%)
Diabetes	32 (28%)	20 (17%)	52 (23%)
SOFA score			
Score (mean, SD)	6.3 (2.5)	6.3 (2.6)	6.3 (2.5)
Score (median, IQR)	6.0 [5.0,8.0]	6.0 [4.0,8.0]	6.0 [4.0,8.0]
Ward			
Medical ICU	37 (32%)	39 (34%)	76 (33%)
Surgical ICU	78 (68%)	76 (66%)	154 (67%)
Duration of intubation prior to VAP			
Days (mean, SD)	17.7 (14.2)	22.7 (27.4)	20.2 (22.0)
Days (median, IQR)	13.0 [10.0,21.5]	14.0 [10.5,21.0]	13.0 [10.0,21.0]
Reason for intubation			
Heart failure	32 (28%)	21 (18%)	53 (23%)
Metabolic acidosis	9 (8%)	7 (6%)	16 (7%)
Neurological failure	36 (31%)	38 (33%)	74 (32%)
Respiratory failure	11 (10%)	15 (13%)	26 (11%)
Sepsis	23 (20%)	32 (28%)	55 (24%)
Trauma	4 (3%)	2 (2%)	6 (3%)
Carbapenem-resistant bacteria in sputum sample during the index episode of VAP			
Carbapenem-resistant Gram-negative bacteria isolated	18 (16%)	24 (21%)	42 (18%)
Vital signs on enrolment			
Lowest MAP prior to randomisation (mean, SD)	86.9 (11.8)	83.5 (14.3)	85.2 (13.2)
Mean maximum heart rate (mean, SD)	100.0 (18.7)	99.0 (15.9)	99.5 (17.4)
Inotropic support prior to randomisation	2 (2%)	7 (6%)	9 (4%)
SpO2/FiO2 ratio (mean, SD)	252.4 (88.4)	266.8 (93.3)	259.6 (91.0)
SpO2/FiO2 ratio (median, IQR)	247.5 [205.8,250.0]	250.0 [200.0,250.0]	247.5 [200.0,250.0]

Table 3.2: Baseline characteristics of the patients, according to treatment group.

The per-protocol population consisted of 220 participants, which excluded 10 participants who did not receive their allocated interventions. The reasons for exclusion of the 10 participants from the intention-to-treat population are detailed below. Participant characteristics for the per-protocol population are included in Appendix Table B.2).

- Short arm (9)
  - Erroneously randomised at day 8 without the participant having received appropriate antibiotics (1)
  - Miscommunications between study and physician teams (2)
  - Physician non-adherence (6)
- Long arm (1)
  - Error in counting treatment duration by the study team resulting in antibiotics stopped at day 8 (1)

Two-hundred ninety-nine microorganisms were isolated from the 230 episodes of VAP at enrolment in the intention-to-treat cohort (Table 3.3). The majority were Gram-negative bacilli, accounting for 78% of all pathogens isolated, of which the commonest were Gram-negative non-fermenting bacilli such as *Acinetobacter* spp. and *Pseudomonas* spp. (121 isolates, 40% of all isolates). Carbapenem-resistant organisms were isolated from 16% (18/115) of patients in the long treatment arm, and 21% (24/1150) patients in the short arm.

Organisms		All participants (n =230)	Long arm (n =115)	Short arm (n =115)
Gram-negative bacteria		**236 (79)**	**117 (39)**	**119 (40)**
Gram-negative non-fermenters		**121 (40)**	**58 (19)**	**63 (21)**
Acinetobacter spp.	Carbapenem-susceptible	34 (11)	21 (7)	13 (4)
	Carbapenem-resistant	33 (11)	14 (5)	19 (6)
Pseudomonas spp.	Carbapenem-susceptible	35 (12)	15 (5)	20 (7)
	Carbapenem-resistant	10 (3)	5 (2)	5 (2)
Stenotrophomonas spp.		9 (3)	3 (1)	6 (2)
Enterobacteriaceae		**110 (37)**	**55 (18)**	**55 (18)**
Klebsiella spp.	3GC-susceptible†	35 (12)	21 (7)	14 (5)
	3GC-resistant	18 (6)	6 (2)	12 (4)
	Carbapenem-resistant	16 (5)	5 (2)	11 (4)
E coli	3GC-susceptible	4 (1)	2 (1)	2 (1)
	3GC-resistant	1 (0)	1 (0)	0
Other Enterobacteriaceae	3GC-susceptible	29 (10)	18 (6)	11 (4)
	3GC-resistant	4 (1)	1 (0)	3 (1)
	Carbapenem-resistant	3 (1)	1 (0)	2 (1)
Others		**3 (1)**	**2 (1)**	**1 (0)**
Hemophillus spp.	Beta-lactam–susceptible	2 (1)	1 (0)	1 (0)
	Beta-lactam–resistant	1 (0)	1 (0)	0
Gram-positive bacteria		**12 (4)**	**7 (2)**	**3 (1)**
Staphylococcus aureus	Vancomycin-susceptible	8 (3)	5 (2)	3 (1)
Streptococcus spp.	Beta-lactam–susceptible	1 (0)	1 (0)	0
	Beta-lactam–resistant	1 (0)	1 (0)	0
Enterococcus spp.	Vancomycin-susceptible	2 (1)	2 (1)	0
No growth		**53 (18)**	**27 (9)**	**26 (9)**
Total		**299 (100)**	**151 (51)**	**148 (49)**

Table 3.3: Microorganisms isolated from patients' endotracheal-tube cultures during the episodes of ventilator-associated pneumonia. †3GC: Third generation cephalosporin

3.3.2 Antibiotic treatment

In the intention-to-treat population, the short and long duration arms had medians of six (IQR 5–7) and 15 (IQR 11–21.5) days respectively of continuous antibiotics for the treatment of VAP (Figure 3.6). Thirty-two participants (32/115, 28%) in the short arm had antibiotics restarted within five days of randomisation.

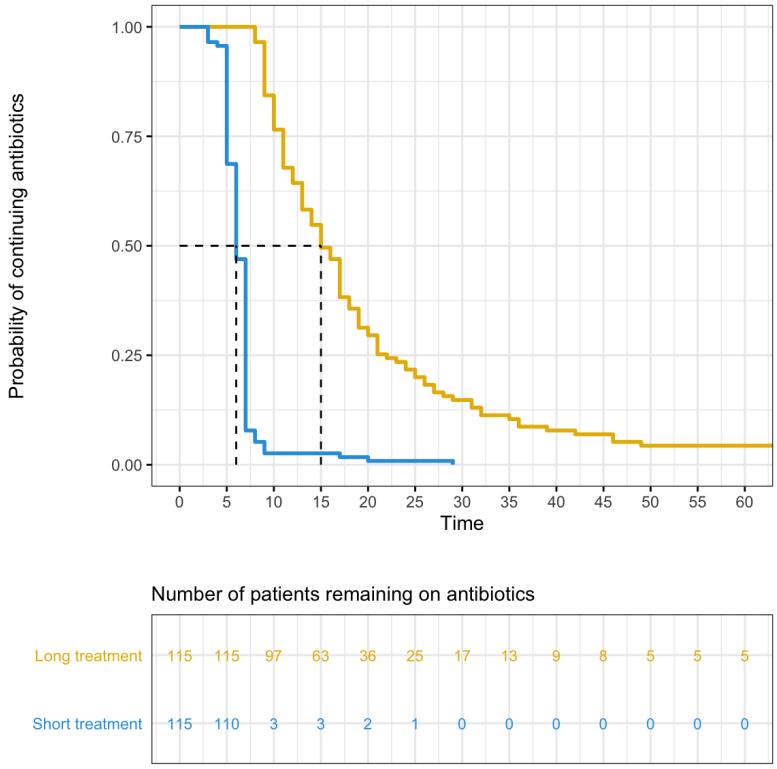


Figure 3.6: Antibiotic treatment duration in the long and short arms. Duration of antibiotic treatment presented as survival curves. Long arm is represented by yellow; short arm is represented by blue. The dotted lines indicate median treatment durations.

Table 3.4 shows the antibiotics used for the treatment of the index episodes of VAP within seven days of enrolment. One hundred and sixty-seven participants (167/230, 73%) had combination antibiotics for the treatment of VAP, with a higher proportion of patients in the long arm receiving combinations than in the short arm (94/115, 82% versus 73/115, 63% respectively). The antibiotics used in the respective arms during the follow-up period are shown in Appendix Table B.3.

Antibiotics	Frequency (%)		
	All participants (n =230)	Long arm (n =115)	Short arm (n =115)
Combinations[†]	**167 (73)**	**94 (82)**	**73 (63)**
Beta-lactams/beta-lactamase inhibitors	**128 (55)**	**67 (58)**	**61 (53)**
Piperacillin/Tazobactam	40 (17)	20 (17)	20 (17)
Amoxicillin/clavulanic acid	8 (3)	5 (4)	3 (3)
Cefoperazone/sulbactam	80 (35)	42 (37)	38 (33)
Carbapenems	**123 (54)**	**70 (61)**	**53 (46)**
Meropenem	66 (29)	34 (30)	32 (28)
Imipenem	52 (23)	35 (30)	17 (15)
Ertapenem	5 (2)	1 (1)	4 (3)
Cephalosporins	**27 (12)**	**24 (21)**	**3 (3)**
Ceftazidime	25 (11)	23 (20)	2 (2)
Cefepime	2 (1)	1 (1)	1 (1)
Aminoglycosides	**24 (10)**	**18 (16)**	**6 (5)**
Amikacin	21 (9)	15 (13)	6 (5)
Gentamicin	2 (1)	2 (2)	0 (0)
Netilmicin	1 (0)	1 (1)	0 (0)
Fluoroquinolones	**9 (4)**	**4 (4)**	**5 (4)**
Ciprofloxacin	6 (3)	2 (2)	4 (3)
Levofloxacin	3 (1)	2 (2)	1 (1)
Colistin	**55 (24)**	**36 (31)**	**19 (17)**
Tigecycline	**23 (10)**	**16 (14)**	**7 (6)**
Vancomycin	**2 (1)**	**2 (2)**	**0 (0)**
Trimethoprim/sulfamethoxazole	**7 (3)**	**4 (3)**	**3 (3)**

Table 3.4: Antibiotics used for the treatment of index episode of ventilator-associated pneumonia within the first seven days of enrolment. Antibiotic classes are highlighted in bold. [†]Antibiotics were prescribed in combination for some patients and hence not all percentages sum to 100%.

3.3.3 Study outcomes

Primary outcomes

Combined primary outcomes of mortality and pneumonia recurrence within 60 days from enrolment were compared in both the intention-to-treat and per-protocol populations. The per-protocol population was further analysed with inverse-probability weighting, using baseline patient characteristics (see section 3.1.4 for variables used in the inverse-probability weighting models). Because these covariate values contributed by the single patient enrolled from Srinagarind Hospital caused a large shift in the treatment estimates (due to the large weight of this single patient on the overall estimates), this patient was categorised as enrolled from Sunpasitthiprasong Hospital for this interim analysis.

More participants in the long arm than the short arm died or had pneumonia recurrence in the 60-day follow-up period (Table 3.5). In the intention-to-treat analysis, 45 (45/115, 39%) in the short arm and 57 (57/115, 50%) in the long arm

died or had pneumonia recurrence. In the per-protocol analysis, 41 (41/106, 39%) in the short arm and 57 (57/114, 50%) in the long arm died or had pneumonia recurrence.

The absolute risk difference estimates (proportion of participants with the primary outcome in the short arm minus that in the long arm) are given in Table 3.6, Figure 3.7 and Figure 3.8. Adjustment with inverse probability weighting shifted the upper limit of one-sided 95% confidence interval to less than zero, suggesting that short treatment duration is potentially superior to long duration in terms of the primary outcome.

		Mortality (%)	Recurrences (%)	Primary outcome (%)
ITT (n=230)	**Short (n=115)**	40 (35)	15 (13)	45 (39)
	Long (n=115)	50 (43)	16 (14)	57 (50)
	Total (n=230)	90 (39)	31 (13)	102 (44)
PP (n=220)	**Short (n=106)**	37 (35)	14 (13)	41 (39)
	Long (n=114)	50 (44)	16 (14)	57 (50)
	Total (n=220)	87 (40)	30 (14)	98 (45)

Table 3.5: Primary outcomes for the intention-to-treat and per-protocol patients. ITT: Intention-to-treat; PP: Per-protocol

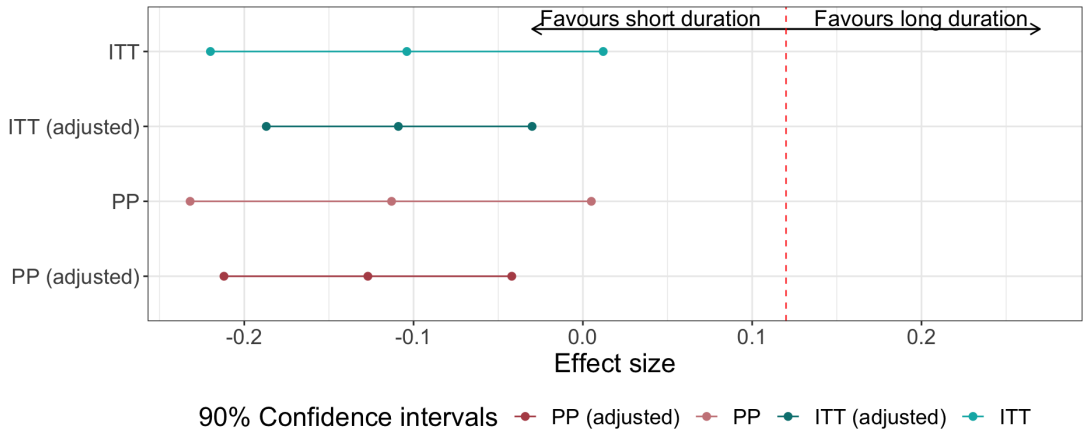


Figure 3.7: Absolute risk difference estimates. The non-inferiority margin is marked with the dashed red line at 0.12. Negative estimates indicate a higher proportion of participants suffered mortality or pneumonia recurrence in the short arm.

	Absolute risk difference	Upper limit of one-sided 95% CI
ITT	−0.104	0.012
ITT (adjusted)	−0.109	−0.030
PP	−0.113	0.005
PP (adjusted)	−0.127	−0.042

Table 3.6: Absolute risk difference estimates. The estimates are calculated as the proportion of participants with the primary outcome in the short arm minus that in the long arm. ITT: Intention-to-treat; PP: Per-protocol

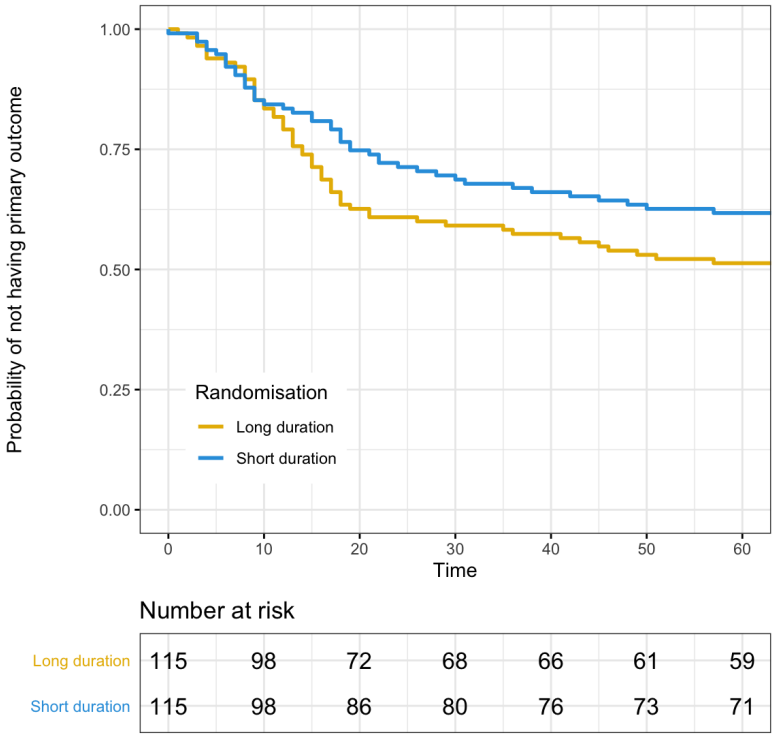


Figure 3.8: Survival curve over 60 days of follow-up in terms of primary outcome.

All mortality events were reviewed. The final diagnoses for eight participants (8/40, 20%) in the short duration arm and 12 participants (12/50, 24%) in the long duration arm were pneumonia. Amongst these patients whose deaths were attributed to pneumonia, nine had at least one multi-drug resistant organism grown in their last respiratory culture prior to demise (four versus five in the short and long arms respectively). Out of the patients who died, two patients had a positive blood culture in the 7 days prior to demise (one in each of the short and long arms). Both of these organisms were multi-drug resistant.

In terms of recurrences, 5 out of 15 patients (33%) in the short arm and nine out of 16 patients (56%) in the long arm had a multi-drug resistant organism in the

sputum culture during recurrence. Amongst all the patients who had recurrences, nine patients had the same organism(s) detected from the sputum culture as the index episode of VAP (four versus five in the short and long arms respectively).

Two subgroup analyses were pre-planned which focus on Gram-negative non-fermenting and carbapenem-resistant bacilli (Table 3.7). The interaction tests for these subgroup analyses found odds ratios of 0.95 (95% CI 0.8 to 1.14, p-value 0.61) and 0.85 (95% CI 0.64 to 1.12, p-value 0.25) for Gram-negative non-fermenting and carbapenem-resistant bacilli respectively.

			Mortality (%)	**Recurrences (%)**	**Primary outcome (%)**
Subgroup 1: Carbapenem-resistant bacilli	**ITT (n=42)**	Short (n=24)	10 (42)	3 (12)	11 (46)
		Long (n=18)	11 (61)	3 (17)	11 (61)
		Total (n=42)	21 (50)	6 (14)	22 (52)
	PP (n=37)	Short (n=19)	9 (47)	3 (16)	10 (53)
		Long (n=18)	11 (61)	3 (17)	11 (61)
		Total (n=37)	20 (54)	6 (16)	21 (57)
Subgroup 2: Gram-negative non-fermenting bacilli	**ITT (n=100)**	Short (n=51)	20 (39)	10 (20)	24 (47)
		Long (n=49)	23 (47)	8 (16)	25 (51)
		Total (n=100)	43 (43)	18 (18)	49 (49)
	PP (n=94)	Short (n=45)	18 (40)	9 (20)	21 (47)
		Long (n=49)	23 (47)	8 (16)	25 (51)
		Total (n=94)	41 (44)	17 (18)	46 (49)

Table 3.7: Primary outcomes for patients with ventilator-associated pneumonia due to Carbapenem-resistant bacilli and Gram-negative non-fermenting bacilli. ITT: Intention-to-treat; PP: Per-protocol

Secondary outcomes

Secondary outcomes were compared in the per-protocol population, and analysed with both unadjusted and adjusted methods. Patients in the short arm had fewer days of antibiotics compared to those in the long arm during 60 days of follow-up (18 ± 14 versus 26 ± 16 days respectively, $p < 0.001$, in Table 3.8). Otherwise, there were no evidence of differences between the two groups in terms of duration of mechanical ventilation, ICU stay, total hospitalisation, readmissions, ventilator-associated events, bloodstream infections and colonisation or infection with carbapenem-resistant bacteria after enrolment. Adjustment using inverse probability weighting did not substantially modify the findings with unadjusted estimates.

3.3.4 Hypothesis testing and stopping rules

The null hypothesis H0: $\theta \geq \delta$ was tested against the alternative H1: $\theta < \delta$, where θ is the treatment effect difference between the proportion of participants suffering

	Long arm	Short arm	Overall	Unadjusted estimates† (95% CI; p)	Adjusted estimates with IPW† (95% CI; p)
Total duration of antibiotics					
Mean (SD)	26.2 (15.9)	18.3 (13.9)	22.4 (15.4)	−8.0	−8.8
Median [IQR]	21.0 [14.0, 33.0]	14.0 [8.0, 24.0]	17.0 [11.0, 30.0]	(−11.9, −4.0; <0.001)	(−11.9, −5.7; <0.001)
Duration of mechanical ventilation					
Mean (SD)	46.3 (46.5)	48.9 (53.4)	47.5 (49.8)	2.5	2.6
Median [IQR]	33.5 [13.0, 71.8]	27.5 [12.0, 73.0]	31.5 [13.0, 73.0]	(−10.8, 15.9; 0.71)	(−8.9, 14.1; 0.71)
Duration of stay in ICU					
Mean (SD)	40.8 (29.7)	46.3 (52.8)	43.5 (42.4)	5.5	4.3
Median [IQR]	31.0 [18.0, 63.2]	27.0 [14.0, 64.8]	30.0 [15.0, 64.2]	(−6.0, 17.0; 0.35)	(−5.6, 14.3; 0.48)
Duration of stay in hospital					
Mean (SD)	51.1 (39.2)	62.6 (69.6)	56.6 (56.1)	11.5	8.7
Median [IQR]	37.0 [25.0, 73.0]	40.5 [21.2, 77.0]	37.5 [22.8, 74.0]	(−3.7, 26.7; 0.14)	(−3.6, 21.0; 0.24)
Readmission to an acute care hospital					
Number of readmissions (%)	21 (18%)	23 (22%)	44 (20%)	0 (−0.1, 0.1; 0.66)	0.1 (0.0, 0.1; 0.28)
Ventilator-associated events					
Number of events (%)	21 (18%)	20 (19%)	41 (19%)	0 (−0.1, 0.1; 1.00)	0 (−0.1, 0.1; 0.96)
Bloodstream infection after enrolment					
Number of bloodstream infections (%)	10 (9%)	9 (8%)	19 (9%)	0 (−0.1, 0.1; 1.00)	0 (−0.1, 0.0; 0.69)
Colonised/infected with carbapenem-resistant bacteria after enrolment					
Number of colonisation/ infection (%)	55 (48%)	44 (42%)	99 (45%)	−0.1 (−0.2, 0.1; 0.39)	−0.1 (−0.2, 0.0; 0.17)

Table 3.8: Secondary outcomes. Ventilator-associated events were identified as suspected episodes of pneumonia recurrences that fulfilled the US CDC criteria but were not confirmed by two independent assessors. †Estimates are absolute risk differences (proportion of participants with the outcome in the short arm minus that in the long arm) when proportions are reported; and differences between the means (means in participants in the short arm minus that in the long arm) when means are reported.

the primary outcome in the short and long treatment strategy groups, and δ is the fixed non-inferiority margin of 12%.

The upper bounds of the one-sided 95% confidence intervals for both the unadjusted and adjusted intention-to-treat and per-protocol analyses were less than the 12% non-inferiority margin, supporting the alternative hypothesis that short duration is non-inferior to long duration (Table 3.9). The associated Z values of these analyses exceeded the Fleming-Harrington-O'Brien boundaries (Figure 3.9).

3.3.5 Discussion

In this second interim analysis, the trend from the first analysis continued, in that more participants in the long antibiotic treatment arm died or suffered pneumonia recurrences. The final diagnoses for 8 participants (8/40, 20%) in the short duration

Analyses	Z values					α error					Power
	Boundary	ITT unadjusted	ITT adjusted	PP unadjusted	PP adjusted	Spending	ITT unadjusted	ITT adjusted	PP unadjusted	PP adjusted	
1	3.75	2.005	–	2.145	–	0.00009	0.02248	–	0.01598	–	0.006
2	**2.54**	**3.447**	**3.742**	**3.498**	**3.731**	**0.00549**	**0.00028**	**0.00009**	**0.00023**	**0.00009**	**0.219**
3	2.02					0.01805					0.348
4	1.72					0.02637					0.227
Total						**0.05**					**0.80**

Table 3.9: Upper bound boundaries proposed by Fleming, Harrington and O'Brien, α error spending and power at each interim analysis.

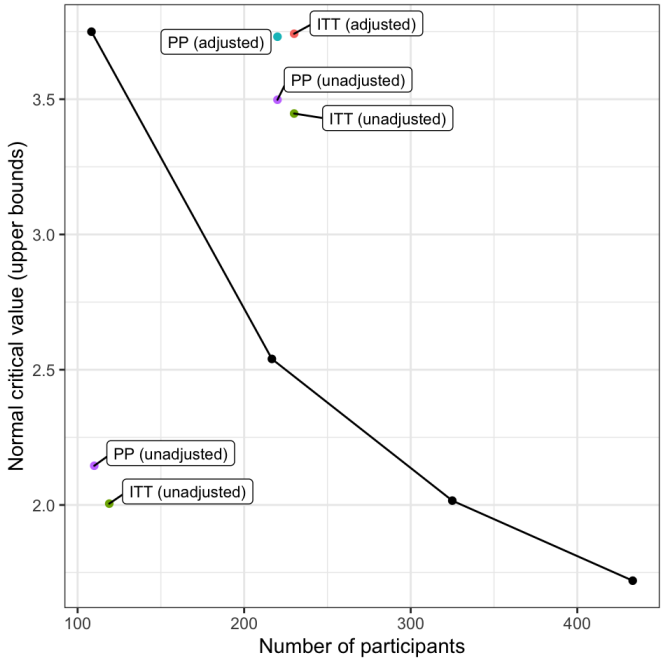


Figure 3.9: Calculated Z values. The calculated Z values for the intention-to-treat and per-protocol populations are represented by coloured dots. These are compared with the upper bound boundaries proposed by Fleming, Harrington and O'Brien represented by the black line.

arm and 12 participants (12/50, 30%) were pneumonia. This 10% difference for pneumonia-attributable mortality between the arms was similar to the all-cause mortality difference of 8% (43% in long arm and 35% in short arm). The participants' baseline characteristics were carefully checked between the two arms to ensure that the randomisation was adequate. No major differences were found in the comorbidities, disease severity (indicated by the SOFA score and vital signs on randomisation) or causative organisms.

Since Z values calculated with both the unadjusted and adjusted intention-to-treat and per-protocol patient populations exceeded the pre-defined Fleming-Harrington-O'Brien boundaries, the study was considered for termination, as the pre-specified termination conditions were met according to the initial data analysis plan.

However, given the result suggesting that short duration may actually be superior

to long duration, as well as the earlier first interim analyses which consistently showed fairly large and increasing differences between the two treatment durations in favour of the short duration, the investigators deemed that there was a potential for the trial to demonstrate superiority of the short treatment duration in addition to non-inferiority. There is a profound difference in the potential impact on clinical practice of a trial that demonstrates only non-inferiority and one that demonstrates superiority. The former suggests that short treatment duration *can* be adopted, the latter that it *should* be adopted. In this trial, the strong suggestion that short duration may actually be superior was unexpected and, because of the much greater potential for clinical impact, would itself justify continuation. In addition, unlike early termination, there were no concerns that continuing will introduce bias.

There is a reasonable probability that the finding of more patients in the long antibiotic duration arm suffering deaths and recurrences than in the short arm is due to chance. On the other hand, other trials have reported favourable clinical outcomes in those receiving short compared to long antibiotic treatment.[151–153] A biologically plausible reason for this is that patients who are exposed to longer antibiotics might become more susceptible to multi-drug resistant bacteria colonisation and infection. Previous observational data have reported associations between increased antibiotic use with infection and hospitalisation.[154, 155] *In vitro* experiments have also shown the critical role of gut microbiota on the innate immunity system.[156]

Proposed change in protocol

In view of the above, the investigators proposed to change the trial hypothesis from non-inferiority to stepwise noninferiority–superiority. The noninferiority–superiority design is well-described [54] and recommended by the Infectious Diseases Society of America to overcome the practical barriers for the conduct of superiority studies in antibiotic trials.[157] The CONSORT statement for non-inferiority trials recommended that 'if a superiority conclusion is drawn for outcome(s) for which noninferiority was hypothesized, provide justification for switching.'[149] An example of such a trial is the randomised controlled trial for linezolid in Methicillin-resistant *Staphylococcus aureus* nosocomial pneumonia [158] which was a non-inferiority trial with a nested superiority hypothesis.

In addition to the test for non-inferiority, the following was proposed to be added to the statistical methods in the study protocol:'If the upper bound of the one-sided 95% confidence interval for the absolute risk difference not only lies below the non-inferiority margin, but also below zero, then superiority can be concluded at an alpha level of 0.05'. In accordance with the latest protocol, all four types of analyses, namely intention-to-treat, adjusted intention-to-treat, per-protocol and adjusted per-protocol, must show similar conclusions, i.e. all four analyses must produce one-sided 95% confidence intervals with upper bounds that are below zero, for a conclusion of superiority to be made. There is no concern about multiplicity here because tests of non-inferiority and superiority both correspond to a simple closed test procedure.[54, 73] Adopting the closed testing procedure would also

mean that the sample size remains the same as initially proposed based on the non-inferiority hypothesis and a non-inferiority margin of 12%.

Ethical considerations

There would be no ethical concerns with continuing the trial as REGARD-VAP is a non-inferiority trial, and premature termination would not provide the evidence that short duration should be preferred over long duration (in contrast to a superiority trial). The practical result of concluding non-inferiority might be little change in prescribing behaviour. Because of the potential for showing superiority, a strong ethical argument was made for continuing; i.e. that long duration might actually be detrimental to patient outcomes, which this trial might potentially show.

Discussion with data safety and monitoring committee and trial steering committee

Both the data safety and monitoring committee (DSMC) and trial steering committee (TSC) agreed to change the trial to have a hierarchical noninferiority-superiority hypothesis. Continuation of the trial would also allow more patients to be enrolled from sites other than Sunpasitthiprasong Hospital to improve generalisability of the findings. The committee members highlighted that strategic public health clinical trials focusing on appropriate antibiotic use such as duration and choices are a top research priority. Robust evidence from clinical trials is urgently needed to inform global guidelines.

Another discussion point was on the appropriateness of the 12% non-inferiority margin. Both the FDA and major society guidelines by the Infectious Diseases Society of America, the American College of Chest Physicians, the American Thoracic Society, and the Society of Critical Care Medicine recommended a 10% non-inferiority margin in VAP and hospital-acquired pneumonia drug trials.[159, 160] Previous VAP trials on antibiotic treatment duration also considered 10% as the non-inferiority margin.[45, 50] However, in REGARD-VAP 12% was chosen because the trial adopted both mortality and pneumonia recurrence as a combined outcome. Hence, the DSMC and TSC advised keeping the same 12% non-inferiority margin. Reducing this margin to 10% would fail to address the trial's composite endpoint, since the FDA and ATS/IDSA guidelines specifically stated that the 10% margin is for mortality endpoints.

4

Effect of reducing antibiotic treatment duration on antimicrobial resistance in the hospital setting

Contents

4.1 Introduction

The primary motivation behind shortening treatment duration is the expectation that it will reduce antibiotic selective pressure for antimicrobial resistance in treated individuals, and over time will lead to lower prevalence of resistance at a population level. However, while antibiotic selective pressure as a driver of resistance is not in doubt, there are many gaps in our understanding of the relationship between specific antibiotic exposures, varying by type, frequency, dose, spatial density, and duration, and the prevalence of resistance.[60, 61] Specifically, clinical trials of reduced antibiotic treatment duration have reported conflicting effects on resistance carriage at both individual and population levels (Table 4.2). A theoretical framework explaining these differences is lacking.

Antibiotic agents promote resistance by killing or inhibiting the growth of susceptible microbes while allowing resistant ones to grow. This selective pressure for resistance most obviously applies to pathogens directly targeted by the antibiotic when used to treat or prevent infections. Examples of pathogens that may become resistant with inappropriate treatment include *Mycobacterium tuberculosis* and human immunodeficiency virus. More recently, it has been recognised that for most important bacterial pathogens, antibiotic selective pressure results largely from 'bystander' selection, which occurs when colonising bacteria are exposed to antibiotics that are not intentionally targeting them.[161] The impact of bystander exposures to antibiotics depends on the agent's spectrum of coverage, pharmacokinetic and pharmacodynamic properties, and bioavailability where the bacteria are located. The importance of bystander selection for driving resistance in hospitals is a consequence of widespread use of broad-spectrum antibiotics which distribute widely in the body.[162] Antibiotic resistance in colonising bacteria is clinically important, as asymptomatic carriage of these bacteria typically precedes invasive infection, and preventing colonisation will reduce associated morbidity and mortality.[163–165] For these reasons, this study focuses on the bystander selective pressure resulting from different durations of antibiotic treatment on colonising bacteria.

The effect of treatment duration on bystander selective pressure is potentially affected by pathogen, host, antibiotic and environmental factors. Some important considerations include how quickly antibiotic use leads to resistance, e.g. through selective growth of sub-populations of resistant bacteria within-host versus *de novo* resistance mutations, if and how quickly the treated individual becomes colonised by resistant bacteria through transmissions, the antibiotic's spectrum of coverage, and the growth rates of the susceptible and resistant bacteria in the presence and absence of antibiotics. The complex interplay of these within- and between-host factors determines the effect of treatment duration on the prevalence of resistant bacteria in a given setting.

In this study, to gain a mechanistic understanding of the relationship between antibiotic treatment duration and the prevalence of colonisation by antibiotic-resistant bacteria in hospitalised patients, I modelled within- and between-host

dynamics of susceptible and resistant bacteria in response to systemic antibiotic treatment. I then sought to compare the model findings with empirical observations. To do this, I conducted a systematic review and meta-analysis of randomised controlled trials that studied the effect of antibiotic treatment duration on patient carriage of antibiotic-resistant bacteria. Using a Bayesian meta-regression model, I quantified the daily risk of acquiring resistant bacteria colonisation with increasing antibiotic treatment duration.

4.2 Methods

4.2.1 Modelling dynamics of colonising bacteria in response to clinically-indicated antibiotic treatment

Three stochastic agent-based models with increasing complexity and biological realism were constructed. From here on, they will be referred to as the exclusive colonisation model, co-colonisation model and within-host growth model (Figure 4.1). I aimed to compare the model outputs and identify the assumptions that accounted for differences in the outcomes.[166]

In all the models, I simulated individual patients in a hospital ward environment. The patients' colonisation status changed from day to day during hospitalisation due to i) differential growth/killing rates of resistant and sensitive bacterial populations within a patient, ii) transmission events between patients, and iii) loss of carriage (Figure 4.1). All these processes are assumed to be affected by antibiotic use. Because the population size under consideration is small and local fade-out events (when resistant populations reach zero) are potentially important, I considered stochastic implementations of these models (Figure 4.2). The details of the model mechanisms can be found below in Box 4.1.

Within the same ward, patients who were administered antibiotics received treatment with the same mean duration. I considered two scenarios. In the first, the resistant phenotype in the colonising bacteria was assumed to be susceptible to one of the administered antibiotics, e.g. carbapenems against third generation cephalosporin resistance. Patients in this scenario could be prescribed with two types of antibiotics; either a narrow-spectrum antibiotic which was assumed to have activity against the susceptible but not the resistant bacteria, or a broad-spectrum antibiotic which was effective against both the susceptible and resistant bacteria. In the second scenario, there was no effective antibiotic available against the resistance phenotype in the colonising bacteria, e.g. colistin or carbapenem resistance. Here, patients could only be prescribed with antibiotics which were effective against susceptible bacteria but not against resistant bacteria. Subsequently, I refer to these two scenarios as 'administered antibiotics have activity against resistant and susceptible organisms', and 'administered antibiotics have activity only against susceptible organisms'.

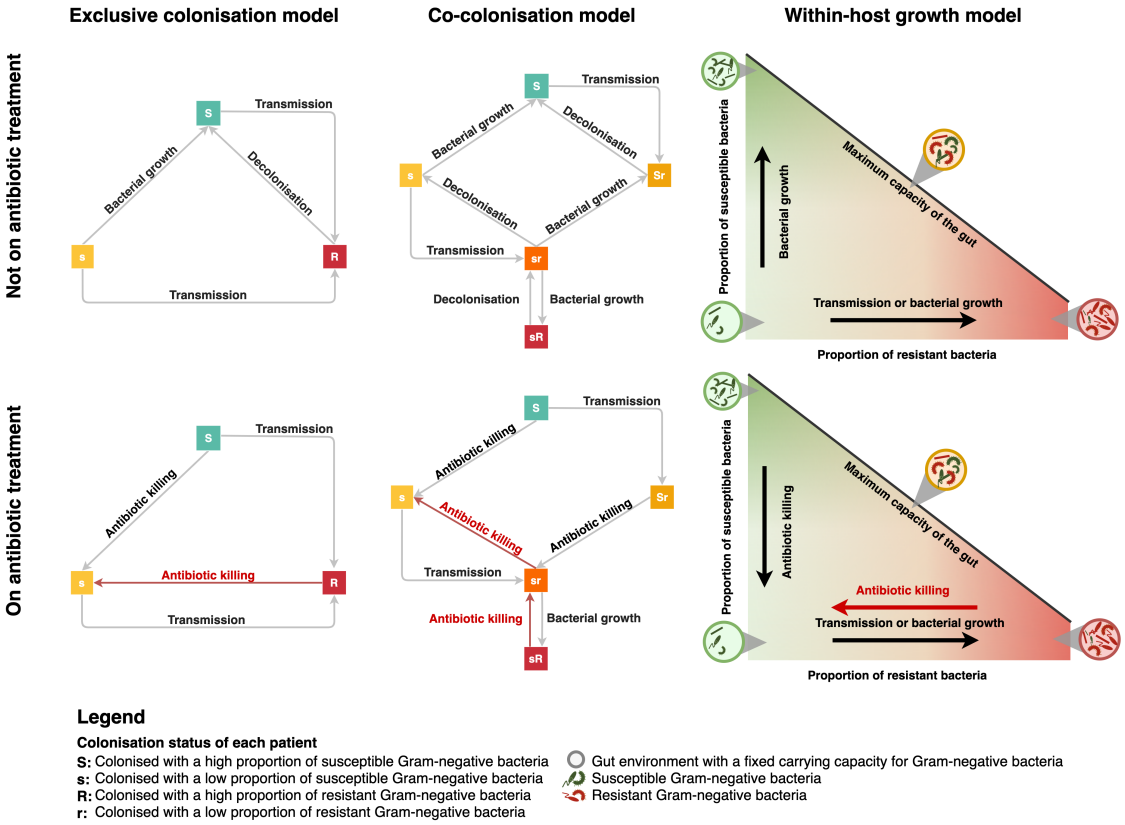


Figure 4.1: Host states and transitions in the exclusive colonisation, co-colonisation and within-host growth models. In the three models, all patients were assumed to be colonised with bacteria, but in the exclusive colonisation model there is complete bacterial interference, and patients can be colonised with resistant bacteria (R), and high or low levels of sensitive bacteria (S or s), but not both resistant and sensitive at the same time. The co-colonisation model makes the more realistic assumption that patients may be colonised with both resistant and sensitive bacteria at different levels of abundance. Similarly in the within-host growth model, patients carry both resistant and sensitive bacteria but their combined populations cannot exceed a maximum carrying capacity. Green, orange and red squares/circles represent the host states when the host is carrying susceptible, a mix of susceptible and resistant, and resistant bacteria respectively. The red arrows represent antibiotic killing of resistant bacteria when a patient is treated with an antibiotic effective against both susceptible and resistant bacteria.

The models assumed that antibiotics act within-host by selectively promoting the growth of existing resistant bacteria in the gut,[167, 168] and between-host by predisposing the treated individual to become more likely to transmit resistant bacteria as a consequence of a higher bacterial load (Figure 4.1).[169] Spontaneous decolonisation of resistant bacteria was assumed to occur only when a patient was not receiving antibiotics to which the colonising bacteria were resistant.

In the exclusive colonisation model, patients could be carriers of either susceptible or resistant bacteria but not of both at the same time. In this model, for a carrier of

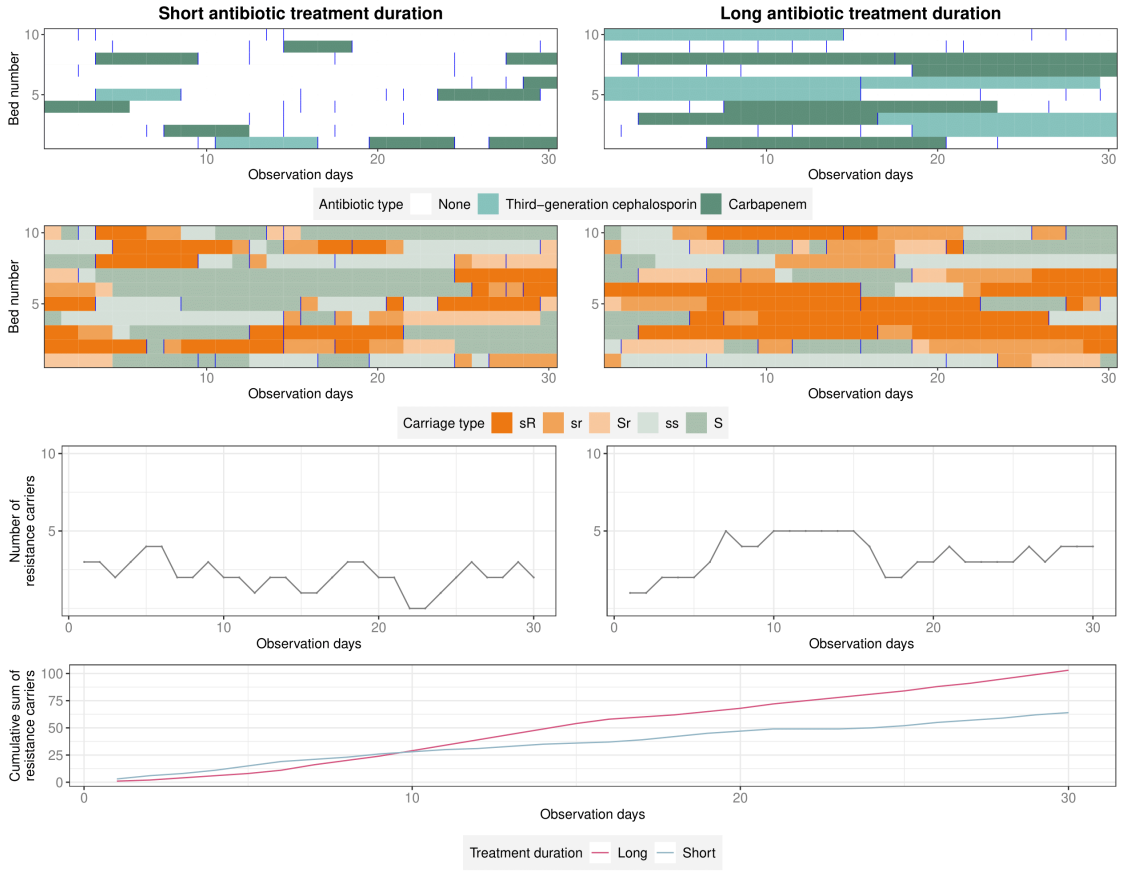


Figure 4.2: Example output from the co-colonisation model. This example simulation shows two 10-bed wards with the same transmission risk of resistant bacteria, patient lengths of stay, proportion of patients who required antibiotic treatment and proportion of patients who were resistance carriers upon admission. The left panels represent the ward where antibiotic durations were short (mean of five days), while the right panels represent the ward where antibiotic durations were long (mean of 20 days). Top row panels: Antibiotic treatment types and duration. Each square in the graphs represents one patient day. Light and dark green squares indicate one day of third-generation cephalosporin or carbapenem administered respectively. Vertical blue lines represent new admissions to the ward. Second row panels: The carriage status for each patient on a particular day. The increasingly darker shades of orange and green indicate an increasing amount of resistant and susceptible colonising bacteria carried by each patient respectively. Third row panels: The number of antimicrobial resistance carriers in each ward per day. Fourth row panel: The cumulative number of resistance carriers in each ward over 30 observation days.

susceptible bacteria to become a carrier of resistant bacteria required transmission of resistant bacteria from another carrier in the ward (i.e. I did not consider *de novo* resistance emergence or horizontal transfer from other components of the flora). In the other two more complex models, patients could carry both susceptible and resistant bacteria simultaneously with time-varying levels of abundance. In the co-colonisation model these abundances were dichotomised into *high* and *low*, and only bacteria carried at a high level were able to be transmitted to other patients.

In the within-host growth model bacterial population sizes could vary continuously, but had to be above a threshold in order to be capable of transmission to others. I also assumed a total carrying capacity for Gram-negative bacteria [170] and note that newly admitted patients may carry fewer bacteria than the carrying capacity as I would expect some patients were prescribed antibiotics prior to admission.

While the modelling framework I have adopted is quite general, here I focus on areas of parameter space that are appropriate for Gram-negative bacteria such as Enterobacteriaceae and Gram-negative non-fermenting bacilli, as they are the commonest cause of urinary tract and bloodstream infections, and frequently associated with multidrug resistant hospital-acquired infections. The predominant mechanisms for resistance dissemination in Gram-negative bacteria colonising the gut are horizontal transfer of mobile genetic elements and clonal expansion.[171]

I first explored the models by varying pairs of parameters over a grid of values while holding the other parameter values constant. This was followed by global sensitivity analysis using Latin hypercube sampling and partial rank correlation coefficient (Appendix Section C.1). In the global sensitivity analysis, real-world parameter values were obtained from the literature and used in these explorations (Table 4.1). In all results shown below, each simulation was produced from at least 40,000 iterations (100 repeats of 400 unique sets of parameters selected within the ranges shown in Table 4.1). The computer code is publicly available (https://github.com/moyinNUHS/abxduration_abm).

The models were used to assess how changing duration of antibiotic treatment affect the risk of colonisation with resistant bacteria at both the individual and population levels. Three types of outcomes were considered: resistance carriage amongst i) patients who were treated with antibiotics and therefore directly affected by different treatment durations; ii) overall carriage of resistance within a ward population that consisted of patients who were treated and not treated (i.e. indirectly affected by the treatment duration received by those treated patients); and, iii) patients who were not carriers of resistant bacteria when admitted to the ward. To evaluate reducing antibiotic duration as an intervention, all the model outputs were assessed as the absolute difference between the short and long duration wards.

All simulations are performed with R version 4.0.4 (2021-02-15)[86] using packages `msm`,[172] `pse`,[173] and `spartan`.[174] Code review was done using unit tests for each function's base scenarios, which can be found under `/unit_tests` in the source code.

Box 4.1 Details of the simulation models

The agent-based models were set up to simulate a randomised controlled trial involving two identical hospital wards. Patients who were admitted to the hospital were randomly assigned to either ward. These two wards shared

similar ward and patient characteristics, i.e. lengths of stay, bed capacity, transmission rate of resistant bacteria, proportion of patients who required antibiotic treatment and proportion of patients who were resistance carriers at admission. The wards were assumed to be fully occupied at all times.

Patients were prescribed antibiotics for clinical infections at admission and during hospitalisation. The prescribed antibiotics could be active against both susceptible and resistant bacteria or active against only susceptible bacteria. The wards differed only in antibiotic duration. For example, in the scenario where antibiotics administered were active against both susceptible and resistant bacteria, patients admitted to the long duration ward who required antibiotics received this type of antibiotic for a long duration while the patients in the short duration warded received the same type of antibiotic for a shorter duration.

The patients' colonisation statuses were updated daily, depending on whether antibiotics were prescribed, whether the antibiotics prescribed were active against resistant bacteria or selectively promoted the growth of resistant bacteria, transmission of antibiotic resistant bacteria and bacterial growth (Figure 4.1). The time step was kept sufficiently small at one-day intervals so that the sum of probabilities for competing events was less than one. This allowed for only one event to take place during each time step.

A new resistance carrier may be added to the ward in three ways: i) transmission events, i.e. newly acquiring resistant bacteria from existing resistance carriers in the ward; ii) within-host selection of pre-existing resistant bacteria due to receiving antibiotics which are only active against susceptible bacteria; or iii) resistance carriers are admitted to the ward. The probability of acquiring resistant bacteria by transmission depended on the number of patients in the ward who were existing resistance carriers:

$$\zeta_{i,t} \sim \text{Bernoulli}\left(1 - (1-\varphi)^{n_t^r}\right)$$

where $\zeta_{i,t}$ is a binary indicator which takes the value 1 if a transmission event occurs and 0 otherwise, for an individual i at time step t. φ is the probability of a patient being transmitted resistant bacteria from one existing carrier in one time step. n_t^r is the number of patients in the ward who were resistant carriers at time step t. Hence, the probability that no one infects the patient under consideration in one time step is $(1-\varphi)^{n_t^r}$, and the probability that at least one person infects the patient is $1-(1-\varphi)^{n_t^r}$.

The three models are described in detail below.

- **Exclusive colonisation model**

  Patients may assume any of the three carriage statuses: high proportion of susceptible bacteria (S), low proportion of susceptible bacteria (s), or

resistant bacteria (R). Co-carriage of susceptible and resistant bacteria was not allowed in this model.

When administered antibiotics which were active against both susceptible and resistant bacteria, an individual in S or R state might change to s at the next time step due to effective antibiotic killing. Without antibiotics, an individual in s state might change to S due to regrowth of the susceptible bacteria, and an individual in R state might decolonise spontaneously to S. An individual now in either s or S states became a resistant carrier if there was a transmission event. Individuals in state S, by having a high proportion of susceptible bacteria, were protected from acquiring antibiotic-resistant bacteria through the 'bacterial interference factor', b:[175]

$$\varphi_{S\to R} = \varphi_{s\to R}(1-b)$$

- **Co-colonisation model**

  Co-carriage of susceptible and resistant bacteria was incorporated in this model. Patients might be in any of the five carriage statuses: high proportion of susceptible bacteria (S), low proportion of susceptible bacteria (s), dominant population of susceptible bacteria with some resistant bacteria (Sr), low proportion of both susceptible and resistant bacteria (sr) or dominant population of resistant bacteria with some susceptible bacteria (sR).

  When on effective antibiotics, an individual in S, Sr or sR state might change to s or sr due to antibiotic killing. However when the administered antibiotics were only active against susceptible bacteria, selective pressure might promote the growth of resistant bacteria and enable an individual to change from sr to sR. Without antibiotics, susceptible and resistant bacteria might grow and change from s or sr to S or Sr states respectively, or sr to sR.

  Spontaneous decolonisation might occur in the absence of antibiotics, resulting in individuals changing from sr or Sr states to s or S respectively. Transmission events for individuals in S or s states might change their status to Sr or sr respectively. Only patients who carried a high proportion of resistant bacteria (sR) could transmit resistance to others and were considered resistance carriers. Similar to the exclusive colonisation model, individuals in S state were protected from acquiring drug resistant bacteria through the bacterial interference factor.

- **Within-host growth model**

The within-host growth model explicitly calculated the amounts of susceptible and resistant bacteria in the gut with the logistic growth function, which describes the typical bacterial within-host growth in lag, exponential growth, and stationary phases. The initial bacterial growth rate was proportional to the existing population size (giving exponential growth) and decreased as the population size approached the maximum carrying capacity, imposed by limited resources in the environment. I assumed that both susceptible and resistant bacteria occupied a similar ecological niche and hence shared the same maximum carrying capacity. The total numbers of resistant and sensitive bacteria at each time step were a consequence of bacterial growth, antibiotic killing, and transmission events.

In each iteration, a newly-admitted individual i has a total carrying capacity for bacteria of $K_i \sim \mathcal{N}(\mu_k, 1)$. The amount of resistant bacteria, N_i^r carried by individual i at the time of admission, $t = 0$, is $N_{i,t=0}^r = K_i \rho_e \rho_{r,i}$, where ρ_e is the proportion of the total carrying capacity of bacteria present in the gut and $\rho_{r,i}$ is the proportion of the existing amount of bacteria present in the gut that is resistant with $\rho_{r,i} \sim \text{Beta}(0.2, 2)$. The distribution $\text{Beta}(0.2, 2)$ is taken from an analysis of within-host extended-spectrum beta-lactamase resistance dynamics by Niehus *et al.*[170], where the copies of CTX-M beta-lactamases detected per stool sample were reported. The amount of susceptible bacteria, N_i^s, carried by individual i at the time of admission, $t = 0$, is $N_{i,t=0}^S = K_i \rho_e - N_{i,t=0}^r$.

The change in the amount of susceptible bacteria after admission for individual i per time step t is determined by the following processes:

i) Growth, which is modelled with the logistic growth function:

$$G_{i,t}^s = c^s N_{i,t-1}^s \left(1 - \frac{N_{i,t-1}^s + N_{i,t-1}^r}{K_i}\right),$$

where c^s is the growth constant for susceptible bacteria.

ii) Antibiotic killing:

$$A_{i,t}^s = \delta_{i,t-1} \alpha^s \left(N_{i,t-1}^s + G_{i,t}^s\right),$$

where $\delta_{i,t-1}$ takes a value of 1 if the individual i was on antibiotics at time step $t-1$ and 0 otherwise, and α^s is the antibiotic killing constant.

Hence, the change in the population of susceptible bacteria within an individual i per time step t is

$$\Delta N_{i,t}^s = G_{i,t}^s - A_{i,t}^s.$$

The change in the amount of resistant bacteria after admission for individual i per time step t is determined by the following processes:

i) Growth, which is modelled with the logistic growth function

$$G^r_{i,t} = c^r N^r_{i,t-1}\left(1 - \frac{N^s_{i,t-1} + N^r_{i,t-1}}{K_i}\right),$$

$$c^r = fc^s,$$

where c^r is related to c^s by f, a fitness factor for resistant bacteria.

ii) Antibiotic killing:

$$A^r_{i,t} = \delta'_{i,t-1}\alpha^r\left(N^r_{i,t-1} + G^r_{i,t}\right),$$

where $\delta'_{i,t-1}$ takes a value of 1 if the individual, i, was on antibiotics at time step, $t-1$, and 0 otherwise, and α^r is the antibiotic killing constant.

iii) Transmission:

$$T^r_{i,t} = \zeta_{i,t}\tau N^r_{i,t},$$

where $\zeta_{i,t}$ is a binary indicator which takes the value of 1 if a transmission event occurs and 0 otherwise for an individual i at time step t, and $\tau N^r_{i,t}$ is the amount of resistant bacteria transmitted to the receiving individual, i.

Hence, the change in resistant bacteria within an individual i per time step t is

$$\Delta N^r_{i,t} = G^r_{i,t} - A^r_{i,t} + T^r_{i,t}.$$

When $N^s_{i,t} + N^r_{i,t} > K_i$, I set $N^r_{i,t} = \frac{N^r_{i,t}}{N^s_{i,t}+N^r_{i,t}}K_i$ and $N^S_{i,t} = \frac{N^S_{i,t}}{N^s_{i,t}+N^r_{i,t}}K_i$ i.e. ensuring that the total carrying capacity is not exceeded.

An individual carrying a proportion of resistant bacteria higher than the threshold, Γ, is considered a resistance carrier. This equates to an absolute number of bacteria $N^t = \Gamma K \rho_e$.

Symbol	Description	Models	Range†	Unit	Reference
Ward level					
n	Number of beds in the ward	1, 2, 3	5-50	Beds	
l	Maximum length of stay for each patient in the ward	1, 2, 3	3–20	Days	
Individual level					
Individual carriage status on initial admission to the ward					
p_R	Probability of carrying resistant bacteria (R)	1, 2, 3	0.01–0.8		
p_r	Probability of carrying low proportion of resistant bacteria (Sr or sr)	2	0–1		
p_{Sr}	Probability of carrying high proportion of susceptible and low proportion of resistant bacteria (sR)	2	0–1		
p_S	Probability of carrying susceptible bacteria (S)	1, 2	0–1		
ρ_e	Proportion of the total carrying capacity occupied	3	0.1–0.9		[170, 176]
μ_k	Mean total carrying capacity for bacteria	3	e^{18-24}	Bacteria/ml	[177]
Growth of bacteria					
g_s	Probability of susceptible bacteria repopulating ($s \rightarrow S$ and $sr \rightarrow Sr$)	1, 2	0.02–0.12		[178–180]
c_s	Exponential growth constant for susceptible bacteria	3	0.1–2		[181]
f	Growth fitness factor for resistant bacteria (multiplied by c_s to derive exponential growth constant for resistant bacteria)	2, 3	0.5–2		[182]
Transmission of resistant bacteria					
Γ	The threshold proportion of resistant bacteria carried for the carrier to transmit resistant bacteria to others	3	0.01–0.2		
τ	Proportion of resistant bacteria carried that is transmitted during a transmission event	3	0.01–0.5		
φ	Probability of transmitting resistant bacteria from a resistance carrier to non-resistance carriers	1, 2, 3	0.001–0.3	/day	[183, 184]
b	Bacterial interference factor, a protection factor against being transmitted resistant bacteria in those carrying a high proportion of susceptible bacteria compared to those carrying a low proportion of susceptible bacteria on a relative scale	1, 2	0–1		
Decolonisation of resistant bacteria					
μ	Probability of resistance decolonisation ($R \rightarrow S$)	1	0.002–0.02	/day	[185]
μ	Probability of resistance decolonisation ($sR \rightarrow sr$, $Sr \rightarrow S$ and $sr \rightarrow s$)	2	0.002–0.02	/day	

Table 4.1 continued from previous page

Symbol	Description	Models	Range†	Unit	Reference
Antibiotic killing					
α_s	Probability of bacterial killing by an antibiotic active against susceptible bacteria (S and s)	1, 2	0.1–0.5	/day	[186–188]
α_s	Mean proportion reduction in susceptible bacteria due to effective antibiotic killing	3	0.1–1	/day	
α_r	Probability of bacterial killing by an antibiotic active against resistant bacteria (R and r)	1, 2	0.1–0.5	/day	
α_r	Mean proportion reduction in resistant bacteria due to effective antibiotic killing	3	0.1–1	/day	
Number of antibiotic prescriptions					
ω_{day1}	Probability of being prescribed any antibiotics on admission day one	1, 2, 3	0.1–1		
$\omega_{\text{day1.r}}$	Probability that the antibiotic prescribed on admission day one is active against resistant bacteria	1, 2, 3	0.1–1		
ω_{after}	Probability of being prescribed any antibiotics during admission	1, 2, 3	0.01–0.13	/day	
$\omega_{\text{after.r}}$	Probability the antibiotic prescribed during admission is active against resistant bacteria	1, 2, 3	0.1–1		

Table 4.1: Parameters used in the models and their respective ranges obtained from literature review. †Parameter values were drawn from uniform distributions on the specified ranges.

4.2.2 Systematic review

To compare model findings with empirical evidence, I performed a systematic review of randomised controlled trials which studied the effect of antibiotic treatment duration on patient carriage of antibiotic resistant bacteria. I searched MEDLINE and EMBASE for randomised controlled trials published from 1 January 2000 to 20 May 2021 which allocated participants to varying durations of systemic antibiotic treatments. Studies that compared antibiotics to no antibiotic treatment were excluded.

The study methodologies and outcomes were expected to be highly diverse and unlikely to be appropriate for including in a single analysis. Hence I identified a homogeneous group defined by studies which i) measured asymptomatic carriage of resistant bacteria as an outcome, and ii) collected cultures at pre-defined time point(s) for surveillance (as oppose to clinically indicated) during follow-up. The other specific inclusion criteria and search strategy were the same as the literature review reported in Chapter 1 (Table 1.1). The indications for antibiotic treatment in the trials included treatment or prevention of bacterial infections, decolonisation of resistant bacteria, and anti-inflammation for autoimmune diseases. The sites of colonisation included the digestive tract, respiratory tract and urinary tract.

The specific aims of the systematic review were to: i) evaluate the quality of evidence in the current literature of antibiotic duration effects on resistance carriage, ii) describe and summarise the findings from these studies, and iii) compare the findings to the model outputs.

4.2.3 Meta-analysis

From the studies identified in the systemic review, I selected for a meta-analysis those which monitored resistant Gram-negative bacteria carriage. The meta-regression analysis was performed using a Bayesian mixed-effects model to estimate the change in the daily risk of acquiring resistant bacteria colonisation per day of antibiotic consumption. Due to the small number of studies identified from inpatient settings, I included studies conducted in both inpatient and outpatient settings in the meta-analysis. The dependent variable in the meta-regression analysis was the number of patients colonised with resistant bacteria in a given arm of a given trial (together with the associated denominator). These data were analysed with a logistic regression using antibiotic duration administered to the patients in each arm, elapsed time from antibiotic administration to surveillance culture, and healthcare setting (inpatient versus outpatient) as independent variables. I allowed slopes and intercepts to vary between trials (Model 1 in Box 4.2).

I also considered two other models. In the first alternative model, I excluded the healthcare setting as an independent variable (Model 2 in Box 4.2). In the second alternative model, I assumed each arm in each trial to be independent (Model 3 in Box 4.2). Model comparison was done using the Widely Applicable Information

Criterion (WAIC) where lower values indicate improved model fit.[189] Details of the models are described below in Box 4.2.

Box 4.2 Details of the meta-regression models

The meta-regression analysis was performed using a Bayesian regression model to estimate the change in the daily risk of acquiring resistant bacteria colonisation per day of antibiotic consumption. A linear relationship was assumed between the log of the daily acquisition risk and days of antibiotic intake.

The number of patients colonised with resistant bacteria, Y_{ij}, in arm j of study i was modelled by a binomial distribution:

$$Y_{ij} \sim \text{Bin}(n_{ij}, p_{ij}),$$

where n_{ij} is the total number of participants in arm j of trial i, and p_{ij} is the probability of colonisation over the follow-up period.

I compared three models for p_{ij}:

i) **Model 1**

$$\text{logit}(p_{ij}) = \alpha_i + \beta_i t_{ij} + c_i f_{ij} + d_i w_{ij}$$

where t_{ij} is the mean duration of antibiotic use in arm j of trial i and p_{ij} is the probability of colonisation over a follow-up period of f_{ij} days. The setting where the trial was conducted, i.e. inpatient or outpatient setting, was indicated by a binary variable, w_{ij}.

ii) **Model 2**

$$\text{logit}(p_{ij}) = \alpha_i + \beta_i t_{ij} + c_i f_{ij}$$

In models 1 and 2, the intercepts, α_i, and slopes, β_i, c_i, and d_i were allowed to vary between trials and assumed to be normally distributed.

iii) **Model 3**

$$\text{logit}(p_{ij}) = \alpha_{ij} + \beta_{ij} t_{ij} + c_{ij} f_{ij} + d_{ij} w_{ij}$$

In model 3, intercepts, α_i, and slopes, β_i and c_i were allowed to vary between arms of each trial and assumed to be independent and normally distributed.

I implemented the above meta-regression model in JAGS using the `R2jags` package,[190] and performed all analyses in R version 4.0.4 (2021-02-15).[86]

4.3 Results

4.3.1 Individual and population resistance dynamics with antibiotic treatment

I first consider as an illustrative example of model behaviour a ward where 25% of the patients carried resistant bacteria on admission and half of the patients were given antibiotics upon admission (Figure 4.3). In the scenario where the patients were given an antibiotic with activity against both susceptible and resistant colonising bacteria of interest (Figure 4.3, top 2 panels), in all three models carriage of resistant bacteria population in treated individuals decreased over the period of treatment. At a population level, this led to a reduction in the number of carriers of resistant bacteria, resulting in a ward-level prevalence lower than the baseline carriage prevalence of resistant bacteria on admission. Because there were fewer carriers of resistant bacteria, less transmission of resistant bacteria between patients occurred.

In the other case, where administered antibiotics had activity only against susceptible organisms (Figure 4.3, bottom 2 panels), the antibiotics prescribed to the patients resulted in a decline in the susceptible bacterial population but selectively favoured the growth of resistant bacteria. This bystander selection led to an increased prevalence of resistance carriers. More resistance carriers in the ward then acted as a reservoir for further spread of resistance to other non-carriers.

4.3.2 Key parameters in the relationship between antibiotic duration and resistance carriage

The above within- and between-host dynamics of resistance (i.e. competitive bacterial growth, selective pressure for resistance, antibiotic killing, and transmission) jointly determine the impact of antibiotic duration on resistance carriage. The global sensitivity analysis enabled the exploration of how the parameters in the three models modified the differential effect of short versus long antibiotic treatment on resistance at the individual and population level (Figure 4.4). This differential effect can be considered as the effect of implementing a ward-wide antibiotic stewardship policy on the resistance carriers in the ward, where antibiotic durations were reduced for all patients admitted.

The global sensitivity analysis showed that for population-level outcomes, when considering treatment with an antibiotic to which the colonising resistant bacteria were susceptible, the more frequently and longer antibiotics were prescribed, the greater the reduction in resistance in long duration wards compared to short duration. In contrast, when considering treatment with an antibiotic to which colonising resistant bacteria were not susceptible, increased frequency and duration of antibiotic prescribing led to more resistance in the long duration ward relative to the short duration ward (Figure 4.4 top row, middle panel). However, amongst those

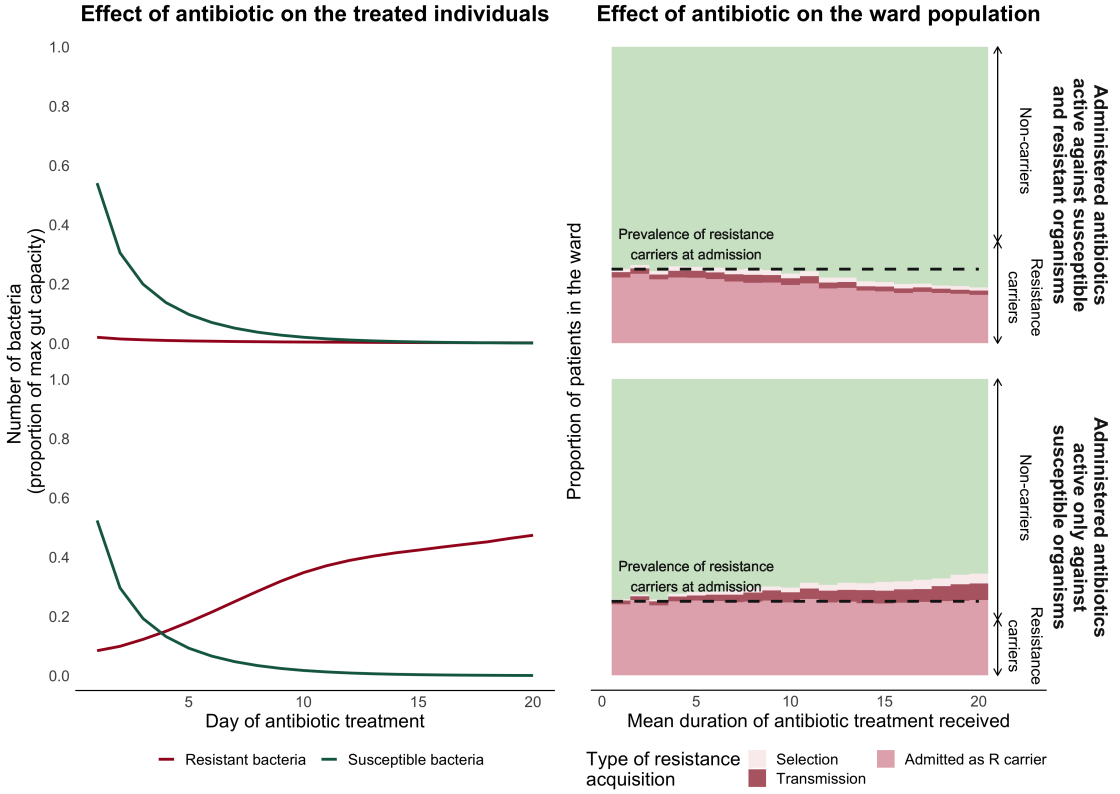


Figure 4.3: Effect of antibiotic duration on antibiotic resistance carriage at individual and population levels. A 50-bed ward was simulated in which 25% of the patients were already colonised with resistant bacteria when admitted to the ward. Half of the warded patients were administered antibiotics assumed to start as soon as they were admitted. The panels on the left depict the mean within-host dynamics of susceptible (green line) and resistant (red line) bacteria with increasing days of antibiotic treatment amongst the treated patients who received 20 days of antibiotics. The panels on the right show the equilibrium prevalence of resistance carriers on the ward as a function of antibiotic treatment durations. The proportion of resistance carriers (represented by shades of red) and non-carriers (represented by green) in the ward is plotted against antibiotic treatment time. The top panels show the scenario where administered antibiotics have activity against both susceptible and resistant bacteria. The bottom panels show the scenario where administered antibiotics have activity only against susceptible organisms. Results shown are from the within-host growth model, but qualitatively similar results are obtained with the other two models.

who were non-carriers on admission, more patients gained resistance carriage with increased frequency and duration of treatment even when the colonising resistant bacteria were susceptible to the antibiotic prescribed (Figure 4.4, top row, right-side panel). This is because, despite the absence of within-host bystander selection, those treated patients could still become increasingly at risk of acquiring resistant bacteria from other existing carriers with longer antibiotic treatment.

The analysis identified both transmission rates and prevalence of resistance on

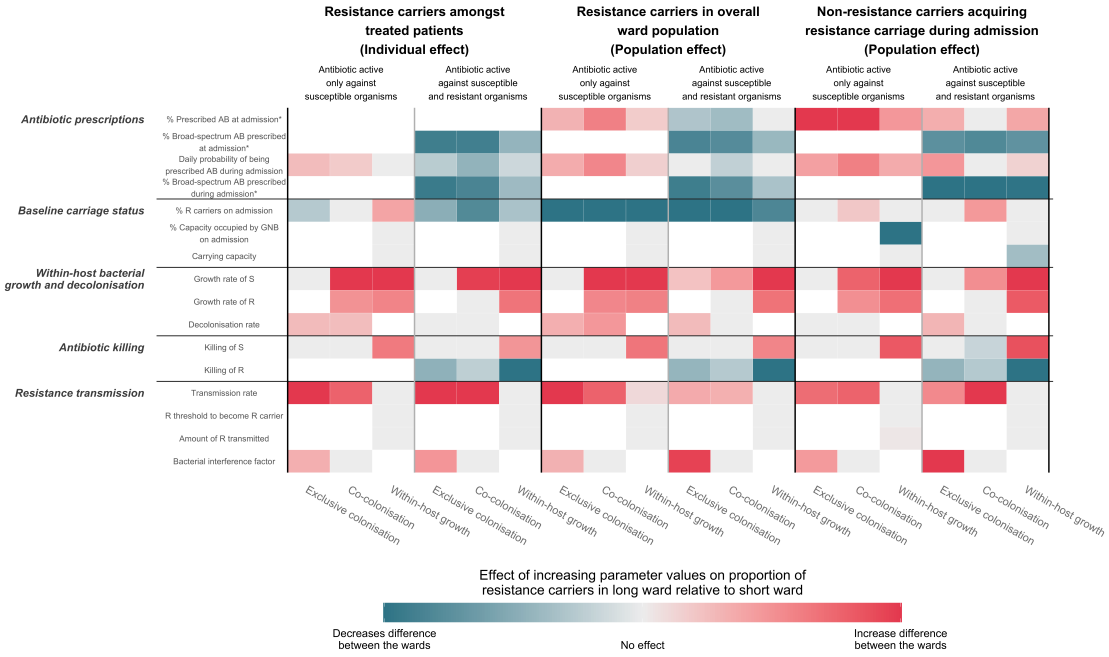


Figure 4.4: Global sensitivity analysis. Partial rank correlation coefficients for the various parameters are shown in the heatmap. Parameters from the three different models are grouped according to the main within- and between-host processes they describe. The three panels (separated by black vertical lines) represent the three types of outcomes assessed. Within each panel, the first three columns refer to the scenario where administered antibiotics have activity only against susceptible organisms; the last three columns refer to the scenario where administered antibiotics have activity against resistant and susceptible organisms. Red, blue and grey rectangles indicate that as the parameter value increased, the difference in resistance carriers between the long and short duration wards increased, decreased and did not change respectively. GNB: Gram-negative bacteria; AB: Antibiotic; S: Susceptible; R: Resistant. White rectangles indicate that the parameter was not found in the model or not applicable. *Parameters for the percentage of patients who were administered antibiotics are not relevant for the outcome observed in the treated patients and are therefore left white.

admission as important effect modifiers of the relationship between treatment duration and resistance prevalence (Figure 4.4, rows labelled 'baseline carriage status' and 'resistance transmission'). Reducing antibiotic treatment duration was more effective in reducing resistance carriage in the overall ward population when the transmission rate was high across all model outcomes and regardless of whether the administered antibiotics were active against resistant organisms. As for the prevalence of resistance on admission, the lower the resistance prevalence on admission, the more effective reducing treatment duration was in reducing resistance carriage at a population level. This is because importation of resistance carriers is not affected by antibiotic treatment. Hence for the ward population, the lower the prevalence of resistance carriage on admission, the greater effect antibiotic duration has on overall resistance. However amongst the patients who were treated and who were non-carriers on admission, reducing antibiotic treatment duration might be

more effective when the prevalence of resistance carriage on admission was high. This is because in these subsets of patients, the gain in resistance carriage was mainly through transmission events from existing carriers rather than within-host bystander selection.

Within-host, a rapid growth rate of both susceptible bacteria (i.e. recovery after antibiotics had been stopped and competitively outgrowing resistant bacteria) and resistant bacteria (i.e. high selective pressure and low killing rate during antibiotic treatment) resulted in more resistance carriers in the long duration ward compared to the short duration ward (Figure 4.4, rows labelled 'within-host bacterial growth and decolonisation'). The intuition is that faster growth of susceptible bacteria would lead to less resistance carriers regardless of antibiotic duration due to competitive growth in the same ecological niche, but because this reduction was more prominent in patients from the short duration ward, so there would be comparatively more resistance carriers in the long duration ward.

A difference between the models in this global sensitivity analysis is that rapid growth of susceptible bacteria was an important parameter associated with reducing resistance favouring short duration in the co-colonisation and within-host growth models but not the exclusive colonisation model (Figure 4.4 row labelled 'growth rate of S'). This is because in the co-colonisation and within-host growth models, stopping antibiotics early led to fast regrowth of susceptible bacteria, which offered protection against colonisation by resistant bacteria. Since the exclusive colonisation model only allowed carriage of either susceptible or resistant bacteria, this protection offered by susceptible bacteria was encapsulated in the 'bacterial interference factor' (Figure 4.4, row labelled 'bacterial interference factor'). The bacterial interference factor was an important factor in the relationship between antibiotic duration and resistance across all outcomes in the exclusive colonisation model. The higher the bacterial interference factor, the more effective reducing antibiotic treatment duration was in reducing resistance carriage in the overall ward population.

Since resistance phenotypes that are extensively resistant to most available antibiotics are of most clinical concern, in the following sections I focus on this scenario and consider the effects of long versus short antibiotic duration on both within-host processes in the treated individuals and on population level processes.

4.3.3 Within-host dynamics of colonising bacteria with varying antibiotic treatment duration

Bacterial growth and decolonisation rate in the absence of antibiotics were important within-host factors in the relationship between treatment duration and resistance carriage regardless of whether administered antibiotics were active against resistant organisms. Under both fast and slow growth scenarios I found that long duration treatment allowed the resistant population to become established at a much larger proportion of the total bacterial load at the end of treatment, which in turn

resulted in a longer time for the resistant population to decline to low levels once treatment ceased (Figure 4.5).

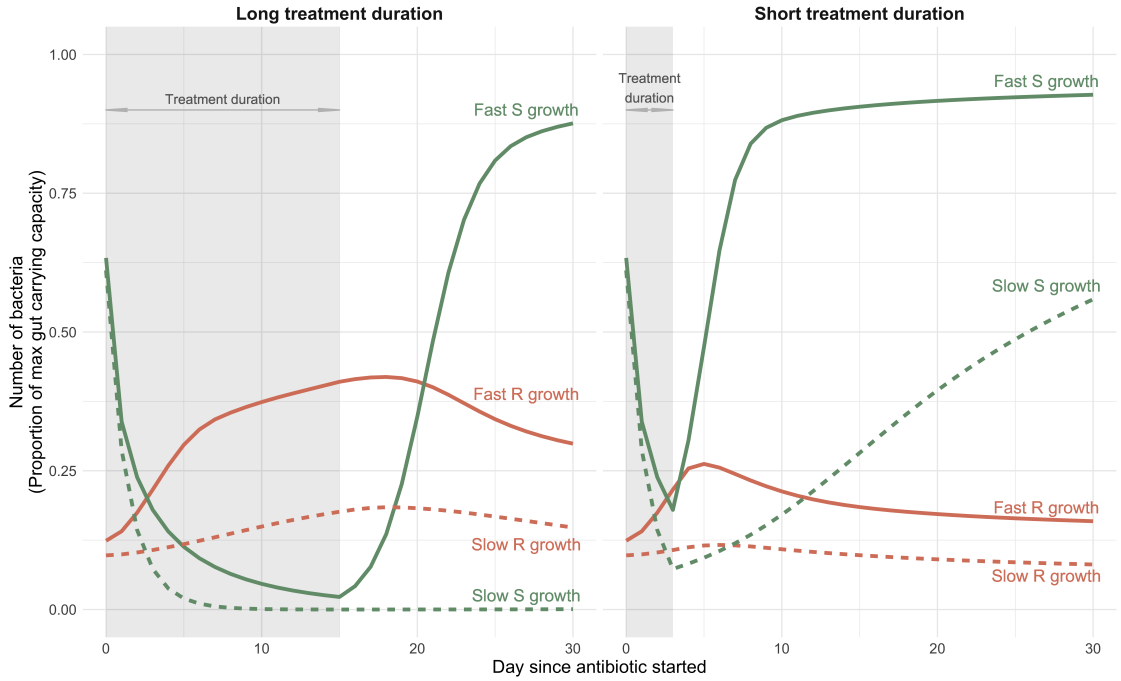


Figure 4.5: Example output from the within-host growth model showing effect of antibiotic duration on competitive growth of susceptible and resistant colonising bacteria at different growth rates. These graphs illustrate an example where susceptible and resistant bacteria within the same host were exposed to 15 days (left panel) versus three days (right panel) of antibiotic treatment. This antibiotic was capable of killing susceptible bacteria but ineffective against resistant bacteria. All parameter values other than bacterial growth rates were kept the same. Rapid bacterial growth is represented by solid lines. Slow bacterial growth is represented by dashed lines.

During antibiotic treatment, susceptible bacteria became rapidly suppressed. Resistant bacteria, on the other hand, were able to grow. The more rapidly these resistant bacteria were able to grow beyond a threshold, the sooner hosts would be able to transmit resistant bacteria to others in the ward (Figure 4.5, solid red lines). If the resistant bacteria grew at a slow rate, a resistant population of sufficient size to pose a significant transmission risk might not become established even with prolonged antibiotic treatment (Figure 4.5, dashed red lines).

Upon termination of antibiotics, there was no longer a selective growth advantage for resistant bacteria, and the population of susceptible bacteria rebounded. When antibiotics were not being used, the more rapidly the susceptible bacteria outgrew the resistant bacteria, the quicker the host was cleared of carriage of resistant bacteria (Figure 4.5, solid green lines). However when the susceptible bacteria recovered slowly after termination of antibiotics, and especially coupled with rapid acquisition of resistance soon after starting antibiotics, patients who were administered antibiotics could remain colonised with resistant bacteria for prolonged

periods regardless of the treatment duration (Figure 4.5, dashed red lines).

4.3.4 Resistance transmission and importation on the effects of antibiotic duration

In the scenario where administered antibiotics have activity only against susceptible organisms, antibiotic treatment had different effects on treated non-carriers and carriers (Figures 4.6 and 4.7). For the treated non-carriers, antibiotics increased these patients' risk of acquiring resistant bacteria by reducing the amount of colonising bacteria to below the carrying capacity. For the existing carriers, antibiotics preferentially promoted the growth of resistant bacteria, increasing these patients' probability of transmitting resistance to others. Hence, a higher transmission rate amplified the effect of treatment duration on increasing resistance carriers by bridging the transfer of resistance from the carriers to the at-risk non-carriers.

On the other hand, resistance could be introduced to the ward by newly admitted patients who were already resistance carriers. These patients contributed to the reservoir of resistance carriers, independent of antibiotic treatment. Considering the whole ward, when the baseline prevalence of resistance carriers admitted to the ward was high, prolonged antibiotic treatment duration did not contribute many new resistance carriers, as many patients admitted were already carriers. This was also true when the baseline prevalence of resistance carriers admitted to the ward was negligible. Given that the main mechanisms in our models through which a patient became a resistance carrier were bystander selection and transmission (rather than *de novo* mutations), treatment duration could only contribute to new resistance carriers when there were existing resistant bacteria in the ward population.

It is worth noting that when administered antibiotics were active against the resistant organisms, there remained areas of parameter space where longer duration led to more resistance than shorter duration (red pixels in Figure 4.7, top row). This was observed at high resistance importation and transmission rates, especially for the treated patients and patients who were non-carriers at admission. This is because without bystander selection, prolonged antibiotic treatment increased the treated patients' risk of acquiring resistance. High numbers of imported carriers and transmission rates would therefore contribute to more resistance acquisition events in the long duration ward.

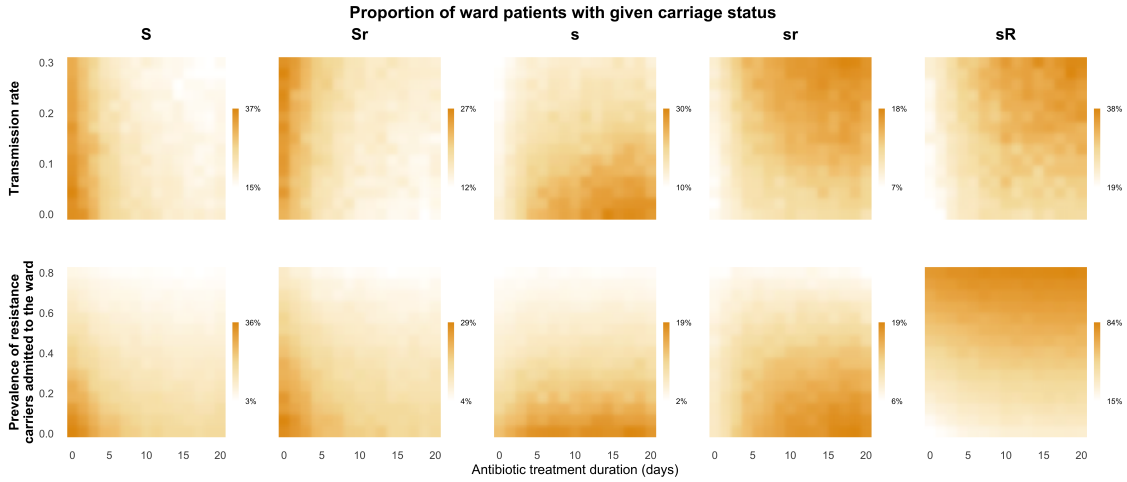


Figure 4.6: Effects of transmission rate and baseline prevalence of resistance carriers at admission on resistant bacteria carriage in the scenario where administered antibiotics have activity only against susceptible organisms I. Output from co-colonisation model under various parameter values (y-axis) for i) transmission rate and ii) prevalence of resistance carriers admitted to the ward with increasing antibiotic duration (x-axis). The coloured pixels indicate the proportion of ward patients (at equilibrium) with a given carriage status where *S/s* represents high/low proportions of susceptible bacteria and *R/r* represents high/low proportions of resistant bacteria. Increasing antibiotic duration increased the treated patients' risk of acquiring resistance carriage by suppressing the susceptible population. This resulted in shifts with increasing duration of treatment from *S* and *Sr* states to predominantly state *s* at low transmission rates and to predominantly states *sr* and *sR* at higher transmission rates.

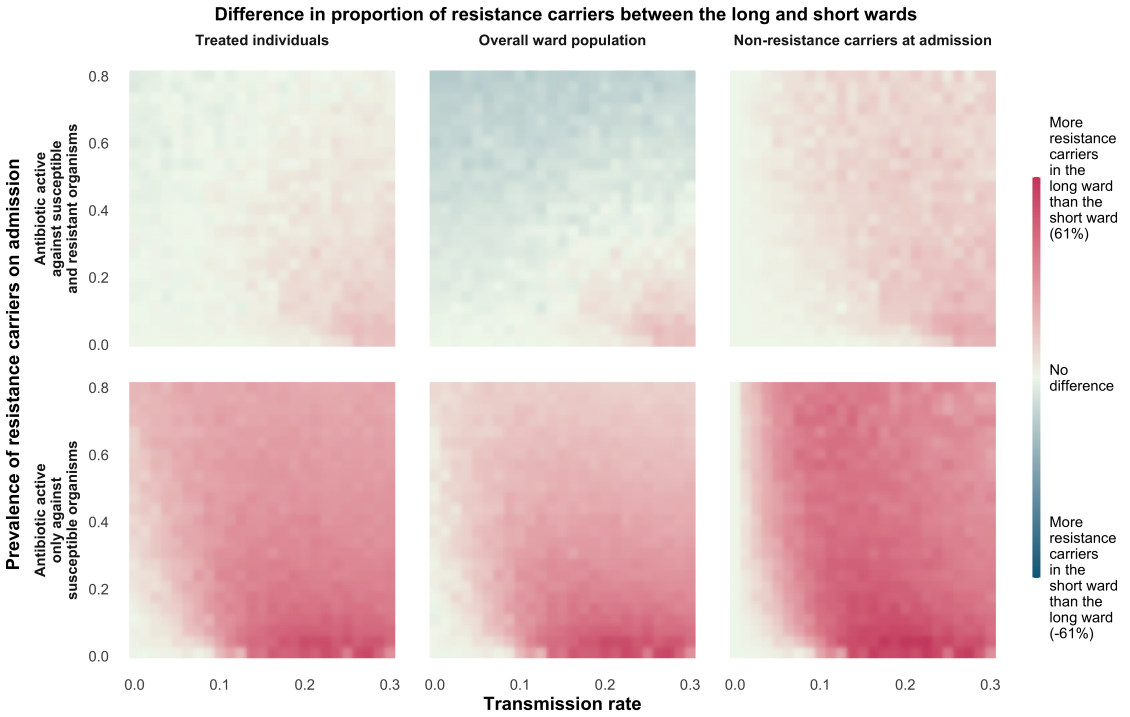


Figure 4.7: Effects of transmission rate and baseline prevalence of resistance carriers at admission on resistant bacteria carriage in the scenario where administered antibiotics have activity only against susceptible organisms II. Output from the exclusive colonisation model. i) Transmission rate (x-axis) and ii) prevalence of resistance carriers admitted to the ward (y-axis) were varied while other parameters were fixed to explore the difference in proportion of resistance carriers between the wards administered long and short antibiotic durations (coloured pixels). Red, white and blue pixels indicate respectively that there were more resistance carriers in the long ward (i.e. shortening treatment duration was effective at reducing resistance carriers); no difference between the short and long ward (i.e. shortening treatment duration had no effect on resistance carriers); and more resistance carriers in the short ward (i.e. prolonged treatment duration was effective at reducing resistance carriers). The top and bottom rows describe the scenarios where administered antibiotics have activity against both susceptible and resistant organisms and only susceptible organisms respectively. The columns represent the three populations in which the outcomes were assessed: i) treated individuals, ii) overall ward population and iii) non-carriers on admission.

4.3.5 Overall effect of antibiotic duration on resistance carriage

The models showed generally modest and highly heterogeneous effects of antibiotic duration on resistance carriage (Figure 4.8). When administered antibiotics have activity only against susceptible organisms, longer antibiotic duration resulted in higher prevalence of resistance carriage (Figure 4.8, right-side panels). In contrast, when administered antibiotics have activity against both resistant and susceptible organisms, longer antibiotic duration resulted in either an increase or decrease in the prevalence of resistance carriers depending on the parameter

space (Figure 4.8, right-side panels).

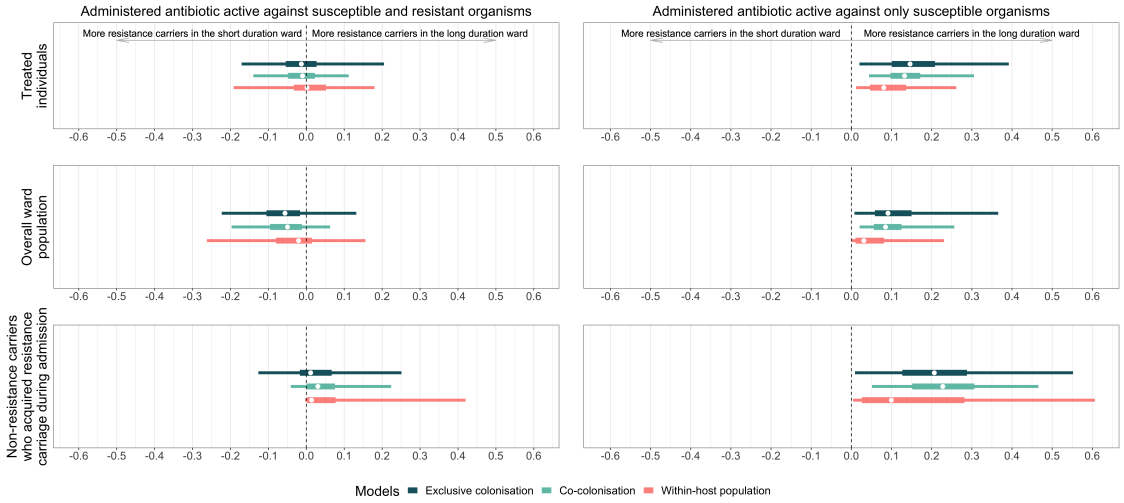


Figure 4.8: Difference in proportion of resistance carriers between wards administered long and short antibiotic duration (x-axis) under different antibiotic treatment. The three panels show the outcomes assessed: i) treated individuals, ii) overall ward population and iii) non-carriers on admission. Five hundred parameter combinations were used in the simulations sampled from uniform distributions (Table 4.1). For each sampled parameter combination, 100 iterations of the stochastic model were run, and the outputs were averaged. Each coloured bar represents the average output distribution for each set of parameter values. In each bar, the white dot represents the median and the thick and thin bars represent the 50% and 95% interval ranges.

4.3.6 Evidence from antibiotic duration randomised controlled trials

An initial search of the MEDLINE and EMBASE databases returned 2407 unique publications. Out of these, 187 were randomised trials which compared antibiotic treatment durations. Seven of these trials collected surveillance cultures for colonising bacteria during follow-up visits and were included in the qualitative synthesis (Figure 4.9). Complete data extracted from these randomised trials can be found at https://github.com/moyinNUHS/abxduration_abm. I did not find any meta-analysis in the literature that quantified the risk of colonisation by resistant bacteria with antibiotic duration.

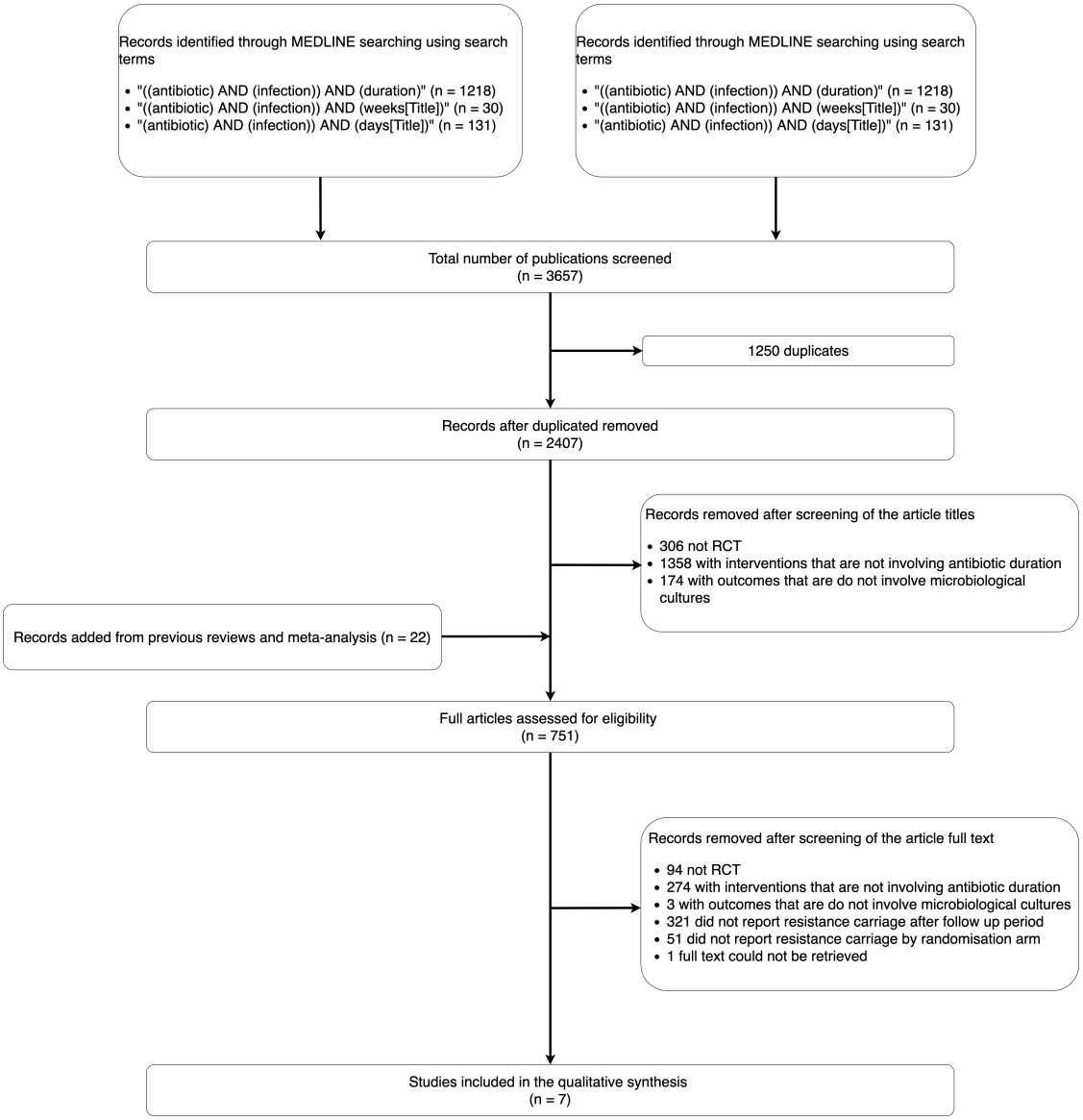


Figure 4.9: PRISMA flow diagram for systematic review on antibiotic duration trials which reported resistance carriage as an outcome. Note that reasons for record exclusion may overlap.

Ref	Indication for antibiotic treatment	Country	Healthcare setting	Age group	Antibiotic duration (days)		Antibiotic prescribed	Bacteria detected in surveillance culture	Colonisation site	Follow up (days)	Proportion of resistance carriers at end of study (resistance carriers/ participants who provided samples, %)		Difference between short vs long arm† (long - short, 95% CI)
					Short	Long					Short	Long	
Lutsar, 2020[191]	Neonatal late-onset sepsis	European countries	Inpatient	Children	7	8	Various	Carbapenem-resistant Gram-negative bacteria	Gut	28	5/70 (7)	5/124 (4)	−3 (−11, 5)
Hoberman, 2016[47]	Otitis media	USA	Outpatient	Children	5	10	Amxocillin/ clavulanic acid	Beta-lactamase producing *H influenzae*	Respiratory tract	14	28/222 (13)	38/233 (16)	4 (−3, 11)
Cera, 2010[192]	Uncomplicated urinary tract infection	Turkey	Outpatient	Adult	1	5	Fosfomycin vs ciprofloxacin	Fosfomycin-resistant Enterobacteriaceae	Urinary tract	7	7/77 (9)	12/65 (18)	9 (−3, 22)
Merode, 2005[193]	Uncomplicated urinary tract infection	Netherlands	Outpatient	Adult	3	5	Trimethroprim	Trimethroprim-resistant *E coli*	Urinary tract	7	8/64 (14)	9/60 (13)	2 (−11, 16)
Dow, 2004[194]	Urinary tract infection in patients with spinal cord injury	Canada	Inpatient	Adult	3	14	Ciprofloxacin	Fluroquinolone-resistant Gram-negative bacteria	Urinary tract	42	8/30 (27)	5/30 (17)	−10 (−34, 14)

Table 4.2: Details of the antibiotic trials included in the meta-analysis. †Difference between short and long arm was calculated with two-proportion Z-test.

Out of these seven trials, the indications for antibiotic treatments in six were for treatment of infections (urinary tract infections (3), otitis media (1), sepsis (1), acute respiratory illness (1)) and in one was for *Pseudomonas* spp. in non-cystic fibrosis bronchiectasis. Five were performed in outpatient settings with mean follow-up periods ranging from 3 to 365 days, two were in hospital settings with mean follow-up periods 28 and 41 days. Four trials enrolled only adults, while the other three enrolled only children as participants. All studies were carried out in high or upper-middle income countries. Antibiotic durations in the short arms ranged from 5 to 14 days with a median of 5 days, while in the long arms the durations ranged from 5 to 104 days with a median of 10 days (Figure 4.2).

A meta-regression analyses was performed using five out of the seven trials. One was excluded because the antibiotic duration in the long arm and the follow-up time were very prolonged compared to the other studies (104 days of treatment compared to 5 to 14 days, and 365 days follow-up compared to 3 to 41 days). Another was excluded because the study only collected Gram-positive bacteria in surveillance cultures. I considered resistance carriage when the antibiotic prescribed in the trial was not effective against the specific type of resistance e.g. ciprofloxacin was prescribed in a trial that monitored fluoroquinolone-resistance carriage.

The mixed-effect model (Model 2 in Box 4.2) was selected for the meta-analysis (Appendix Section C.2.2 Table C.2). The meta-analysis found the odds ratio of acquiring Gram-negative resistance carriage with one additional day of antibiotics to be 1.05 (80% credible interval 0.90 to 1.23%). This translates to a 4% increase in the daily probability of acquiring resistant carriage given a baseline daily probability of 0.1 (Figure 4.10). Sensitivity analyses using data only from studies with surveillance cultures collected up to 30 days after antibiotic treatment and using different priors produced similar results (Appendix Table C.3).

4.4 Discussion

In this study, I used three agent-based models with different assumptions to evaluate the efficacy of shortening antibiotic duration on resistant bacterial colonisation due to bystander selection from prescribed antibiotics. These three models reached broadly similar conclusions. Amongst the treated individuals, shortening antibiotic duration is most effective at reducing resistance carriage when antibiotic treatment is the main driver of the acquisition and subsequent maintenance of resistant bacteria carriage. This happens when bacterial growth and killing dynamics were rapidly responsive to antibiotic treatment, i.e. rapid resistant bacterial growth and killing of susceptible bacteria in the presence of antibiotics, and rapid regrowth of susceptible bacteria to outcompete resistant bacteria in the absence of antibiotics. When stopping antibiotics quickly leads to loss of the resistance strain, shortening duration is more likely to result in a substantial reduction in the prevalence of resistance carriers compared to a situation when colonisation persists months beyond

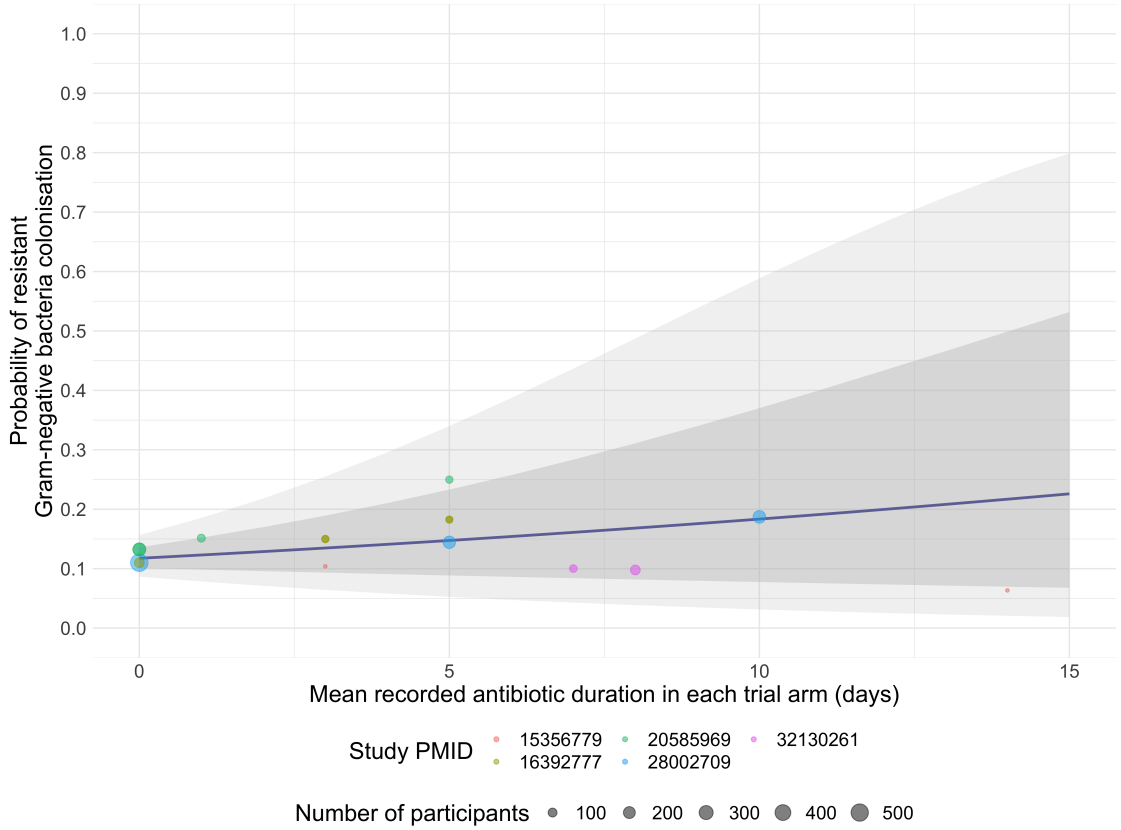


Figure 4.10: Daily risk of antibiotic-resistant Gram-negative bacteria carriage given days of antibiotics prescribed reported by the randomised controlled trials included in the meta-analysis. Daily probability of colonisation by antibiotic resistant Gram-negative bacteria (y-axis) is shown against mean duration of antibiotics (x-axis) reported in each trial. Each colour represents one trial. Each bubble represents a single arm in one trial, where the diameter of the bubble corresponds to the number of participants for the arm in the trial. The analysis allowed the relationship between colonisation and antibiotic duration to vary across each trial (Model 2 in Box 4.2); the blue line corresponds to the mean relationship between the two, considering all included trials. The light grey shaded areas are the associated 80% and 50% credible intervals.

discharge after stopping antibiotics.[168] At a population level, high transmission of resistant bacteria between hosts was a key factor in the efficacy of shortening treatment duration at reducing resistance carriage.

In addition, longer antibiotic treatment may decrease or increase the prevalence of resistance carriers, depending on the availability of effective antibiotics against particular resistance phenotypes. The meta-analysis, using data from randomised controlled trials, showed broad agreement with this finding from the model, and highlighted the increased incidence of resistance colonisation with treatment duration. The models also showed that even when antibiotics administered are active against a resistance phenotype, longer duration may increase resistance carriers, especially amongst treated individuals and non-carriers when the prevalence of resistance

carriers on admission and transmission rates are high.

There is an important caveat when comparing the model and the meta-analysis results. In the models, all patients who were prescribed antibiotics were administered a similar duration in the same ward and the models therefore captured both individual and population level effects of antibiotic selection pressure. However, the trials included in the meta-analysis were individually randomised, where most other individuals in contact with the study participants would not have been administered antibiotics. Hence, the meta-analysis reflected the direct individual antibiotic selection effects.

I did not find any cluster-randomised trials of antibiotic duration that monitored resistance colonisation at the population level. Such trials would have allowed us to compare the individual versus population effects from antibiotic selection. Since most resistance selection in common bacterial pathogens is due to bystander selection,[162] trials that only look at resistance in bacteria that are causing infections being treated are potentially missing a big part of the picture. Future trials should collect surveillance cultures to evaluate effect of duration on resistance carriage.

The modest effect of shortening antibiotic duration on reducing resistant bacteria colonisation can be partially explained by the patient and antibiotic prescribing characteristics unique to the hospital setting. These include relatively short length of stay and the slow decolonisation of resistant bacteria in those who were previously exposed to antibiotics.[167, 168] In this context, admitted patients under antibiotic treatment could potentially be quickly predisposed to acquire resistant bacteria carriage on treatment but remain as carriers during the relatively brief hospitalisation periods even after antibiotic treatment is stopped. This initial rapid increase in the risk of acquiring resistance bacteria carriage during the first few days of antibiotic treatment could reduce the effectiveness of shortening treatment duration at reducing resistance, especially with longer treatment durations.

The above findings highlight an important synergy between shortening antibiotic treatment and reducing transmission of resistant bacteria through infection prevention and control. This is intuitive because in addition to selecting for within-host resistance, antibiotic use also increases the risk of colonisation with resistant bacteria and subsequent prolonged colonisation.[167, 168] Shortening treatment duration has the potential to reduce a patient's risk of acquiring resistant bacteria, while infection prevention and control measures reduce transmission of resistant bacteria to these at-risk patients.

There are important limitations in this study. Even though a comprehensive systemic review was undertaken, some parameter values were not available in the literature or were taken from animal or *in vitro* experiments. For those parameter values not found in the literature, I explored the largest reasonable range of parameter values. As for the meta-analysis, there were few antibiotic duration trials which

monitored resistance carriage as an outcome. This has contributed to the large credible intervals in the daily resistance colonisation risk with antibiotic treatment, especially at longer treatment durations. These limitations highlight important gaps in the existing literature and the conduct of antibiotic duration trials. Lastly, our models did not account for *de novo* mutations during treatment, which are an important source of resistance for certain organisms, such as *Mycobacterium tuberculosis*.[195]

Understanding the key factors driving the variations in the antibiotic duration effect on resistance carriage will inform future research study designs, antimicrobial stewardship interventions and resource allocation in the overall set of control strategies. The practical implications from these findings are that interventions for shortening antibiotic treatment duration are potentially the most effective when antibiotics are stopped as early as feasible without compromising treatment success for the target pathogen, especially in high transmission settings. To ensure antibiotic treatment durations are minimised, using individualised antibiotic stopping rules that cater to individual response to treatment could be useful, especially when the diagnostic criteria is non-specific such as in catheter-associated urinary tract infections or ventilator-associated pneumonia.

5 Conclusion

Contents

5.1 Key points of the thesis

Treatment duration is a fundamental component of antibiotic prescribing. Optimal treatment duration is an intricate balancing act between maximising treatment success while minimising side effects and cost to the individual, as well as the emergence of AMR in the population. Most of the available literature on defining antibiotic treatment duration focuses on the first two outcomes, i.e. treatment success and antibiotic side effects to the treated individual, through randomised controlled trials.

There are unique considerations in the design of antibiotic duration trials. Central to this is that there should be a clear causal pathway between the disease, the treatment and the monitored outcome. Disease can be defined by various bacterial pathogens and/or sites of infection. The intervention for duration can be arbitrarily chosen or customised to individual responses to treatment by monitoring clinical or biomarker trends. Lastly, outcomes can be clinical or microbiological eradication. When there are factors causally linked with treatment adherence which also have a causal link with outcomes not mediated by treatment, without appropriate adjustment estimates of treatment effects will be biased. The effect of this in non-inferiority trials, the commonest trial hypothesis design in antibiotic duration trials, is that the probability of concluding non-inferiority when the shorter treatment regime may actually be inferior is increased. When deciding on the appropriate effect measure to determine non-inferiority, investigators should consider the context of the trial (i.e. consider whether the inclusion criteria, intervention, and outcome are generalisable to other settings) and also consider the perspective of the user. When the primary interest is effectiveness under real-world conditions, measures should be taken to ensure the reproducibility of the trial in other settings where the healthcare delivery system and practices are different.

In addition to the above design issues for non-inferiority antibiotic duration trials, there are also unique problems in the implementation of such trials. In particular, non-adherence is especially challenging to account for due to frequent direct cross-overs between long and short arms and difficulties in blinding treatment duration. By performing a simulation study, I found that the effect of allocation and effect of treatment differ with various patterns of non-adherence and analysis methods, and hence affect the determination of non-inferiority. With ITT analysis, if just 10% of participants cross over to the opposite arm the probability of claiming non-inferiority when the null hypothesis of inferiority is true can be inflated to as much as 10% compared to the nominal value of 2.5%. It is important to determine potential confounders during the trial design and to collect appropriate data from both adherent and non-adherent participants to adequately adjust for these factors. An adjusted PP analysis may be done as a primary analysis when treatment efficacy estimates are more pertinent, while it may be more appropriate as a supplementary analysis in other contexts where the ITT estimate is more relevant.

With the above in mind, I designed and implemented the REGARD-VAP trial in a

network of seven hospitals across Asia. The trial adopted the widely recognised and practised US CDC criteria to diagnose VAP. This set of criteria is based on a set of objective clinical features that can be easily reproduced even in most resource-limited settings. The intervention incorporated the patients' clinical response, allowing individualisation of treatment duration. The outcomes included both mortality and pneumonia recurrence at day 28 and 60. Careful steps were taken to blind the assessors when determining recurrence episodes. In addition, there were ample preparations to ensure adherence to intervention including a pilot study and continuous engagement with the local investigators and intensive care physicians. The ITT and PP study populations were pre-defined in the protocol. Potential confounders were identified, and adjustment will be done with inverse probability weighting using baseline patient characteristics as independent variables. To show non-inferiority, conclusions from all four adjusted and unadjusted ITT and PP analyses should agree.

The REGARD-VAP interim analysis of the data from 230 patients showed that more participants in the long antibiotic treatment arm died or suffered pneumonia recurrences within 60 days of enrolment. The upper bounds of the one-sided 95% confidence intervals for both the unadjusted and adjusted ITT and PP analyses were less than the pre-defined 12% non-inferiority margin. The associated Z-values of these analyses exceeded the Fleming-Harrington-O'Brien boundaries. This supported the alternative hypothesis that short duration is non-inferior to long duration, and the study termination conditions were met according to the initial data analysis plan. However the investigators, data safety and monitoring committee, and trial steering committee reached an agreement to continue the trial and change the trial design to hierarchical noninferiority-superiority. This allowed the trial to continue enrolment, given its potential to demonstrate superiority of the short duration. In addition, it would allow more patients to be enrolled from other sites to improve generalisability of the findings.

Despite reducing antibiotic treatment duration being widely adopted as a stewardship strategy, its effectiveness in reducing antimicrobial resistance is uncertain and a clear theoretical rationale of the approach is lacking. To investigate this, I constructed three increasingly complex stochastic models which incorporated both between- and within-host dynamics of susceptible and resistant bacteria in response to antibiotic exposure. Using parameters appropriate for Gram-negative bacteria, I performed a sensitivity analysis to understand under what circumstances shortening antibiotic duration would lead to the most reduction in resistance carriage. The three models reached broadly similar conclusions. Shortening antibiotic duration is most effective at reducing resistance carriage for treated individuals when there is rapid selective pressure for resistant bacteria during antibiotic treatment, and decolonisation when treatments are stopped. On a population level, a short duration strategy is most effective in high transmission settings. In addition, longer antibiotic treatment may decrease or increase the prevalence of resistance carriers depending on the availability of effective antibiotics targeting particular resistance mechanisms.

The meta-analysis found the odds ratio of acquiring Gram-negative resistance carriage with one additional day of antibiotics is 1.05 (80% credible interval 0.90 to 1.32%). This translates to a 4% increase in daily probability of acquiring resistant carriage given a baseline daily probability of 0.1.

5.2 Ongoing collaborations in the research trial network

The fundamental groundwork that enabled the REGARD-VAP trial was the establishment of an inclusive trial research network. This network is located in areas with disproportionately high AMR burden and costs, and presents opportunities to carry out further trials to optimise treatment for AMR infection and inform global policies. This is highly relevant, as a single centre or country may lack the sample size required to produce results with adequate power, while AMR hotspot areas may not have adequate expertise and resources to carry out large-scale trials.

To ensure that the local sites benefit from the collaboration and for sustainability of the network, numerous capacity-building activities were conducted during the REGARD-VAP trial. This included conference meetings, educational seminars, site visits, and language courses. These have led to other collaborative projects which are currently being conducted in the network (Figure 5.1).

5.2.1 Qualitative study

Through the REGARD-VAP trial, the site investigators and physicians raised numerous issues regarding local antibiotic prescribing practices. Some of these were the crux of why antibiotic stewardship programmes are yet to be established in their respective ICUs. This is in line with previous surveys that showed that fewer than two-thirds of the hospitals surveyed in Asia reported having ASPs.[196]

While REGARD-VAP measures the clinical effectiveness of shortening treatment duration for VAP, this quantitative evidence is only one of the numerous considerations a prescriber might have in a situation where antibiotics may be needed. A qualitative study will offer important insights into other determinants of prescribing behavior.

Hence, I conducted a qualitative study at the REGARD-VAP study sites to understand the implementation process of antibiotic stewardship interventions in the ICU, with a focus on duration of antibiotic courses. The results are currently being analysed. The study aims are to:

i. measure uptake and healthcare worker reactions to reducing antibiotic duration as an ASP intervention;

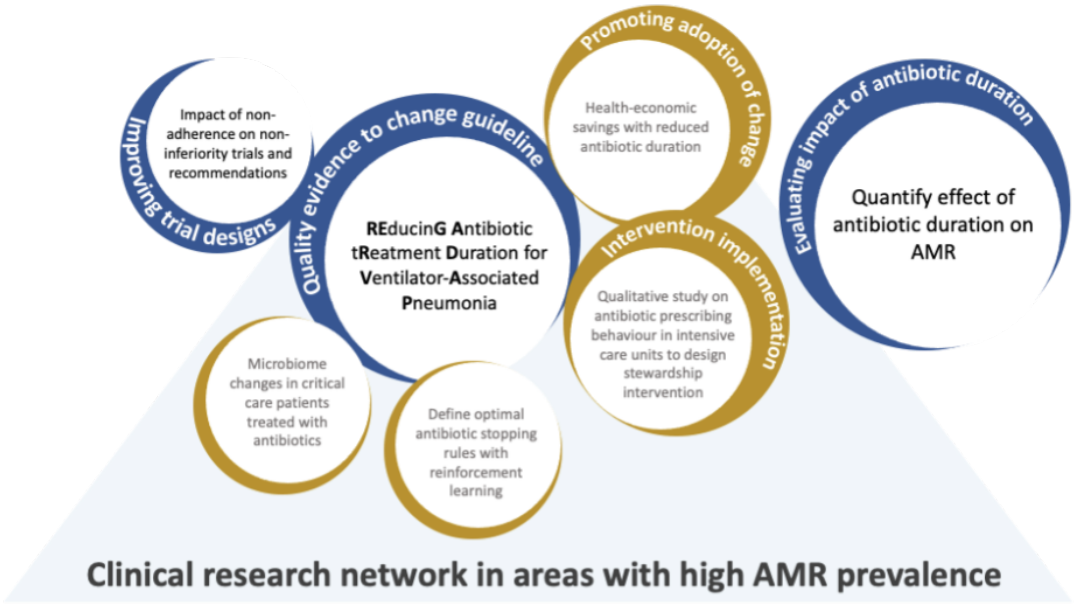


Figure 5.1: Ongoing projects within the REGARD-VAP network. Each bubble represents an ongoing project. Blue bubbles are the projects described in the thesis. Gold bubbles are the other ongoing projects that have received funding.

ii. investigate and describe decision pathways that determine/influence the choice and duration of antibiotics for common infections in the ICU;

iii. evaluate support and acceptance of current and potential future ASP interventions in the ICU; and,

iv. define the contextual determinants of ASP, and how to use this knowledge to develop sustainable interventions.

Since the beginning of 2019, all research staff at the respective study sites have been trained via online and in-person qualitative courses to conduct semi-structured interviews and ethnographies. Initial interviews were observed with immediate feedback to the research staff afterwards to improve on questioning techniques. To date, over 200 interviews have been completed, transcribed and translated. These scripts are currently being thematically coded for further analysis.

The initial findings highlight the following:

i) *Prescribing behaviour in the ICU*: When balancing the immediate concern of patients' illness with potential emergence of AMR infections in the future, ICU

physicians tend to prioritise the former and usually prescribe combinations of broad-spectrum antibiotics empirically. Antibiotics are deemed safe and often given as part of a treatment package for ICU patients either for suspected infections or prophylaxis. The power dynamics between the healthcare workers are highly hierarchical.

ii) *Interactions with the REGARD-VAP intervention*: Most welcome the intervention, and pointed out that protocolised antibiotic stopping rules can help to empower non-prescribers to participate in stewardship activities.

iii) *Barriers to antibiotic stewardship programmes*: Most healthcare workers are occupied with high workloads and do not have incentives to organise quality improvement programmes. Widespread lack of confidence in microbiology diagnostic capabilities, infection control programmes, and the quality of locally manufactured antibiotics was also expressed, especially in sites with fewer resources.

iv) *Opportunities for change*: There needs to be commitment at the ministry and hospital level to dedicate resources to stewardship programmes. Influential senior figures in the hospital can be sought to act as champions to lead changes in practice. A multi-pronged approach should be taken, which includes improving infection prevention and control strategies and microbiology laboratory standards. These are critical for building trust in stewardship programmes. To encourage these policy changes, research to quantify cost savings from stewardship programmes is urgently needed in order to engage health authorities and ministries.

5.2.2 Health economics modelling

As highlighted in the qualitative study above, healthcare costs are major concerns in both developed and developing countries. Economic evaluations of antimicrobials and interventions such as the stopping rules adopted in the REGARD-VAP trial require a balance between the quality-adjusted life years gained from successful treatment and the associated costs including AMR emergence, in order to maximise net health benefit.[197] Such work is challenging and thus not widely available in the current literature, especially in low-and-middle income countries. With our prospective trial data, we aim to quantify the cost of AMR associated with the consumption of each standard unit of the various classes of antibiotics. This will help guide target classes of antibiotics for control and monitoring, empirical antibiotic usage policies to minimize emergence of AMR, and prioritisation of infection control interventions in the ICUs.

5.2.3 Further analysis of REGARD-VAP trial data to individualise treatment duration

Individualisation is a key step in reducing antibiotic treatment duration, due to highly variable host-pathogen interactions. One of the most common concerns

expressed by clinicians when deciding when to stop antibiotics is that it 'depends on clinical progression'. While REGARD-VAP addresses this by using pre-defined stopping rules, it is difficult to identify the optimal stopping rules for individual patients from randomised controlled trials due to the sheer number of permutations of clinical parameters that can inform duration.

The REGARD-VAP trial provides a rich dataset consisting of daily hemodynamic parameters, laboratory test results, antibiotics prescribed, and detailed microbiological and clinical outcome assessments, in an ICU population that received various antibiotic durations. I aim to develop and pilot a novel reinforcement learning methodology to identify setting-specific antibiotic stopping rules. The method will have the potential to be extrapolated to larger observational datasets for optimal stopping rules in other common bacterial infections.

5.2.4 Mediating role of gut microbiome in the development of bloodstream infections post antibiotic treatment

Bloodstream infection (BSI) is one of the most common nosocomial infections in intensive care units, and is associated with high morbidity and mortality. Gut dysbiosis plays a critical role in the development of nosocomial BSI.[198] We collected multiple sputum and stool samples from the REGARD-VAP study participants from all study sites. With these samples I am conducting a pilot study that works towards defining the mediating role of the gut microbiome on the causal pathway between antibiotics and multidrug resistant nosocomial bloodstream infection. Over fifty samples from patients who had bloodstream infections and their matched controls post–antibiotic treatment have been sequenced using shotgun sequencing.

I hope to build on a recently developed dynamic Bayesian model to study changes in the human gut flora,[170] decomposing serial data into observational variation and ecological signal, enabling us to quantify the effects of specific classes of antibiotics on the emergence and persistence of resistance. This pilot will inform the final sample size required to make confident conclusions, and provide preliminary data for further grant applications and collaborations.

5.3 Other published works during the DPhil

I have also contributed to the following publications during the DPhil. Abstracts of selected works are included in Appendix Section D.1.

5.3.1 Publications as first author

1. Y Mo, DW Eyre, SF Lumley, TM Walker, RH Shaw, D O'Donnell, L Butcher, K Jeffery, CA Donnelly. Oxford COVID infection review team, Cooper BS. Transmission of community- and hospital-acquired SARS-CoV-2 in hospital settings in the UK: A cohort study. PLoS Med. 2021 Oct 12;18(10):e1003816.D.1.1

2. Y Mo, PA Tambyah, EN Perencevich. Infection, antibiotics, and patient outcomes in the intensive care unit. JAMA. 2020 Apr 21;323(15):1451-1452.

3. Y Mo, A Hernandez-Koutoucheva, P Musicha, D Bertrand, D Lye, O Ng, et al. Duration of carbapenemase-producing Enterobacteriaceae carriage in hospital patients. Emerg Infect Dis. 2020;26(9):2182-2185. D.1.2

4. Y Mo, CK Lee, TP Loh, ESC Koay, JW Tang, CC Lee. Next-generation sequencing identifies multi-drug resistant herpes simplex virus-associated scrotal ulceration. J Infect. 2020 Feb;80(2):232-254.

5. Y Mo, I Low, SK Tambyah, PA Tambyah. The socio-economic impact of multidrug-resistant nosocomial infections: A qualitative study. Journal of Hospital Infection. 2019 Aug;102(4):454-460.

6. Y Mo, I Seah, PSP Lye, et al. Relating knowledge, attitude and practice of antibiotic use to extended-spectrum beta-lactamase-producing Enterobacteriaceae carriage: Results of a cross-sectional community survey. BMJ Open. 2019 Mar 5;9(3):e023859.

5.3.2 Publications as co-author

1. C Lim, Y Mo, P Teparrukkul, M Hongsuwan, NPJ Day, D Limmathurotsakul, BS Cooper. Effect of delays in concordant antibiotic treatment on mortality in patients with hospital-acquired *Acinetobacter* spp. bacteremia: Emulating a target randomised trial with a 13-year retrospective cohort. Am J Epidemiol. 2021 May 27:kwab158.

2. K Marimuthu, Y Mo, ML Ling et al. Household transmission of carbapenemase-producing Enterobacteriaceae (CaPES-C): A prospective cohort study. J Antimicrob Chemother. 2021 Apr 13;76(5):1299-1302.

3. A Henderson, DL Paterson, MD Chatfield, PA Tambyah, DC Lye, PP De, RTP Lin, KL Chew, Y Mo. MERINO Trial Investigators and the

Australasian Society for Infectious Disease Clinical Research Network (ASID-CRN). Association between minimum inhibitory concentration, beta-lactamase genes and mortality for patients treated with piperacillin/tazobactam or meropenem from the MERINO study. Clin Infect Dis. 2020 Oct 27:ciaa1479.

4. SC Ong, JX Yap, TYF Tay, Y Mo, SC Loon, V Koh. Considerations in the use of slit lamp shields to reduce the risk of respiratory virus transmission in coronavirus disease 2019. Curr Opin Ophthalmol. 2020 Sep;31(5):374-379.

5. R Niehus, E Van Kleef, Y Mo, A Turlej-Rogacka, C Lammens, Y Carmeli, H Goossens, E Tacconelli, B Carevic, S Malhotra-Kumar, BS Cooper. Quantifying antibiotic impact on within-patient dynamics of extended-spectrum beta-lactamase resistance. Elife. 2020 May 7;9:e49206.

6. KB Pouwels, Y Mo, CC Butler, BS Cooper, S Wordsworth, AS Walker, JV Robotham. Optimising trial designs to identify appropriate antibiotic treatment durations. BMC Med. 2019 Jun 21;17(1):115.

7. PNA Harris, PA Tambyah, DC Lye, Y Mo. MERINO Trial Investigators and the Australasian Society for Infectious Disease Clinical Research Network (ASID-CRN). Effect of piperacillin-tazobactam vs meropenem on 30-Day mortality for patients With *E coli* or *Klebsiella pneumoniae* bloodstream infection and ceftriaxone resistance: A randomized clinical trial. JAMA. 2018 Sep 11;320(10):984-994.

8. QX Ng, W Loke, NX Foo, Y Mo, WS Yeo, AYS Soh. A systematic review of the use of rifaximin for *Clostridium difficile* infections. Anaerobe. 2018 Nov 2;55:35-39.

Appendices

A

Statistical considerations in non-inferiority trials with non-adherence

Contents

A.1 Effect estimate and type 1 error graphs for all simulated scenarios.

Drivers of non-adherence	Non-adherent population	Actual treatment received	Influence of increasing confounder values on adherence and outcome	Figure
Non-confounding factors	Both groups	Cross-over	NA	Figure A.1/ main text figure 2.5
		Inferior to experimental and control treatments		Figure A.2
	Experimental group	Cross-over		Figure A.3
		Inferior to experimental and control treatments		Figure A.4
	Control group	Cross-over		Figure A.5
		Inferior to experimental and control treatments		Figure A.6
Confounding factors	Both groups	Cross-over	Chance of treatment failure is increased Chance of taking up experimental treatment is decreased in both the experimental and control groups Chance of taking up inferior alternative treatment is increased in the experimental group and decreased in the control group	Figure A.7/ main text figure 2.6
		Inferior to experimental and control treatments		Figure A.8
	Experimental group	Cross-over		Figure A.9
		Inferior to experimental and control treatments		Figure A.10
	Control group	Cross-over		Figure A.11
		Inferior to experimental and control treatments		Figure A.12
	Both groups	Cross-over	Chance of treatment failure is increased Chance of taking up experimental treatment is increased in both the experimental and control groups Chance of taking up inferior alternative treatment is decreased in the experimental group and increased in the control group	Figure A.13/ main text figure 2.8
		Inferior to experimental and control treatments		Figure A.14
	Experimental group	Cross-over		Figure A.15
		Inferior to experimental and control treatments		Figure A.16
	Control group	Cross-over		Figure A.17
		Inferior to experimental and control treatments		Figure A.18

Table A.1: Index of effect estimate and type 1 error graphs for all simulated scenarios.

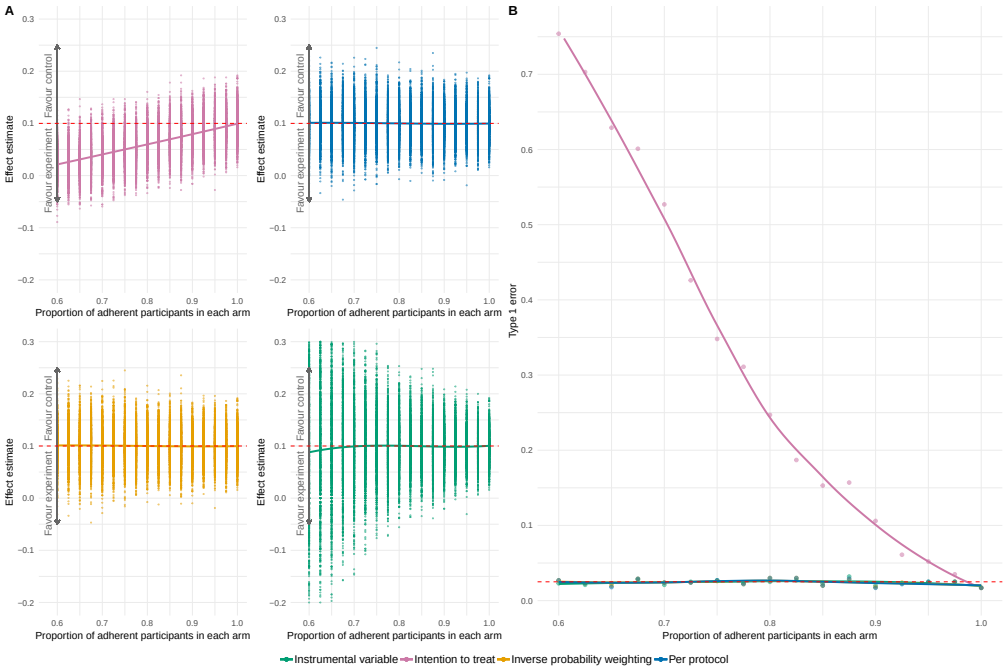


Figure A.1: Simulation scenario 1. Non-adherence due to non-confounding factors, where non-adherent participants from both groups cross over to the opposite arm.

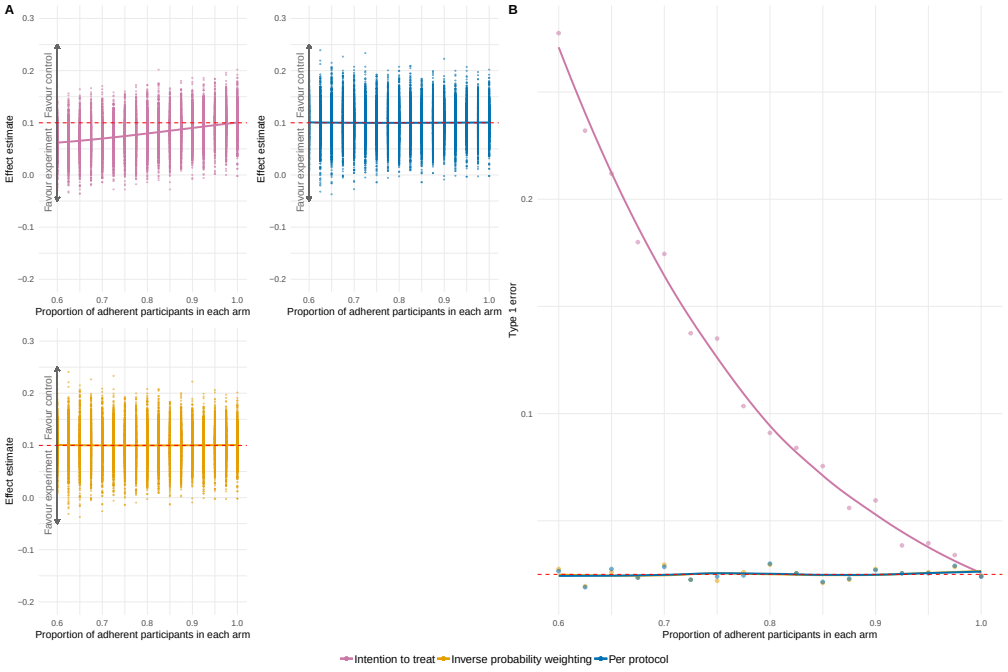


Figure A.2: Simulation scenario 2. Non-adherence due to non-confounding factors, where non-adherent participants from both groups receive inferior treatments compared to the experimental and control treatments.

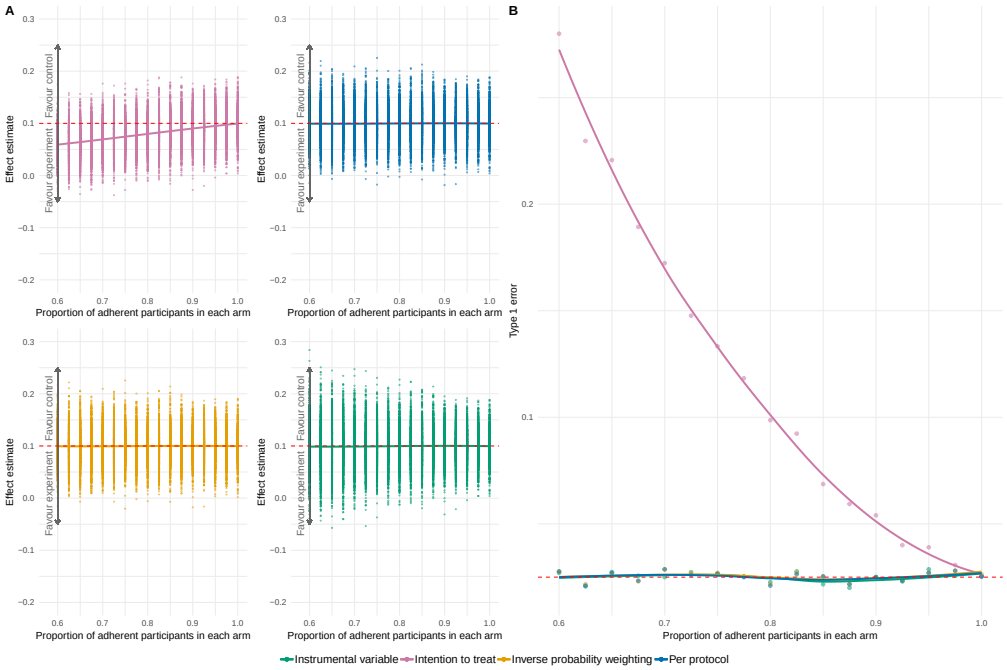


Figure A.3: Simulation scenario 3. Non-adherence due to non-confounding factors, where non-adherent participants from the experimental group cross over to the control treatment group.

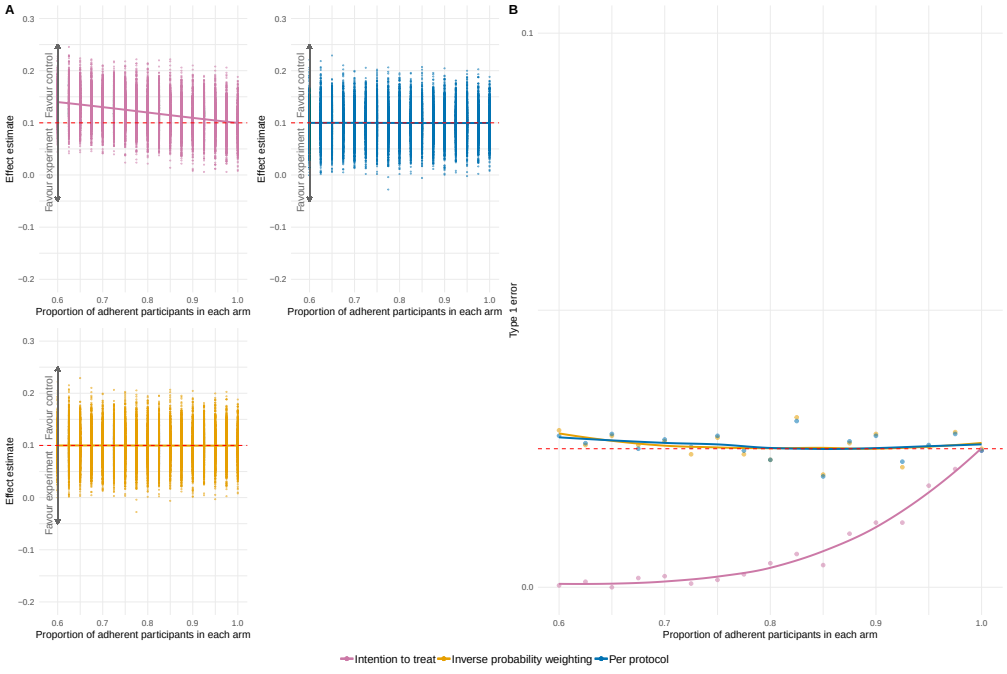


Figure A.4: Simulation scenario 4. Non-adherence due to non-confounding factors, where non-adherent participants from the experimental group receive inferior treatments compared to the experimental and control treatments.

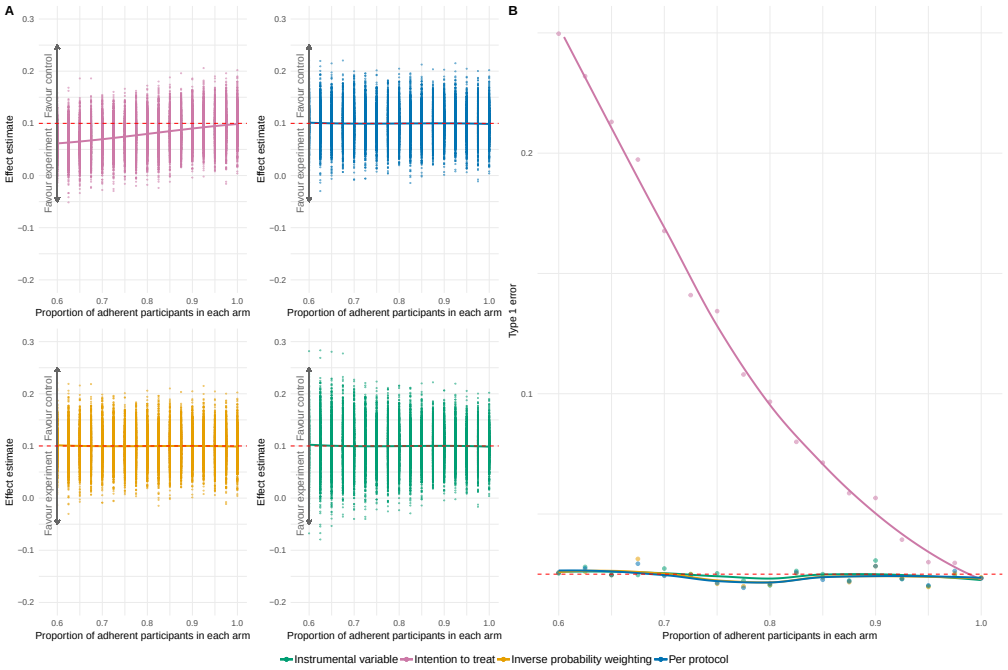


Figure A.5: Simulation scenario 5. Non-adherence due to non-confounding factors, where non-adherent participants from the control group cross over to the experimental treatment group.

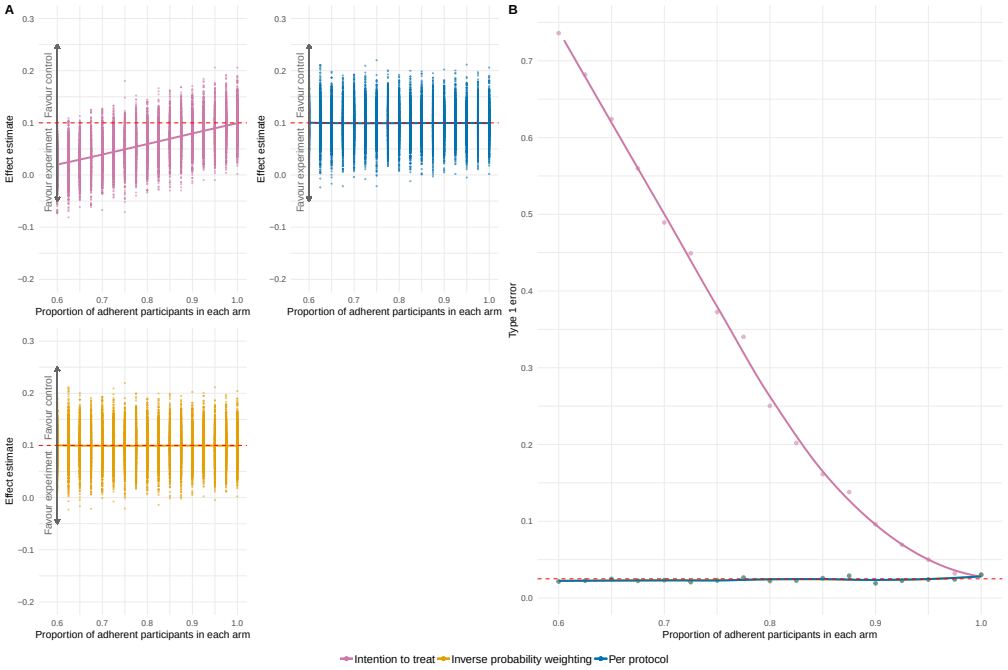


Figure A.6: Simulation scenario 6. Non-adherence due to non-confounding factors, where non-adherent participants from the control group receive inferior treatments compared to the experimental and control treatments.

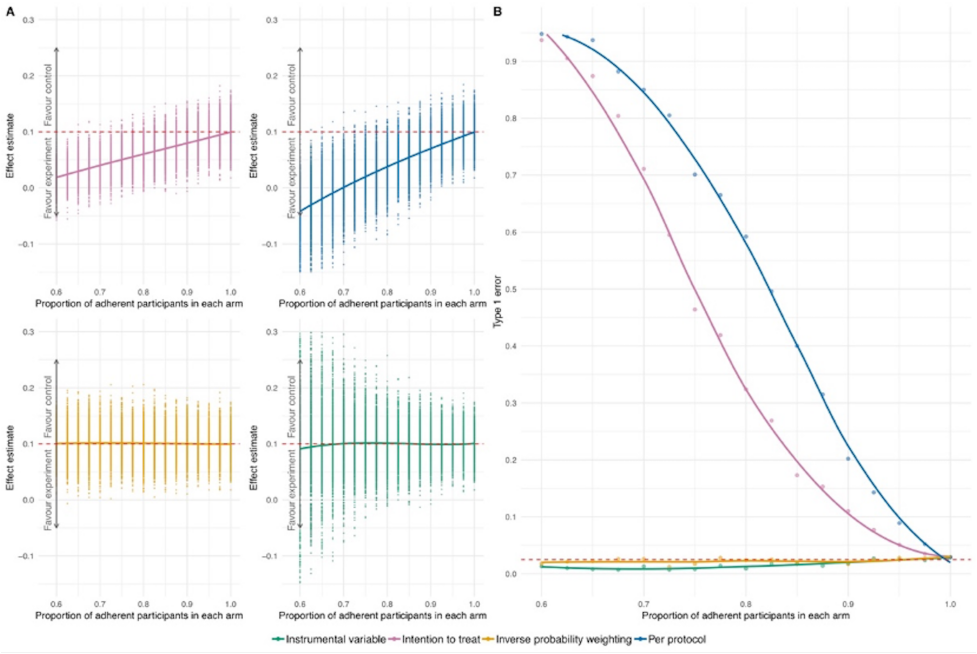


Figure A.7: Simulation scenario 7. Non-adherence due to confounding factors, where non-adherent participants from both groups cross over to the opposite arm. Participants with higher confounder values have lower probability of taking up the experimental treatment regardless of the allocation, and increases the probability of treatment failure.

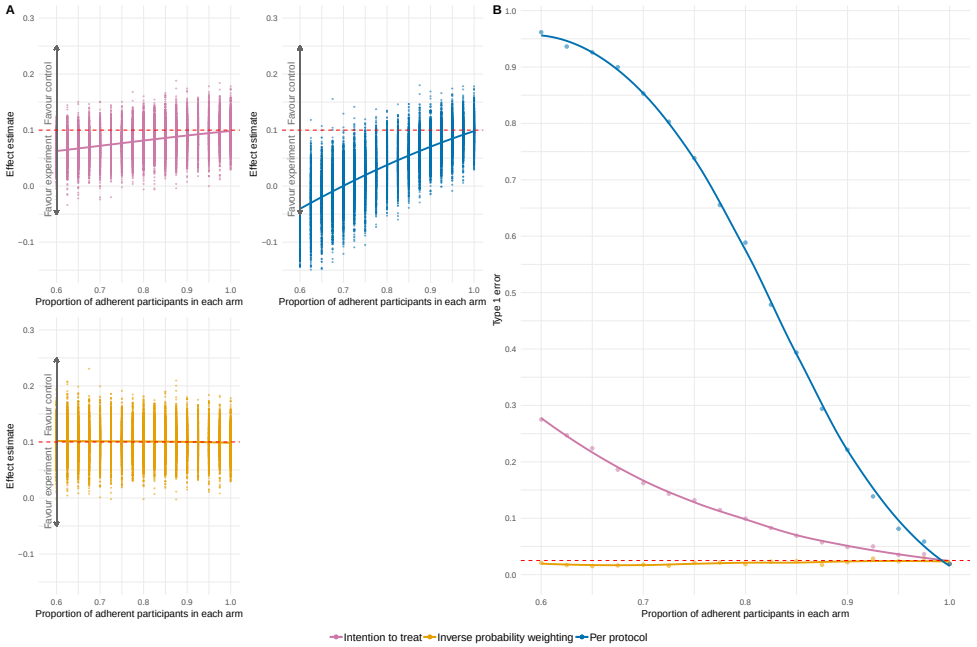


Figure A.8: Simulation scenario 8. Non-adherence due to confounding factors, where non-adherent participants from both groups receive inferior treatments compared to the experimental and control treatments. Participants with higher confounder values in the experimental arm and lower confounder values in the control arm are more likely to be non-adherent. Higher confounder values increases the probability of treatment failure.

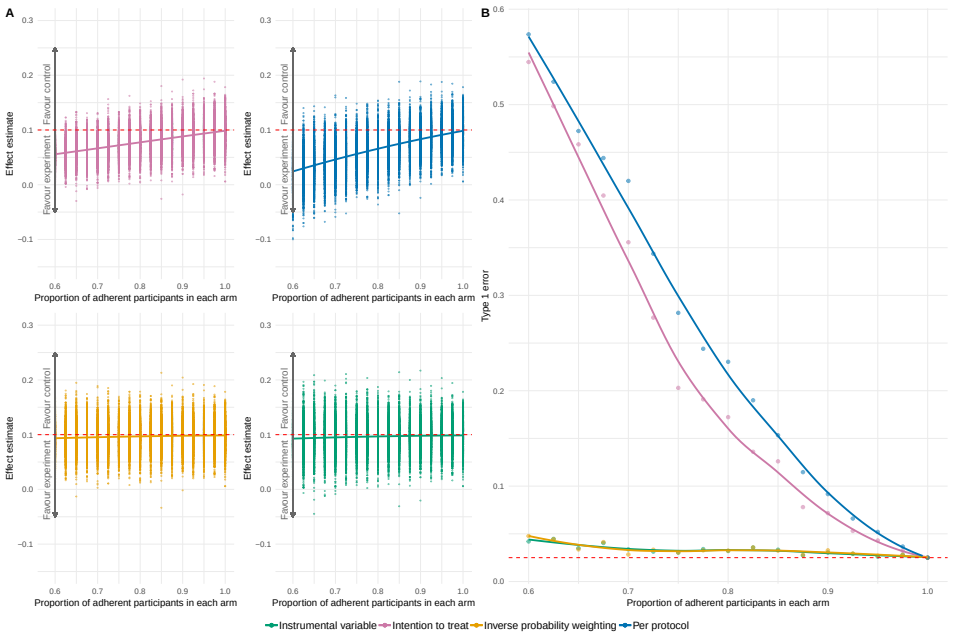


Figure A.9: Simulation scenario 9. Non-adherence due to confounding factors, where non-adherent participants from the experimental group cross over to the control treatment group. Participants with higher confounder values have lower probability of taking up the experimental treatment regardless of the allocation, and increases the probability of treatment failure.

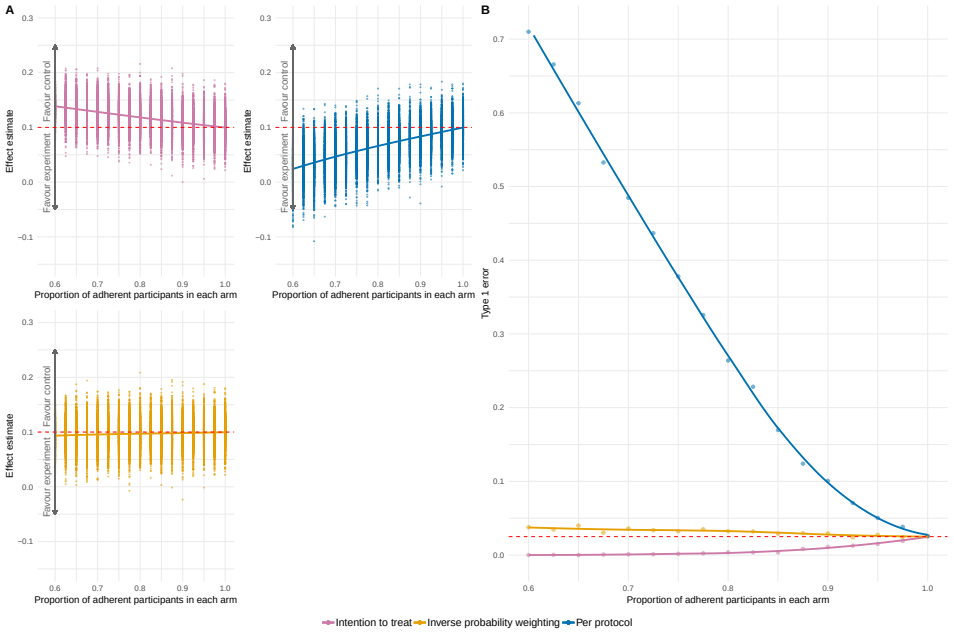


Figure A.10: Simulation scenario 10. Non-adherence due to confounding factors, where non-adherent participants from the experimental group receive inferior treatments compared to the experimental and control treatments. Participants with higher confounder values in the experimental arm and lower confounder values in the control arm are more likely to be non-adherent. Higher confounder values increases the probability of treatment failure.

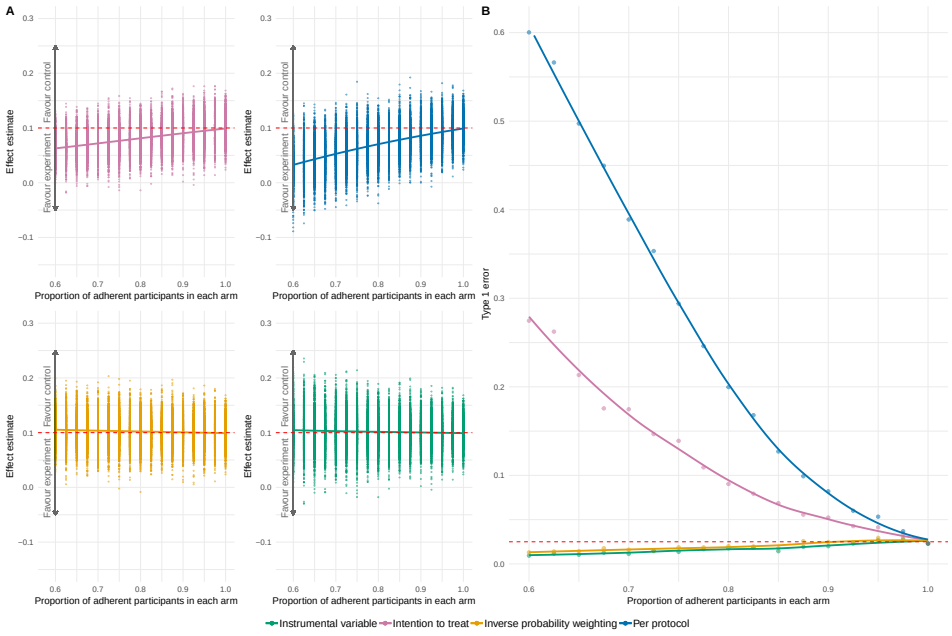


Figure A.11: Simulation scenario 11. Non-adherence due to confounding factors, where non-adherent participants from the control group cross over to the experimental treatment group. Participants with higher confounder values have lower probability of taking up the experimental treatment regardless of the allocation, and increases the probability of treatment failure.

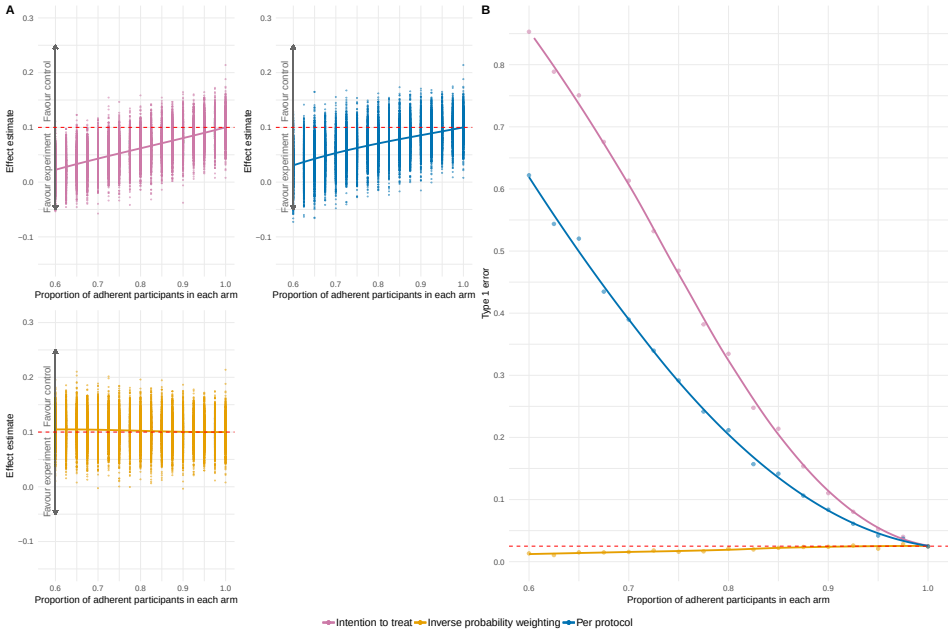


Figure A.12: Simulation scenario 12. Non-adherence due to confounding factors, where non-adherent participants from the control group receive inferior treatments compared to the experimental and control treatments. Participants with higher confounder values in the experimental arm and lower confounder values in the control arm are more likely to be non-adherent. Higher confounder values increases the probability of treatment failure.

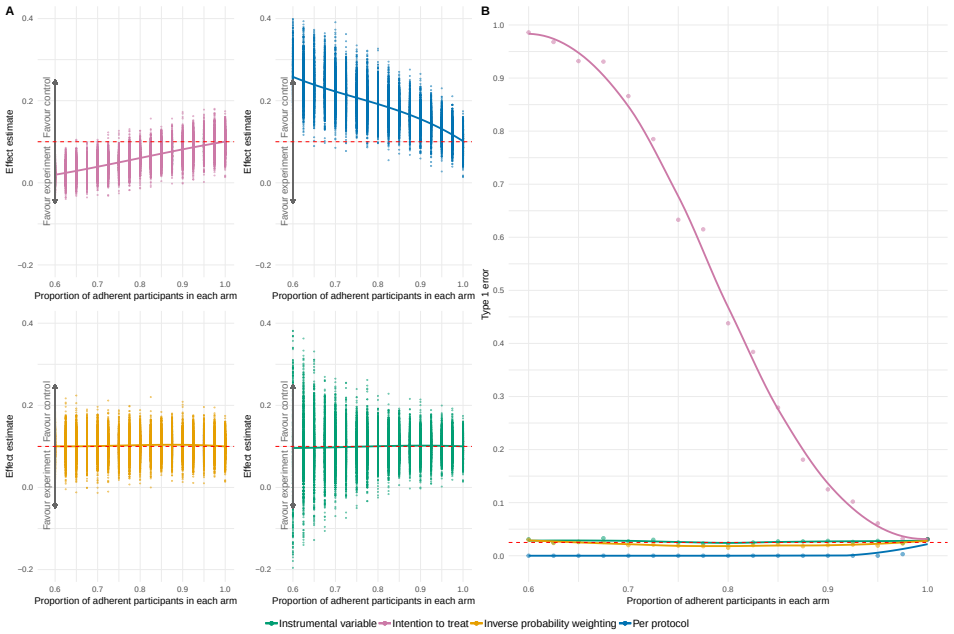


Figure A.13: Simulation scenario 13. Non-adherence due to confounding factors, where non-adherent participants from both groups cross over to the opposite arm. Participants with higher confounder values have higher probabilities of taking up the experimental treatment regardless of the allocation, and treatment failure.

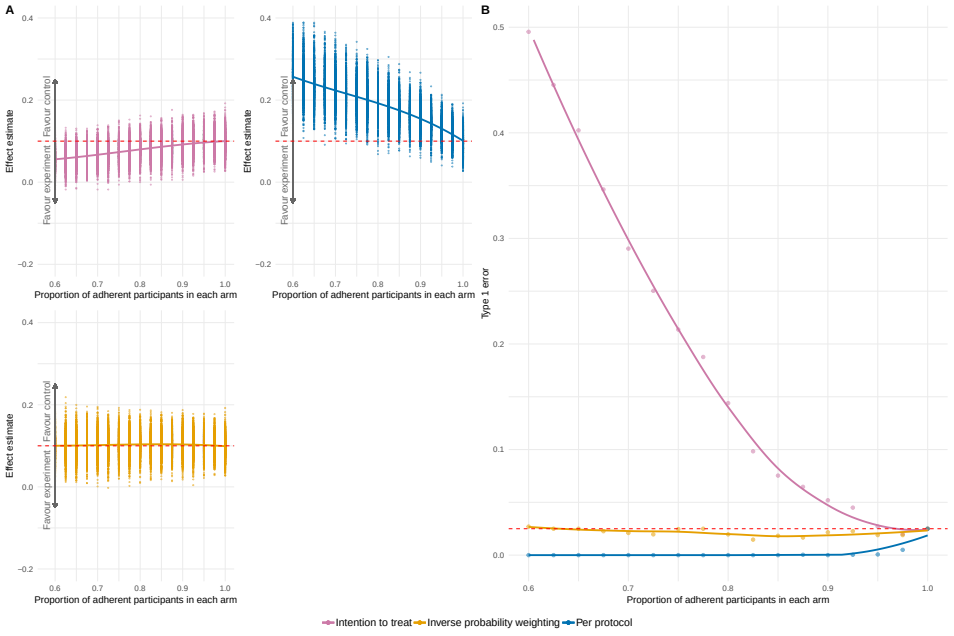


Figure A.14: Simulation scenario 14. Non-adherence due to confounding factors, where non-adherent participants from the both groups receive inferior treatments compared to the experimental and control treatments. Participants with lower confounder values in the experimental arm and higher confounder values in the control arm are more likely to be non-adherent. Higher confounder values increases the probability of treatment failure.

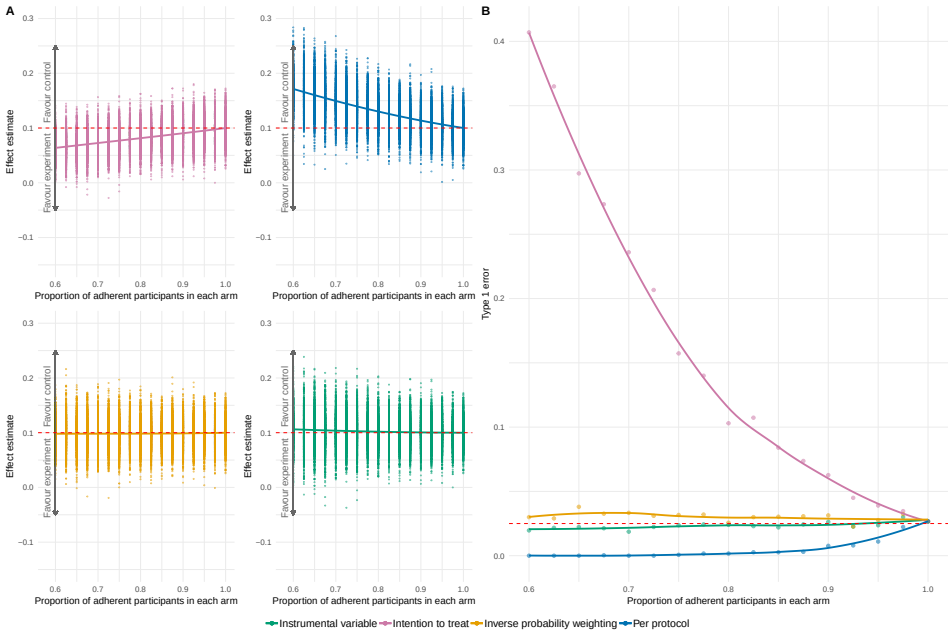


Figure A.15: Simulation scenario 15. Non-adherence due to confounding factors, where non-adherent participants from the experimental group cross over to the control treatment group. Participants with higher confounder values have higher probabilities of taking up the experimental treatment regardless of the allocation, and treatment failure.

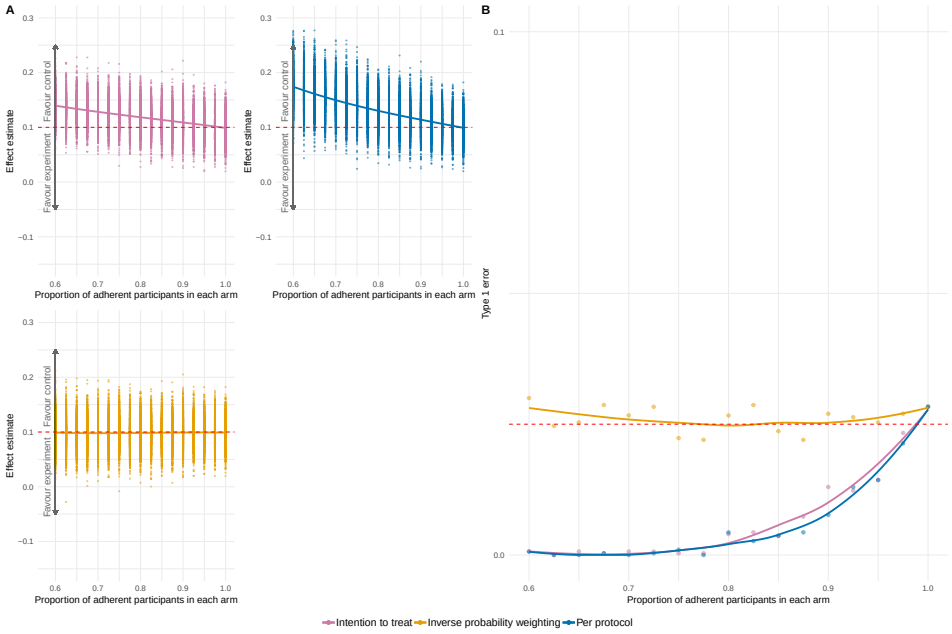


Figure A.16: Simulation scenario 16. Non-adherence due to confounding factors, where non-adherent participants from the experimental group receive inferior treatments compared to the experimental and control treatments. Participants with lower confounder values in the experimental arm and higher confounder values in the control arm are more likely to be non-adherent. Higher confounder values increases the probability of treatment failure.

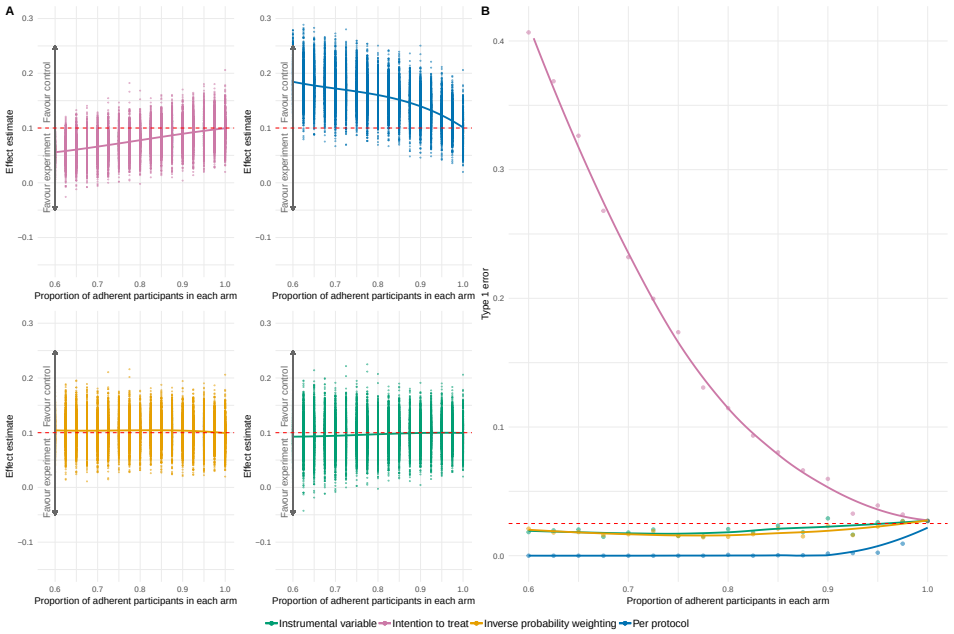


Figure A.17: Simulation scenario 17. Non-adherence due to confounding factors, where non-adherent participants from the control group cross over to the experimental treatment group. Participants with higher confounder values have higher probabilities of taking up the experimental treatment regardless of the allocation, and treatment failure.

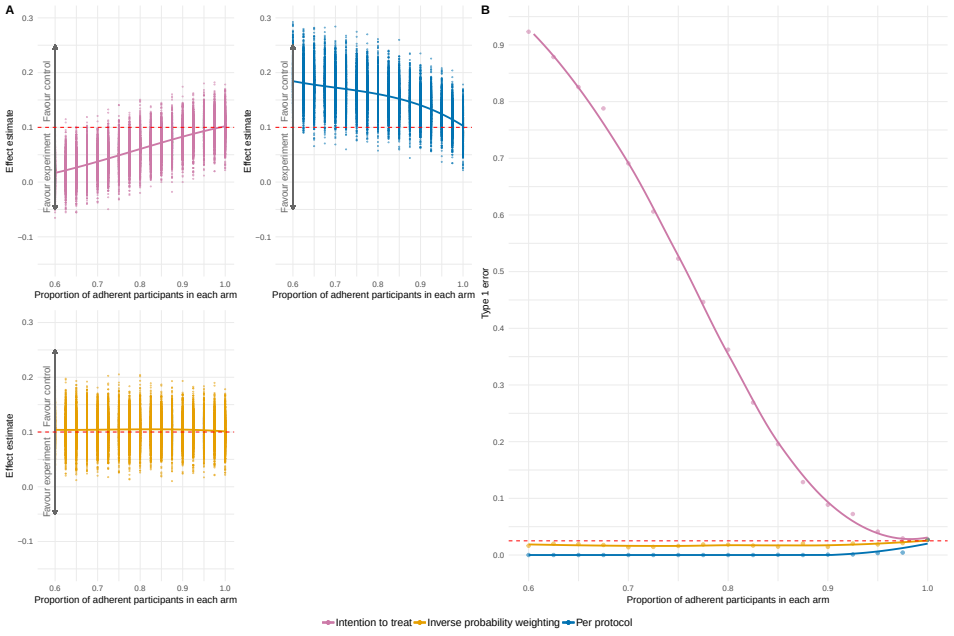


Figure A.18: Simulation scenario 18. Non-adherence due to confounding factors, where non-adherent participants from the control group receive inferior treatments compared to the experimental and control treatments. Participants with lower confounder values in the experimental arm and higher confounder values in the control arm are more likely to be non-adherent. Higher confounder values increases the probability of treatment failure.

B

REducinG Antibiotics tReatment Duration for Ventilator-Associated Pneumonia (REGARD-VAP)

Contents

B.1 Full study protocol version 5.1

B.1.1 Synopsis

Study Title	REducinG Antibiotics tReatment Duration for Ventilator-Associated Pneumonia (REGARD-VAP)	
Protocol no.	BAC17008	
Study Design	Randomised, partially double-blind (blinding may be achieved up to 7 days) controlled trial to assess the non-inferiority, followed by superiority, of a short duration of antibiotics (up to 7 days) versus prolonged antibiotic therapy (as per physician preference) in adult patients with VAP in Asia	
Study Participants	Adult patients (≥18 years old) who satisfy inclusion and exclusion criteria for VAP	
Planned Sample Size	460	
Planned Study Period	2018 January – 2022 October	
	Objectives	**Outcome Measures**
Primary	To demonstrate clinical non-inferiority, followed by superiority, of a shorter antibiotic treatment duration compared to a longer duration in adult patients with VAP	Difference in the proportion of participants with the composite endpoint of death and VAP recurrence rate within 60(±5) days of enrolment
Secondary	To define other clinical benefits of shorter antibiotic treatment duration compared to a longer duration in adult patients with VAP	Ventilator-associated events, duration of mechanical ventilation, duration of hospitalisation, acquisition of multidrug resistant organism (MDRO) infection or colonisation during the hospitalisation, number of days of exposure to antibiotics during the hospitalisation, number and types of extrapulmonary infections during hospitalisation (determined from cultures taken from sterile sites)
	To study the impact of various antibiotic regimen on the microbial community dynamics and diversity in the respiratory and intestinal tracts in individual patients in order to understand colonisation and transmission of MDROs, identify potentially protective components of the microbiome against MDROs	Characteristics of the microbiota will be determined by comparing alpha and beta diversity metrics between the groups of patients, alterations in the repertoire of antibiotic resistant genes, shifts in microbiota functional and metabolic capacity
	To study the impact of antibiotic utilisation on the ward-level multidrug resistant organism emergence and transmission dynamics	Prospective audit of antibiotic utilisation of participating wards, colonising bacteria in respiratory and stool samples, antibiograms of infective and colonising bacteria, whole genome and resistance genes sequencing of multidrug resistant isolates, mathematical modelling using sequencing data to track the spread of pathogens between hosts allowing for unobserved infection times, multiple independent introductions of the pathogen, and within-host genetic diversity
	To construct predictive models with comprehensive clinical and molecular epidemiological data to produce setting-specific and evidence based antibiotic policies	Statistical models and patient screening algorithms in both low-to- middle and high income settings to predict outcomes including proportion of adequate empiric therapy, expected deaths
	To estimate the changes to total costs and health benefits measured by Quality Adjusted Life Years (QALY) from implementing into practice the shorter antibiotic treatment duration	Resources and costs used for clinical implementation of a new policy; changes to costs arising from new policy as shown by length of stay on the ICU and the ward change; antibiotics use; other consumables; tests and diagnostics Risks of adverse events associated with change in duration of ICU stay, particularly other hospital-acquired infections (HAIs). Impact on mortality risk and health related QoL from shortening length of stays in ICU and reducing incidences of other HAIs

B.1.2 Background and rationale

Hospitals and, in particular, intensive care units (ICUs) are epicenters for the emergence and dissemination of multidrug resistant (MDR) bacteria. ICUs have the most vulnerable patient population due to frequent use of invasive devices which break anatomical barriers, bypass host defences and distort normal protective microbiomes, because of high antimicrobial consumption. Consequently, the ICU population has one of the highest rates of nosocomial infections leading to a significant impact on healthcare costs, morbidity and survival.[199] With their controlled settings and well-defined patient populations, ICUs present an opportunity to conduct robust interventional and epidemiological studies on the effect of antimicrobial use on the emergence and spread of MDROs in individual patients as well as on the overall ecology in the ICU environment. Carefully conducted regional multi-centre trials in ICUs will provide important results that can potentially be extrapolated to larger clinical settings via epidemiological modelling to estimate the impact of the measures studied if these are well adopted for the whole Asian region.

VAP is the most common nosocomial infection in patients in ICUs. As demonstrated in a cross-sectional prevalence survey in 1,417 ICUs worldwide, the prevalence of respiratory tract infection was 64% among all patients with nosocomial infections.[112] Estimates of all- cause mortality in patients with VAP range from 20-50%,[40, 113] and can be as high as 94% in low- to middle-income countries.[114] VAP prolongs the length of mechanical ventilation by 7.6-11.5 days and prolongs hospitalisation by 11.5 - 13.1 days compared with similar patients without VAP, and the excess cost associated with VAP has been estimated at USD$40,000 per patient in the United States.[200, 201] Given its high prevalence and associated antibiotic usage, VAP is likely to be a key driver of antimicrobial resistance (AMR) in ICUs.

Over the last few decades, we have continued to rely on the standard clinical, radiographic, and microbiological criteria with a low sensitivity and specificity of between 70 and 75% to diagnose VAP, respectively.[115] Identification of specific causative organisms is also difficult as the respiratory tract is non-sterile. Concordance between tracheal non-quantitative cultures and cultures of lung tissue from open lung biopsy was found to be as low as 40%.[116] These factors result in over-diagnosis and over-treatment of organisms thought to be causing VAP with empirical combinations of broad-spectrum antibiotics. For those patients who are on culture-directed definitive antibiotics, duration of treatment remains controversial. There are two notable clinical trials from France that have suggested that a short course of 8 days has comparable clinical efficacy as a long duration of 15 days.[45, 50] However, these studies could not confidently conclude that the finding can be applied to VAP caused by non-fermenting Gram-negative bacilli due to increased recurrence in such patients. Important potential biases also exist in those studies. For example: long-course therapy is favoured due to the differential time period during which recurrence was assessed, persistent colonisation could have been erroneously classified as recurrent infections, and the observed non-significant

high rate of recurrence in those caused by non-fermenting Gram-negative bacilli could be a spurious finding due to subgroup over-analyses.[45, 50] This severely limits the applicability of the current recommendation in Asia. This is because the majority of VAP in this region are caused by non-fermenting Gram-negative bacilli such as *Pseudomonas aeruginosa* and *Acinetobacter baumannii*.[121, 137, 202, 203] The current median number of days of antibiotic treatment remains at 12-13 days in Thailand.[122, 123] Currently available studies also suffer from narrowly defined inclusion criteria that excluded culture-negative VAP. This limits generalisability of the findings to VAP caused by unknown organisms. Therefore, it is crucial to define whether a short duration of antimicrobial therapy for VAP would provide equivalent clinical outcome and yield lower emergence of MDRO in Asian settings.

Because of the uncertainty about the exact causative organisms, ICUs are likely to utilise a high proportion of inappropriate empirical broad-spectrum antibiotics,[123] which might contribute to the indirect 'collateral damage' of AMR emergence and dissemination to other patients. In Thailand, an alarming 95% of *A. baumannii* and 75% of *P. aeruginosa* isolates were recently found to be resistant to carbapenems, and 47% of *K. pneumoniae* isolates were ESBL-producing in VAP.[204] In Singapore, 91% of *A. baumannii* and 22% of all Gram-negative isolates collected during a 4-month period in 2010 were resistant to carbapenems in the "Comparative Activity of Carbapenem Testing" (COMPACT) study.[205] This emphasizes the importance of antimicrobial stewardship programmes (ASPs) which have shown to be capable of reducing broad-spectrum antibiotic consumption and potentially decreasing antibiotic resistance, particularly among Gram-negative pathogens.[206–208] However, current recommendations on ASP are based on very few conclusive studies as most ASP research was designed to evaluate cost-effectiveness instead of clinical outcomes such as reducing selection of MDR pathogens, mortality, hospital length of stay, and readmission rates.[209–211] In addition, operational delivery of ASPs is often hindered by cost, lack of diagnostics and data collection facilities, poor physician participation and lack of cooperative strategies.[196] Low uptake of ASPs, especially in SEA,[212] calls for more rigorous and methodologically sound evaluations of ASP interventions to initiate practice and policy changes. Mathematical models offer tremendous potential for assessing the effectiveness of ASPs by removing the limitations inherent in human experimentation, including potential health risks, study cohort size and duration.[124] To date, the potential of modelling in evaluating ASPs in control of antimicrobial resistance is largely untapped and much work remains to be done to leverage this potential.

We aim to rationalise and reduce antibiotic consumption in ICUs and assess the benefit of this on both the individual patient and the overall cohort of ventilated patients. The specific aims are:

i) to conduct a randomised controlled trial to assess the clinical efficacy of a short duration of antibiotics (up to 7 days) versus prolonged antibiotic therapy (physician preference) in adults patients with VAP in Asia;

ii) to study the changes in individuals' microbiota and the impact on pathogen

transmission dynamics and ward ecology with the use of various antibiotic regimens; and

iii) to characterise current antibiotic utilisation in the ICU and develop locally optimised, setting-specific empirical antibiotic policies.

This study includes adult patients ($\geq$ 18 years old) who satisfy our inclusion criteria for VAP after $\geq$ 48 hours of mechanical ventilation.

Patient recruitment will be from:

1) National University Hospital, Singapore
2) Tan Tock Seng Hospital, Singapore
3) Sunpasitthiprasong Hospital, Ubon Ratchathani, Thailand
4) Srinagarind Hospital, Khon Kaen, Thailand
5) Patan Academy of Health Science, Patan Hospital, Kathmandu, Nepal
6) Civil Hospital, Kathmandu, Nepal
7) Khon Kaen Hospital, Khon Kaen, Thailand
8) Pontifícia Universidade Católica do Parana, Curitiba, Brazil

The results of the study will be generalisable to adult ventilated patients from both low-to-middle and high-income countries, especially those with VAP caused by Gram-negative bacteria.

B.1.3 Objectives and outcome measures

Objectives	Outcome Measures	Timepoint(s) of evaluation
Primary Objective To demonstrate clinical non-inferiority, followed by superiority, of a shorter antibiotic treatment duration compared to a longer duration in adult patients with VAP	Difference in the proportion of participants with the composite endpoint of death and VAP recurrence rate within 60(±5) days of enrolment	All outcome measures will be assessed within day 60(±5) of enrolment
Secondary Objectives To define other clinical benefits of shorter antibiotic treatment duration compared to a longer duration in adult patients with VAP	Ventilator-associated events, duration of mechanical ventilation, duration of hospitalisation, acquisition of multidrug resistant infection or colonisation during the hospitalisation, number of days of exposure to antibiotics during hospitalisation, number and types of extrapulmonary infections during hospitalisation (determined from cultures taken from sterile sites), difference in the proportion of participants with primary outcome in subgroups of patients with VAP caused by Gram-negative non-fermenters and carbapenem-resistant bacilli	All outcome measures will be assessed within day 60(±5) of enrolment
To study the impact of various antibiotic regimen on the microbial community dynamics and diversity in the respiratory and intestinal tracts in individual patients in order to understand colonisation and transmission of MDROs, identify potentially protective components of the microbiome against MDROs	Characteristics of the microbiota will be determined by comparing alpha and beta diversity metrics between the groups of patients, alterations in the repertoire of antibiotic resistant genes, shifts in microbiota functional and metabolic capacity	Respiratory and stool samples will be collected from the study participants within 24 hours of enrolment, weekly during hospitalisation, and at day 28 (±5) and 60(±5) after enrolment
To study the impact of antibiotic utilisation on the ward-level multidrug resistant organism emergence and transmission dynamics	Prospective audit of antibiotic utilisation of participating wards, colonising bacteria in respiratory and stool samples, antibiograms of infective and colonising bacteria, whole genome and resistance genes sequencing of multidrug resistant isolates, mathematical modelling using sequencing data to track the spread of pathogens between hosts allowing for unobserved infection times, multiple independent introductions of the pathogen, and within-host genetic diversity	These will be measured throughout the duration of the study in participating wards
To construct predictive models with comprehensive clinical and molecular epidemiological data to produce setting-specific and evidence based antibiotic policies	Statistical models and patient screening algorithms in both low-to- middle and high income settings to predict outcomes including proportion of adequate empiric therapy, expected deaths	N.A.
To estimate the changes to total costs and health benefits measured by QALY from implementing into practice the shorter antibiotic treatment duration	Resources and costs used for clinical implementation of a new policy; changes to costs arising from new policy as shown by length of stay on the ICU and the ward change; antibiotics use; other consumables; tests and diagnostics Risks of adverse events associated with change in duration of ICU stay, particularly other HAIs. Impact on mortality risk and health related quality of life from shortening length of stay in ICU and reducing incidences of other HAIs	Data collected during the trial will be used and combined with other data sources from the study hospitals and published literature to predict costs and health outcomes for a time period that exceeds the trial duration

B.1.4 Study design

The study design is a randomised, partially double-blinded (blinding may be achieved up to 7 days) controlled trial to assess the efficacy of short vs. longer antibiotics course in adults with VAP in Asia. The trial has a stepwise noninferiority–superiority hypothesis, i.e., if non-inferiority of short duration, compared with long duration, is shown, statistical tests for superiority will be performed using closed testing methods without requiring adjustment of the significance level for multiple comparisons.

Screening and inclusion criteria

We will apply the US Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN) VAP diagnostic criteria on patients who have been mechanically ventilated for $\geq$ 48 hours as our study subject inclusion criteria.[124] While we acknowledge that there is no "gold standard" diagnostic criteria to VAP and that clinical criteria correlate poorly with autopsy findings (previously determined to be 69% sensitivity and 75% specificity)[115], the CDC NHSN diagnostic criteria is sensitive and practical for use in ICUs of various resources and settings. Only one episode of suspected VAP will be included for this study.

We agree with the 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America (IDSA) and the American Thoracic Society (ATS) guideline that the additional use of infection scores and other biomarkers including procalcitonin, c-reactive protein, and soluble triggering receptor expressed on myeloid cells-1 lack sensitivity and specificity and will not be included in our inclusion criteria.[40]

Microbiological culture results are not part of our inclusion criteria so that we are able to recruit patients who have suspected VAP but respiratory cultures are negative. Management strategies for this group of patients have not been defined by a randomised study. There are two observational studies that compared outcomes among those whose antibiotics were withheld on the basis of negative respiratory cultures to those whose antibiotics were continued.[119, 120] Patients whose antibiotics were discontinued did not have a higher mortality or rate of new respiratory infection compared to patients whose antibiotics were continued.

Randomisation and Intervention

Following screening and fulfilling the inclusion/exclusion criteria, the study subject will be recruited within 72 hours. Antibiotic treatment course will be determined according to randomisation.

Participants in the intervention (short duration) arm will receive antibiotics for up to 7 days. Antibiotics should be stopped from day 3 to 7 if respiratory cultures are negative and the patients fulfil a set of stringent clinical criteria signifying

cardiopulmonary stability for 48 hours. If the respiratory cultures are positive, patients who fulfil the same set of clinical criteria should have their antibiotics stopped from day 5 to 7 (see section 8. Interventions). Antibiotics administered via all routes i.e. intravenous, oral and nebulisation should be stopped. Randomisation will be carried out when the participants fulfil the clinical stop criteria. The rationale for stopping antibiotics at the above time points is as follows:

Group 1: Short Duration arm

(a) **Day 3** Patients who have negative respiratory cultures should have their antibiotics stopped from day 3 to 7 if the patient satisfies a set of criteria signifying clinical stability. As discussed above, though there are no randomised studies to support this, there are 2 observational studies [119, 120] showing that discontinuation of antibiotics in these cases did not result in a higher mortality or rate of new respiratory infections compared to those whose antibiotics were continued. In addition, those whose antibiotics were withheld had a lower rate of total secondary infections by MDROs. One of these studies concluded that antibiotics could be safely discontinued in patients with negative respiratory cultures with no significant impact on mortality, despite persistence of signs and symptoms of pneumonia in 35% of these patients.[120] The above studies have led to the recommendation in the 2016 IDSA/ATS guideline that for patients with suspected VAP whose respiratory invasive culture results are below the diagnostic threshold for VAP, antibiotics should be withheld rather than continued.[40]

(b) **Day 5-7** The 2016 IDSA/ATS VAP guideline recommends a 7-day course of antimicrobial therapy for the treatment of VAP.[40] However, 7-day course is empirical and durations shorter than 7 days have been studied in randomised trials. One of these studies by Singh *et al* [117] evaluated 3 days of empirical ciprofloxacin monotherapy for patients who satisfy a set of clinical criteria signifying low likelihood of active VAP at day 3 of treatment. Compared to those who received longer duration of antibiotics, there was no difference in mortality or ICU length of stay. Patients in the short duration group had less antibiotic resistance and fewer superinfections. Another randomised study by Micek *et al*,[118] adopted an antibiotic discontinuation policy to shorten VAP treatment. Similarly there was no difference between the short (6.0 $\pm$4.9 days) and long duration treatment groups in terms of mortality and VAP recurrences. The majority of the causative pathogens in this study are Gram-negative bacteria including *Pseudomonas* and *Acinetobacter*, which are also the commonest VAP pathogens in Asia.[118] The above evidence supports the use of a set of clinical stability criteria to establish the duration of antibiotic treatment for VAP.

The set of clinical stability criteria used in this study is adopted from Micek *et al*,[118] and IDSA/ATS consensus guidelines on the management of community-

acquired pneumonia in adults.[213] The community-acquired pneumonia guideline is used because a similar set of clinical stability criteria was not defined in the IDSA/ATS VAP guideline.

Group 2: Long Duration arm Participants in the control (long duration) arm will receive standard care, which is antibiotic treatment for at least 8 days with the exact duration decided by the primary physician.

Sample and data collection

Respiratory samples (at least 1 set) will be collected from the participants within 24 hours of enrolment, if not yet collected by the managing physician, for microbiological cultures. Respiratory cultures will be collected either via the endotracheal tube (ETT) or bronchoalveolar lavage (BAL) as ordered by the primary physicians. Both non-invasive (ETT) and invasive (BAL) samples are acceptable methods for microbiological diagnosis of VAP. No difference in mortality in the invasive versus non-invasive groups was shown in large randomised controlled studies including the Canadian Clinical Trials study,[127] and subsequently a Cochrane meta-analysis[128] of 1,367 patients with VAP. Bacterial isolates grown from all other routine clinical cultures such as blood, cerebral spinal fluid, pus etc, will be retrieved from the microbiology laboratory and frozen and stored. These isolates will undergo whole genome sequencing to compare the within-host evolution of resistant organisms.

For microbiome studies, one set of respiratory and stool samples will be collected within 24 hours and weekly during hospitalisation for DNA extraction and microbiome analysis.

We will collect relevant clinical and microbiological data from the participants throughout their hospitalisation. Data collection will be done via review of medical records, microbiological culture results and drug administrative charts. Concurrently, we will collect prospective data on antibiotics utilisation, adaption and compliance of VAP prevention bundles and hand hygiene rates in the participating units. These audits will include patients who are not study participants. These data are important as they are potential confounders to our outcome measures including mortality, VAP recurrence rate, duration of mechanical ventilation, acquisition of MDRO infection, and number and types of extrapulmonary infections during hospitalisation. Consents will not be required for these patients not directly participating in the study because these audits are standard procedures in hospitals.

Follow-up visits

The composite endpoint of death and VAP recurrence rate is assessed at 60($\pm$5) days of enrolment done via an inpatient or outpatient follow-up visit. The time point of day 60($\pm$5) is chosen, in contrast with day 28 in previous studies, to avoid

a bias in favouring long duration of treatment in assessment of VAP recurrences. A longer time point is deemed impractical and may contribute to high loss-to-follow-up rates especially for patients who reside in rural and remote areas. Two follow-up visits are scheduled at day 28(±5) and 60(±5). One set of respiratory and stool sample will be collected from all participants. The expected duration of participant participation is 60(±5) days from recruitment. Please see Figure 3.2 for schedule of study procedures.

B.1.5 Participant identification and recruitment

Study Participants

Adult patients ($\geq$ 18 years old) in participating wards who have been on mechanical ventilation for more than 48 hours.

Inclusion Criteria

a. Patients 18 years and older

b. Invasive mechanical ventilation $\geq$ 48 hours

c. At least one of the following:

   1. temperature $>$ 38°C
   2. white blood cell count $\geq$ 12,000 cells/mm^3 or $\leq$ 4,000 cells/mm^3
   3. altered mental status with no other causes in >70 year-olds; AND

d. Two or more chest imaging tests demonstrating at least one of the following:

   1. new and progressive OR progressive and persistent infiltrate
   2. new and persistent OR progressive and persistent consolidation
   3. new and persistent OR progressive and persistent cavitation, AND

e. At least two of the following:

   1. new onset of purulent sputum, or change in character of sputum, or increased respiratory secretions, or increased in suctioning requirements
   2. new onset or worsening tachypnea or dyspnea
   3. rales or bronchial breath sounds
   4. worsening gas exchange defined by oxygen desaturations (e.g. PaO2/FiO2<240), increased oxygen requirements or increased ventilation demand

Exclusion Criteria

a. Poor likelihood of survival as defined by a Sepsis-related Organ Failure Assessment score (SOFA score) of >11 points[125]

b. Immunocompromised patients (HIV with CD4 <200 cells/mm3, corticosteroids> 0.5 mg/kg per day for > 30 days, received chemotherapy in the past 3 months, solid organ or hematopoietic cell transplant)

c. Patients receiving antibiotic therapy for any other defined extra-pulmonary infections that warrant a duration of antibiotics longer than seven days or complications of pneumonia such as lung abscess or empyema, that warrant a duration of antibiotics longer than seven days (excluding anti-tuberculosis treatment, antifungal medications, antibiotics meant for chronic suppression of chronic infections or chronic obstructive lung disease)

d. Patients who have been treated for VAP for more than seven days from screening

e. Vulnerable patients including prisoners and refugees

B.1.6 Study procedures

Recruitment

Relevant sub-specialists such as general medicine, respiratory, infectious disease, intensive care and anaesthesia physicians, nurses and pharmacists involved in the care of ventilated patients from all participating sites will be briefed on the study protocol in advance. With their support, a site investigator will be nominated to head the study.

The study team will do screening for potential subjects. Potential subjects are identified according to the eligibility criteria stated above. When a potential subject is identified, our study team will approach the primary physician for his/ her permission for the study team to speak to the patient or patient's legal representative or patient's next-of-kin (NOK) for consent. Patient's legal representative or NOK may be then contacted via phone call or in person. Discussion and explanation of the study procedures and signing of the consent will be done face-to-face.

Screening and Eligibility Assessment

Screening will be done on all mechanically ventilated patients in the respective study sites according to the eligibility criteria stated above. Participants should be recruited into the study as soon as possible after screening. A maximum of 72 hours is allowed between screening (fulfilment of the inclusion/exclusion criteria) and recruitment.

Informed Consent

Participants' legal representatives or NOKs will be asked to represent the participants if the participants are sedated and do not have the capacity to make informed decision to participate in the study. The participants, their legal representatives or NOKs must personally sign and date the latest approved version of the Informed Consent form before any study specific procedures are performed.

Written and verbal versions of the Participant Information and Informed Consent will be presented to the participants, their legal representatives or NOKs detailing no less than: the exact nature of the study; what it will involve for the participant; the implications and constraints of the protocol; the known side effects and any risks involved in taking part. It will be clearly stated that the participant, his/her legal representative or NOK (on behalf of the participant) is free to withdraw from the study at any time for any reason without prejudice to future care, and with no obligation to give the reason for withdrawal.

The participants, his/her legal representatives or NOKs will be allowed sufficient time to consider the information, and the opportunity to question the Investigator or other independent parties to decide whether they will participate in the study. Written Informed Consent will then be obtained by means of participants', their legal representatives' or NOKs' dated signatures and dated signature of the person who presented and obtained the Informed Consent. The person who obtained the consent must be suitably qualified and experienced, and have been authorised to do so by the Investigator. A copy of the signed Informed Consent will be given to the participants, their legal representatives or NOKs. The original signed form will be retained at the study site.

When the participants are deemed to have decision-making capacity by the primary physicians, he/she will be re-consented with the same procedures as above. The concurrent prospective data on clinical cultures, antibiotics utilisation, adaption and compliance of VAP prevention bundles and hand hygiene rates will involve all participating wards throughout the period of the study. All patients admitted to these units will be included as part of this audit and consents will not be taken for this purpose. All information collected as part of the audit will be anonymised to ensure patient confidentiality.

Randomisation, blinding and unblinding

Randomisation will be done via stratified block randomisation by the study sites to ensure participants with similar characteristics such as gender and age are distributed equally in the intervention and control groups. Randomisation will be done with a computer programme with a seed to allow reproducibility. Randomisation will be done with a 1:1 ratio. To prevent predictability of the random sequence, generation of the randomisation sequence is performed by an independent statistician and details

of the randomisation generation is unavailable to all investigators. Randomisation will be allocated using sequentially numbered opaque envelopes. Fitness criteria for randomisation (section 8) must be met prior to randomisation. Patients will be blinded to the study, as they will not be informed of the treatment duration and likely to be sedated and unaware of the treatment regimens. Investigators will be blinded during the assessment of the participants for clinical stability based on the above-described criteria to minimise observer bias. Once conditions for stopping antibiotics are satisfied, the investigator will be unblinded and contact the primary physicians to stop antibiotics. The physicians will remain blinded until they are informed that the participant is suitable to stop antibiotics. Independent assessors, who are assigned to determine pneumonia recurrences, will be blinded from the randomisation arms. This will be achieved by blinding all study details, including randomisation arms, for participants with potential recurrences from these independent assessors.

Baseline Assessments

At recruitment, participants' demographics, medical history, antibiotics administration record, chest X-ray or other imaging findings, biochemical, microbiological and haematological results and clinical parameters will be collected. These are routine investigations ordered by the clinical teams taking care of the patients. Stool samples and upper and lower respiratory tract samples will be collected if they have not already been collected in the 24 hours prior to recruitment.

Subsequent Visits

Clinical information will be collected during participant's hospitalisation. Two follow-ups will be done on day 28($\pm$5) and 60($\pm$5) after recruitment. During these visits, similar clinical information will be collected (if available) via hospital record review. Repeat respiratory and stool sample will also be collected. In the event that the participants are not able to attend the follow-ups in person, a telephone interview can be done in place of a physical visit.

Sample Handling

Respiratory cultures will be collected with the usual volume of 5-10ml per sample. Respiratory cultures may be collected via ETT, BAL or any other methods that the primary physician and local microbiology laboratories routinely use. Stool culture will be collected following defecation with the usual volume of 10-50g per sample. Stool samples are collected at recruitment, weekly during hospitalisation and day 28/60($\pm$5) after enrolment.

Microbiology cultures will be processed and reported in the local laboratories, which are expected to have standing quality control measures. Isolates, once identified, will be preserved, frozen and transported to an Oxford or Singapore study laboratory for genomic studies. These genomic results should concur with the initial microbiology laboratories' identification of the isolates.

a) Respiratory cultures
All samples will be cultured and identified via standard microbiological methods. Susceptibility studies will be performed using EUCAST or CLSI agar method and breakpoints. Culture results with antibiogram are usually available within 48 to 72 hours after sample collection.

Pathogens causing VAP and colonising MDROs will be processed for DNA extraction and undergo whole genome sequencing. These sequencing data will be used to track the spread of pathogens between hosts allowing for unobserved infection times, multiple independent introductions of the pathogen, and within-host genetic diversity.

These isolates will be de-identified and stored for up to 15 years for use in future ethically approved studies. We will seek permission from the participants to store their samples and include this in the informed consents.

b) Respiratory and stool samples
For microbiome analysis, respiratory and stool samples will be processed for DNA extraction. Shotgun metagenomics approach will be used to characterise microbial community dynamics in the respiratory and intestinal tract. The characteristics of the microbiota will be determined by comparing alpha and beta diversity metrics between the groups of patients to explore the short- and long- term impact of the various durations of antibiotics in individual patients. The data will also be used to assess if alterations in the repertoire of antibiotic resistant genes can be detected, and to study shifts in functional and metabolic capacity during the time course with a view to getting mechanistic insights into potentially protective components of the microbiome.

These isolates will be de-identified and stored for up to 15 years for use in future ethically approved studies. We will seek permission from the participants to store their samples and include this in the informed consents.

c) Bacterial isolates from other routine clinical cultures such as blood, cerebral spinal fluid, pus etc
All bacterial isolates including MDR *Acinetobacter* spp, *Pseudomonas* spp, Enterobacteriaceae, *Staphylococcus aureus* and *Enterococcus* spp, grown from other routine clinical cultures will be de-identified and stored for up to 15 years for use in future ethically approved studies. In the current study, these isolates will undergo whole genome sequencing to compare the within-host evolution of resistant

organisms. We will seek permission from the participants to store their samples for up to 15 years in the informed consents.

Discontinuation/Withdrawal of Participants from Study

Each participant or NOK has the right to withdraw from the study at any time. The reason for withdrawal will be recorded in the Case Report Form. Withdrawal from the study will not result in exclusion of the data for that participant from analysis. In addition, the Investigator may discontinue a participant from the study at any time if the Investigator considers it necessary for any reason including but not limited to:

- Ineligibility (either arising during the study or retrospectively having been overlooked at screening)
- Significant protocol deviation
- Significant non-compliance with treatment regimen or study requirements
- Loss to follow up
- Failure to meet fitness criteria by day 7

We will compare baseline characteristics between patients who had protocol violations or who were unavailable for follow-up with those who did not. Follow-ups and data collection listed on the standard CRF for discontinued participants, including those who fail to be randomised, will be carried out within day 60($\pm$5) of enrolment.

Discontinued subject can be rescreened (under a new number). If the subject is rescreened, a new subject number will be allocated. The above patients will not be followed up once withdrawn or discontinued, and replaced to achieve the calculated sample size.

Definition of End of Study

The end of study is the date of the last follow-up visit of the last participant.

B.1.7 Interventions/ investigations

Antibiotic treatment for VAP will be tailored to the susceptibility of the pathogen(s) and in accordance to the 2016 IDSA/ATS VAP guideline.[40] Primary physicians are encouraged to convert initial empirical regimen to narrow-spectrum therapy based on culture results. In culture-negative cases, empirical antibiotic choice should be made depending on local antibiogram. The study team will not intervene on the antibiotics choices as this is beyond the objectives of the study. Number of days of antibiotics is calculated from the first day of appropriate coverage according to the susceptibility of at least 1 of the pathogen(s) recovered from respiratory cultures

taken within 48h of screening or VAP symptom onset.

For those participants randomised to the intervention (short duration) arm, site investigators should assess the patients daily. Antibiotics should be stopped at

a) day 3 to 7, if all respiratory cultures during the same episode of VAP are negative, given that the criteria below are satisfied,

b) day 5 to 7, if any of the respiratory cultures are positive is positive attributable to the current episode of VAP, given that the fitness criteria below are satisfied,

    a. body temperature was $\leq 38.3°$ (core body temperature measured orally or rectally) or $\leq 38.0°$ (axillary) for 48 hours, and

    b. hemodynamic instability (systolic blood pressure ≥ 90 mm Hg without inotropic support or no requirement of inotropic support to maintain systolic blood pressure above 90 mm Hg).

In the short duration arm, all antibiotics should be withdrawn by day 7 (short duration) according to the randomisation assignment, except those participants with treatment failure in the case of persistent VAP or a new-onset infection of a different source prior to the last day of assigned duration of antibiotics. Persistent VAP or treatment failure is defined as the lack of improvement of hemodynamic stability or ventilation requirements without an alternative cause other than the same episode of VAP. These patients will be continued to be monitored for the above-described "fitness criteria" and antibiotics should be stopped when they eventually meet the criteria. Recurrent VAP is defined as an additional episode of VAP satisfying the CDC clinical and radiological criteria of VAP [124] following commencement of the primary VAP treatment within 60($\pm$5) days of enrolment. Opinion from two respiratory experts blinded to the randomisation will be sought to diagnose persistence and recurrences. Patients who are diagnosed with an alternative source of infection or a complication from VAP such as lung abscess or empyema that warrant a duration of antibiotics longer than 7 days prior to the last day of assigned duration of antibiotics are not considered treatment failures. These patients will not meet fitness criteria for randomisation and be discontinued from the study and subsequent analysis (see section 7.8). In cases where patients randomised to short duration and satisfy the above-described "stop criteria" but did not have their antibiotics stopped by day 7, they are considered protocol deviations. They will be analysed with intention to treat analysis.

Duration of antibiotics in control group will be at least 8 days with the exact duration decided by the managing physicians.

Adverse events will be recorded throughout the study period, including *C. difficile* infection, acute kidney injury, hepatitis, drug allergies, haematological and other complications.

B.1.8 Safety reporting

Participants in the shorter treatment duration arm may experience an increase in incidence of VAP relapse. Participating in the study will not delay the participants from receiving antibiotic treatment for recurrent VAPs. Decision to treat these participants for VAP recurrence will be made by the primary physicians. For outcome measurement purposes, opinion from 2 respiratory experts blinded to the randomisation will be sought to diagnose recurrences according to the above-described clinical and radiological criteria of VAP. Interim analysis will be done to ensure that participants in the shorter treatment duration arm are not experiencing a significant increase in incidence of ventilator-associated pneumonia relapse.

Definition of Serious Adverse Events

A serious adverse event is any untoward medical occurrence that:

1. results in death
2. is life-threatening
3. requires inpatient hospitalisation or prolongation of existing hospitalisation
4. results in persistent or significant disability/incapacity

Other 'important medical events' may also be considered serious if they jeopardise the participant or require an intervention to prevent one of the above consequences.

NOTE: The term "life-threatening" in the definition of "serious" refers to an event in which the participant was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe.

Reporting Procedures for Serious Adverse Events

Given that VAP in ICU cohorts is associated with high mortality of 13.6–42.8% [142] and high recurrence rate previously shown in Southeast Asia, we would expect 55%-60% of our estimated sample size of 460 patients to fulfil the above definition of SAE. This constitutes up to 276 (138 in each arm) participants that would require SAE reporting.

As these events are very frequent and expected outcomes from ICU cohorts of patients, local Investigators will report mortalities and VAP recurrences that were 'definitely related' (definitely resulted from shortening antibiotic therapy duration), 'probably related' (some evidence that the event was resulted from shortening antibiotic therapy duration), 'possibly related' (little evidence that the event was resulted from shortening antibiotic therapy duration), and 'not related' (no evidence that the event was resulted from shortening antibiotic therapy duration) to the principal investigator(PI) on a monthly basis in the form of collated SAE reports. PI will, in turn, present these collated SAE reports to a nominated member of Data

and Safety Monitoring Committee (DSMC) on a monthly basis.

Local investigator will present these SAE to the relevant local authorities according to local reporting requirements and timelines.

To address the main risk of our study, which is that the participants in the intervention arm (short duration) may have higher recurrence and mortality rates, we have planned for four interim analysis for safety monitoring by the DSMC.

B.1.9 Statistics and analysis

Description of statistical methods for primary and secondary outcomes

Descriptive statistics will include frequency tables for categorical data, means (SDs), or medians (interquartile ranges) depending on the distribution of the data. Categorical variables will be compared with χ^2 and Fisher's exact tests as appropriate and continuous variables with unpaired, 2-tailed t tests or nonparametric Wilcoxon rank sum tests as appropriate.

The primary and secondary outcomes of the study populations will be analyzed using both unadjusted and adjusted methods in both the per-protocol and intention-to-treat populations. Adjustment will be done with inverse probability weighting, using baseline patient characteristics (study site, age, gender, comorbidities, residence prior to admission, type of ICU admitted to, SOFA score, VAP infection with CRE, maximum heart rate and minimum mean arterial blood pressure on randomisation day, duration of intubation prior to developing VAP, reason for intubation, number of days from first respiratory symptom onset to first day of appropriate antibiotics) as independent variables.

Subgroup analysis will be performed using the above primary outcome amongst patients with VAP caused by Gram-negative non-fermenters and carbapenem-resistant bacilli. These groups are of interest because VAPs caused by Gram-negative non-fermenters have been previously shown to be associated increased recurrence [45, 50]; and VAPs caused by carbapenem-resistant bacilli have no standardised treatment. We will compare the primary outcome in these subgroups in both the per-protocol and intention-to-treat populations, and if the sample sizes allow, the same set of baseline patient characteristics stated above will be used in an inverse probability weighting model. We also will adopt the same non-inferiority hypothesis as the main analysis in these subgroups.

This is a non-inferiority trial with a hierarchical noninferiority–superiority hypothesis. The first analysis to be conducted will be for determination of non-inferiority. Only if non-inferiority is established by this primary analysis, a second analysis for superiority will be conducted using closed testing methods without requiring adjustment of the significance level for multiple comparisons.[54] The trial estimates for the primary outcome will be calculated with the absolute risk difference

(proportion of participants with the primary outcome in the short arm minus that in the long arm). Hence, non-inferiority will be concluded if the upper boundaries of the one-sided 95% confidence intervals from both unadjusted and adjusted analyses are below the non-inferiority margin. The purpose of using both adjusted analyses on the intention-to-treat and per-protocol populations to determine non-inferiority is to minimise the inflation of type 1 error associated with non-adherence in non-inferiority trials.[51] Superiority will be declared if the entire confidence intervals for all the trial estimates are above zero.

Definitions of study populations

Intention-to-treat: The intention-to-treat population includes all study participants who have been randomised during the conduct of the study

Per-protocol: The per-protocol population includes all study participants who fulfill eligibility criteria specified in the inclusion/exclusion criteria (section 6.2 and 6.3), fitness criteria for randomisation (section 8) and received 7 days or less of appropriate antibiotics in the short arm, and 8 days or more of appropriate antibiotics in the long arm.

Interim analyses

Four interim analyses will be performed on the primary endpoint whenever 25% of patients has been randomised and have completed the 60 (±5) days follow-up. The interim analyses will be performed by a trial statistician in coordination with the study team. The statistician will report to the independent DSMC. The DSMC will have unblinded access to all data and will discuss the results of the interim-analysis during a DSMC meeting. A trial steering committee will also be constituted and will decide on the continuation of the trial and will report to the central ethics committee. The trial will be terminated if superiority of either short or long treatment durations is shown. We will use the group sequential design adopting the boundaries proposed by Fleming- Harrington- O'Brien (R=0.8) [142] to terminate the trial prematurely once the Z value exceeds the defined boundaries for non-inferiority.

The number of participants

The study is designed to demonstrate the non-inferiority, followed by superiority, on the composite endpoint of mortality and recurrence at 60 (±5) days of the short duration versus the long duration of antibiotic treatment for VAP. A meta-analysis showed that mortality attributable to VAP ranges from 13.6–42.8% in Southeast Asia.[138] Considering that our primary outcome is a composite binary outcome of mortality and recurrence of VAP, we estimate this to be 55%. We derived an absolute non-inferiority margin of 12% with the fixed-margin method, preserving at least 50% of the efficacy of standard treatment in VAP. Using a group sequential design adopting the boundaries proposed by Fleming- Harrington- O'Brien (R=0.8) [51], a maximum of 412 patients will be required to achieve a power of 80% to conclude non-

inferiority between the two groups with an one-sided α risk of 5%. As we anticipate a loss to follow-up of up to 10%, we plan to enrol a maximum of 460 patients.

Analysis of microbiota and whole genome sequencing

Patients will be stratified into groups according to types and duration of antibiotic exposure. Shotgun metagenomics approach can be used to characterise microbial community dynamics in the respiratory and intestinal tract. The characteristics of the microbiota will be determined by comparing alpha and beta diversity metrics between the groups of patients and healthy volunteers to explore the short- and long- term impact of the various durations of antibiotics in individual patients. The data will also be used to assess if alterations in the repertoire of antibiotic resistant genes can be detected, and to study shifts in functional and metabolic capacity during the time course with a view to getting mechanistic insights into potentially protective components of the microbiome.

Health economics modelling

To assess the implementation cost, we will develop a logic model to describe the experience at each site to specify 'inputs', 'activities', 'outputs' and 'outcomes' arising from a policy of shortening duration of treatments. This will be developed with the site teams and managed by the research team for consistency in language and definitions. This exercise will provide an opportunity to identify the incremental resources required for the new policy. We expect most of the extra costs to be staff time, but there might be some other consumable or equipment costs. We will use qualitative methods to elicit information from local administrators and health services leaders.

To assess the changes to costs arising directly from the new policy, we will use the trial data on lengths of stay, use of antibiotics, use of other consumables; and, tests and diagnostics. Unit costs will be sought from local administrators or shadow prices imputed where necessary. Data on adverse events associated with ICU stay will be assessed and augmented with information from the literature about how any unobserved adverse events change with ICU duration in similar populations. The costs of these events measured by excess length of stay will also be harvested from the literature; as will estimates of changes to risk of death and health related quality of life, measured by preference-based utility tools.

The data we assemble will be organised in a decision tree or state-based model to capture chance events or frequently recurring events. For the economic modelling, the parameters that describe costs, health outcomes and probabilities will be fitted to prior uncertain distribution and simulation studies completed to propagate forward uncertainties to joint posterior distributions of change to costs and health benefits. Results will be presented as incremental cost-effectiveness ratios and rearranged as a net monetary benefits framework to show the probability that adoption of the novel

model of care is cost-effective. Plausible thresholds for valuing health benefits will be used for each country. A probability that adoption is cost-effective that exceed 50% means a rational and risk neutral decision maker will adopt. Not adopting in this situation has a lower chance of being the best decision. The perspective of the health service will be adopted for this evaluation and good practice guidelines for reporting cost-effectiveness will be met.

B.1.10 Data management

Access to Data

Direct access will be granted to authorised representatives from the University of Oxford and any host institution for monitoring and/or audit of the study to ensure compliance with regulations.

Data Handling and Record Keeping

All study data will be entered on MACRO, the Clinical Data Management System. The study database will be developed according to the approved CRFs. As for documents that contain patient identifiers, only the site PIs will have access to these documents.

The MORU Data Management Standard Operating Procedures will be followed in the study. After the study is closed, the study documents will be kept in the accessible storage on request for 15 years. The final dataset will be archived in the data repository after the publication.

B.1.11 Quality control and quality assurance procedures

The study will be conducted in accordance with relevant regulations and standard operating procedures.

Study monitoring

An appointed study monitor from the MORU clinical trial support group and a project coordinator will regularly visit the study sites for quality control. All study sites will be assessed prior to initiation of the study for capacity to conduct the randomised controlled trial, during the study and upon completion to ensure data quality. After site initiation, monitoring visits will perform monitoring on informed consent forms, CRF for completeness and accuracy of data and sample storage. There will be a minimum of three monitoring visits per study site: after 3-5 participants enrolled, after 50% of the target sample size enrolled for a particular site and upon study completion. Monitoring reports will be made available to the study sites and investigators after each visit.

Strategies to improve adherence

Regular meetings with the study sites will be carried out to ensure buy-in from the physicians and healthcare providers. Their feedback will be sought throughout the study, and improvements will be made in the study procedures to maintain ongoing support from the local ICUs. The study team will contact the primary physicians prior to enrolment and randomisation to ensure their adherence to allocated interventions.

B.1.12 Ethical and regulatory considerations

Declaration of Helsinki

The PI will ensure that this study is conducted in accordance with the principles of the Declaration of Helsinki.

Guidelines for Good Clinical Practice

The PI will ensure that this study is conducted in accordance with relevant regulations and with Good Clinical Practice.

Approvals

The protocol, participant information sheet and informed consent form will be submitted to OxTREC and local ethics committees for written approval. The Investigator will submit and, where necessary, obtain approval from the above parties for all amendments to the original approved documents.

Participant Confidentiality

The study staff will ensure that the participants' anonymity is maintained. The participants will be identified only by a participant ID number on all study documents and any electronic database, with the exception of the CRF, where participant initials will be added. All documents will be stored securely and only accessible by study staff and authorised personnel. The study will comply with the Data Protection Act, which requires data to be anonymised as soon as it is practical to do so.

Expenses and Benefits

The study team will reimburse reasonable travel expenses for any visits additional to normal care.

Reporting

The PI shall submit an Annual Progress Report to OxTREC and local ethics committees on the anniversary of the date of approval of the study. In addition, the PI shall submit an End of Study Report to OxTREC and local ethics committees.

B.1.13 Finance and insurance

Funding

The project will be jointly funded by Medical Research Council/ Department for International Development (MRC/DfID) (Grant Ref: MR/K006924/1) and Singapore National Medical Research Council (Grant Ref: CTGIIT18MAY-0005).

Insurance

Specialist insurance policy is in place, which would operate in the event of any participant suffering harm as a result of their involvement in the research (Newline Underwriting Management Ltd, at Lloyd's of London).

B.1.14 Publication policy

The PI will coordinate writing and reviewing of drafts of the manuscripts, abstracts, and any other publications arising from the study. Authorship will be determined in accordance with the ICMJE guidelines and other contributors will be acknowledged.

B.2 Additional tables

B.2.1 Detailed definitions of primary and secondary outcomes

	Definition	Method of data capture
Mortality	Death from any cause within 60 days of enrolment	Medical record review and/ or interview
Pneumonia recurrence	Additional episode of pneumonia as determined by two independent infectious disease or respiratory medicine experts blinded to the randomisation within 60 days of enrolment	Medical record review for pneumonia symptoms following the performance of new respiratory culture or change in antibiotic regime by independent experts blinded to randomisation
Ventilator-associated events	Additional episode of respiratory event fulfilling the US CDC NHSN criteria for VAP within 60 days of enrolment, but not determined to be pneumonia recurrence by two independent infectious disease or respiratory medicine experts blinded to the randomisation	Medical record review for the US CDC NHSN criteria for VAP following the performance of new respiratory culture or change in antibiotic regime by independent experts blinded to the randomisation
Duration of mechanical ventilation	Total number of days which a study participant is dependent on a ventilator for respiratory support within 60 days of enrolment	Medical record review
Duration of hospitalisation	Total number of days of hospitalisation which the index episode of VAP that led to the study participant's enrolment took place	Medical record review
Duration of exposure to antibiotics during hospitalisation	Total number of days which the study participant is administered antibiotics while inpatient for the index episode of VAP that led to the study participant's enrolment	Inpatient medicine prescription record review
Acquisition of multidrug resistant infection or colonisation during hospitalisation	Clinical cultures positive for third-generation cephalosporin resistant Enterobacteriaceae, carbapenem-resistant Gram-negative bacteria, methicillin-resistant Staphylococcus aureus and vancomycin-resistant Enterococcus	Review of all clinical cultures performed for the study participants during hospitalisation
Number and types of extra-pulmonary infections identified from sterile sites during hospitalisation	Clinical cultures positive for any organisms taken from normally sterile sites including blood, cerebrospinal fluid, pleural fluid, peritoneal fluid, pericardial fluid, bone and synovial tissue, joint fluid, specimen obtained from surgery or aspirate from lymph node, brain, heart, liver, spleen, vitreous fluid, kidney, pancreas, ovary, vascular tissue and muscle	Review of all clinical cultures performed for the study participants during hospitalisation

Table B.1: Detailed definitions of primary and secondary outcomes and how data will be retrieved.

B.2.2 Baseline characteristics of the per-protocol patients

	Long arm (n = 114)	**Short arm (n = 106)**	**Overall (n = 220)**
Age			
Age (mean, SD)	64.9 (16.7)	60.3 (17.5)	62.7 (17.2)
Age (median, IQR)	66.1 [54.0, 77.2]	62.7 [47.9, 74.4]	64.6 [50.7, 76.0]
Gender			
Female	45 (39%)	44 (42%)	89 (40%)
Male	69 (61%)	62 (58%)	131 (60%)
Study site			
Patan Hospital	8 (7%)	7 (7%)	15 (7%)
NUH	7 (6%)	7 (7%)	14 (6%)
TTSH	2 (2%)	4 (4%)	6 (3%)
Sunpasit Hospital	96 (84%)	88 (83%)	184 (84%)
Srinagarind Hospital	1 (1%)	0 (0%)	1 (0%)
Residence prior to ICU admission			
Own Home	29 (25%)	35 (33%)	64 (29%)
Transfer from another hospital	85 (75%)	71 (67%)	156 (71%)
Charlson comorbidity score			
Score (mean, SD)	3.2 (2.4)	3.1 (2.1)	3.1 (2.3)
Score (median, IQR)	3.0 [1.0, 4.0]	3.0 [1.0, 4.8]	3.0 [1.0, 4.2]
Comorbidities			
Congestive heart failure	6 (5%)	14 (13%)	20 (9%)
Coronary heart disease	10 (9%)	7 (7%)	17 (8%)
COPD	9 (8%)	11 (10%)	20 (9%)
Liver cirrhosis	1 (1%)	1 (1%)	2 (1%)
Chronic kidney disease	13 (11%)	13 (12%)	26 (12%)
Cancer	4 (4%)	4 (4%)	8 (4%)
Diabetes	31 (27%)	19 (18%)	50 (23%)
SOFA score			
SOFA score (mean, SD)	6.2 (2.4)	6.3 (2.6)	6.3 (2.5)
SOFA score (median, IQR)	6.0 [5.0, 8.0]	6.0 [4.0, 8.0]	6.0 [4.0, 8.0]
Ward			
Medical ICU	37 (32%)	34 (32%)	71 (32%)
Surgical ICU	77 (68%)	72 (68%)	149 (68%)
Duration of intubation prior to VAP			
Days (mean, SD)	17.8 (14.3)	21.3 (24.9)	19.5 (20.1)
Days (median, IQR)	13.0 [10.0, 21.8]	14.0 [10.2, 20.8]	13.0 [10.0, 21.0]
Reason for intubation			
Heart failure	32 (28%)	20 (19%)	52 (24%)
Metabolic acidosis	9 (8%)	7 (7%)	16 (7%)
Neurological failure	35 (31%)	33 (31%)	68 (31%)
Respiratory failure	11 (10%)	15 (14%)	26 (12%)
Sepsis	23 (20%)	29 (27%)	52 (24%)
Trauma	4 (4%)	2 (2%)	6 (3%)
Carbapenem-resistant bacteria in sputum sample during the index episode of VAP			
Carbapenem-resistant Gram-negative bacteria isolated	18 (16%)	19 (18%)	37 (17%)
Vital signs on enrolment			
Lowest MAP prior to randomisation (mean, SD)	86.9 (11.9)	83.2 (14.4)	85.1 (13.2)
Mean maximum heart rate (mean, SD)	100.1 (18.8)	99.6 (15.9)	99.9 (17.4)
Inotropic support prior to randomisation	2 (2%)	7 (7%)	9 (4%)
SpO2/FiO2 ratio (mean, SD)	252.8 (88.6)	265.5 (94.3)	258.9 (91.4)
SpO2/FiO2 ratio (median, IQR)	247.5 [213.7, 250.0]	247.5 [200.0, 250.0]	247.5 [200.0, 250.0]

Table B.2: Baseline characteristics of the per-protocol patients.

B.2.3 Antibiotics prescribed during follow-up period

Antibiotics	Frequency†(%)		
	All participants (n =230)	Long arm (n =115)	Short arm (n =115)
Cephalosporins	**112 (49)**	**62 (54)**	**50 (44)**
Ceftazidime	107 (47)	59 (51)	48 (42)
Cefepime	5 (2)	3 (3)	2 (2)
Beta-lactams/beta-lactamase inhibitors	**230 (100)**	**115 (100)**	**115 (100)**
Piperacillin/Tazobactam	92 (40)	44 (38)	48 (42)
Amoxicillin/clavulanic acid	42 (18)	21 (18)	21 (18)
Cefoperazone/sulbactam	123 (53)	65 (57)	58 (50)
Carbapenems	**227 (99)**	**115 (100)**	**105 (91)**
Meropenem	119 (52)	64 (56)	55 (48)
Imipenem	96 (42)	52 (45)	44 (38)
Ertapenem	12 (5)	6 (5)	6 (5)
Fluoroquinolones	**47 (21)**	**19 (17)**	**28 (24)**
Ciprofloxacin	20 (9)	8 (7)	12 (10)
Levofloxacin	27 (12)	11 (10)	16 (14)
Aminoglycosides	**43 (18)**	**26 (23)**	**17 (14)**
Amikacin	33 (14)	20 (17)	13 (11)
Gentamicin	7 (3)	3 (3)	4 (3)
Netilmicin	3 (1)	3 (3)	0 (0)
Colistin	**80 (35)**	**45 (39)**	**35 (30)**
Tigecycline	**39 (17)**	**24 (21)**	**15 (13)**
Vancomycin	**48 (21)**	**29 (25)**	**19 (17)**
Clindamycin	**68 (30)**	**45 (39)**	**23 (20)**
Trimethoprim/sulfamethoxazole	**19 (8)**	**10 (9)**	**9 (8)**

Table B.3: Antibiotics prescribed during follow-up period. Antibiotic classes are highlighted in bold. †Antibiotics were prescribed in combination for some patients and hence not all percentages sum to 100%.

C

Effect of antibiotic treatment duration on antimicrobial resistance in the hospital setting

Contents

C.1 Within- and between-host resistant bacterial carriage model exploration

C.1.1 Optimising number of iterations per parameter combination

I used parameter ranges found in the literature (Table 4.1) and performed global sensitivity analysis using Latin Hypercube sampling and Partial Rank Correlation Coefficient (LHS-PRCC). LHS, a Monte Carlo sampling method, is favoured for its computing efficiency.[214] PRCC is a highly efficient and reliable measure that accounts for non-linear but monotonic relationships between the parameters and model output.[215] We calculated the confidence intervals for the PRCCs using bootstrapping (1000 times).

Prior to calculating PRCC, I examined the correlations of each parameter with the model output for non-monotonicities via visualisation of scatter plots, Spearman's rank correlation measure and Hoeffding's D measure.[216, 217] The optimal sample size for each sensitivity analyses was chosen when additional parameter combinations did not alter the output of the sensitivity analysis significantly. This is calculated using the concordance measure, Symmetrised Blest Measure of Association (SBMA) to confirm the optimal number of parameter combinations).[218]

In addition, I averaged the model outputs from a number of iterations for the same set of parameter combination inputs, which has been shown to improve reproducibility at smaller sample sizes and strengthen correlation between parameter values and model output during sensitivity analyses.[215] To find the optimal number of iterations per simulation, we compared distributions of simulation outputs under identical parameter values using the Vargha-Delaney A-Test, which is a nonparametric measure of the difference between the distributions of model outputs suggesting if the outputs are consistent (Figure C.2).[219, 220]

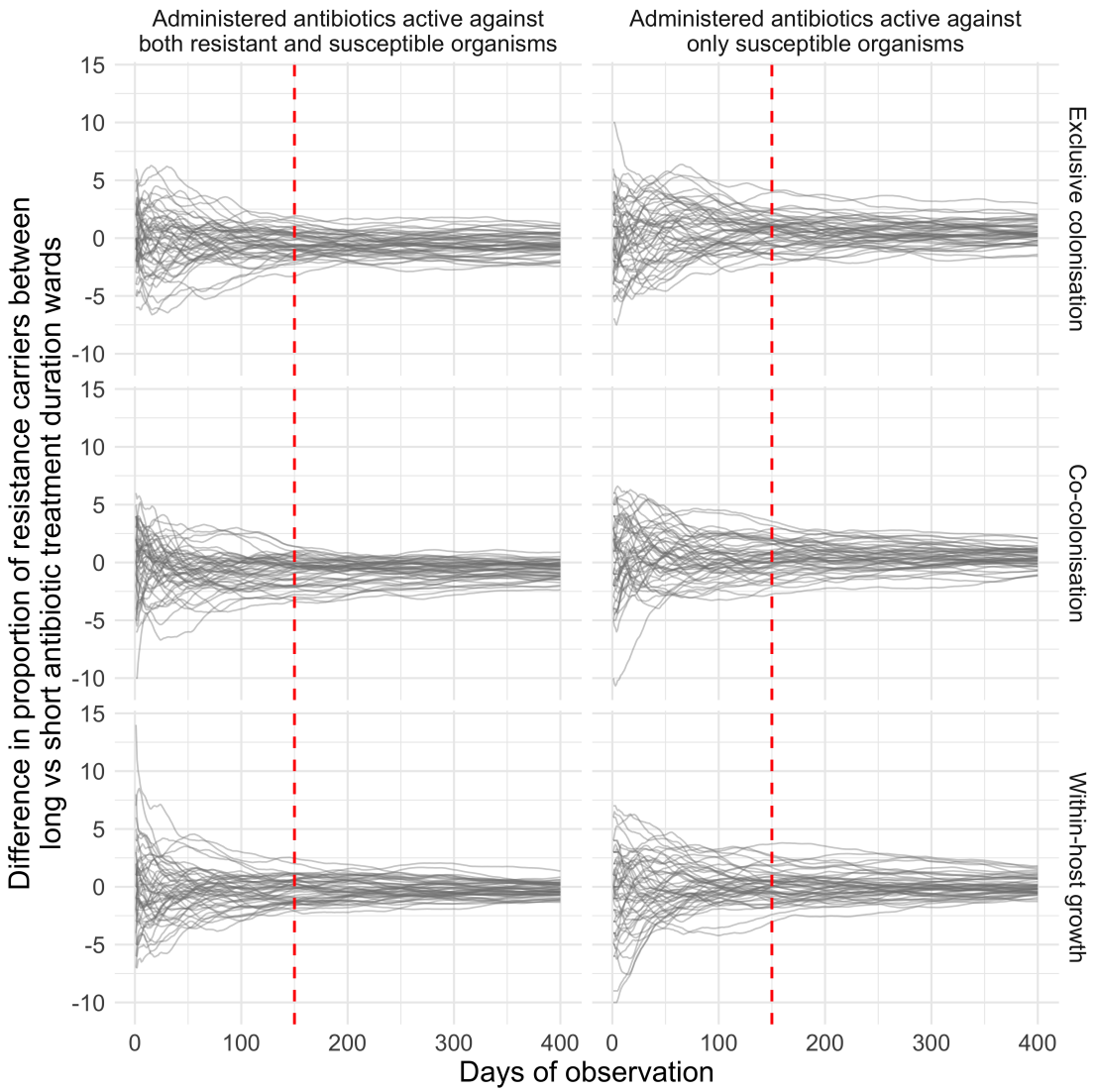


Figure C.1: Checking equilibrium state of the models. To find the number of time steps (days) taken for the models to reach equilibrium, we plotted the output, absolute difference in the proportion of resistance carriers per day between and long and short wards (y-axis), against the number of days of observation (x-axis). Each grey line represents the output from one iteration. The same set of parameter values were used for each iteration. The graphs on the left-side panels are the runs from the scenario where administered antibiotics were active against both resistant and susceptible organisms; right-side panels are from the scenario those from the scenario where administered antibiotics were only active against susceptible organisms. Equilibrium states were generally reached after 150 days for all models (red dashed line). Hence in the simulations, the models were ran over 300 days with the first 150 days discarded.

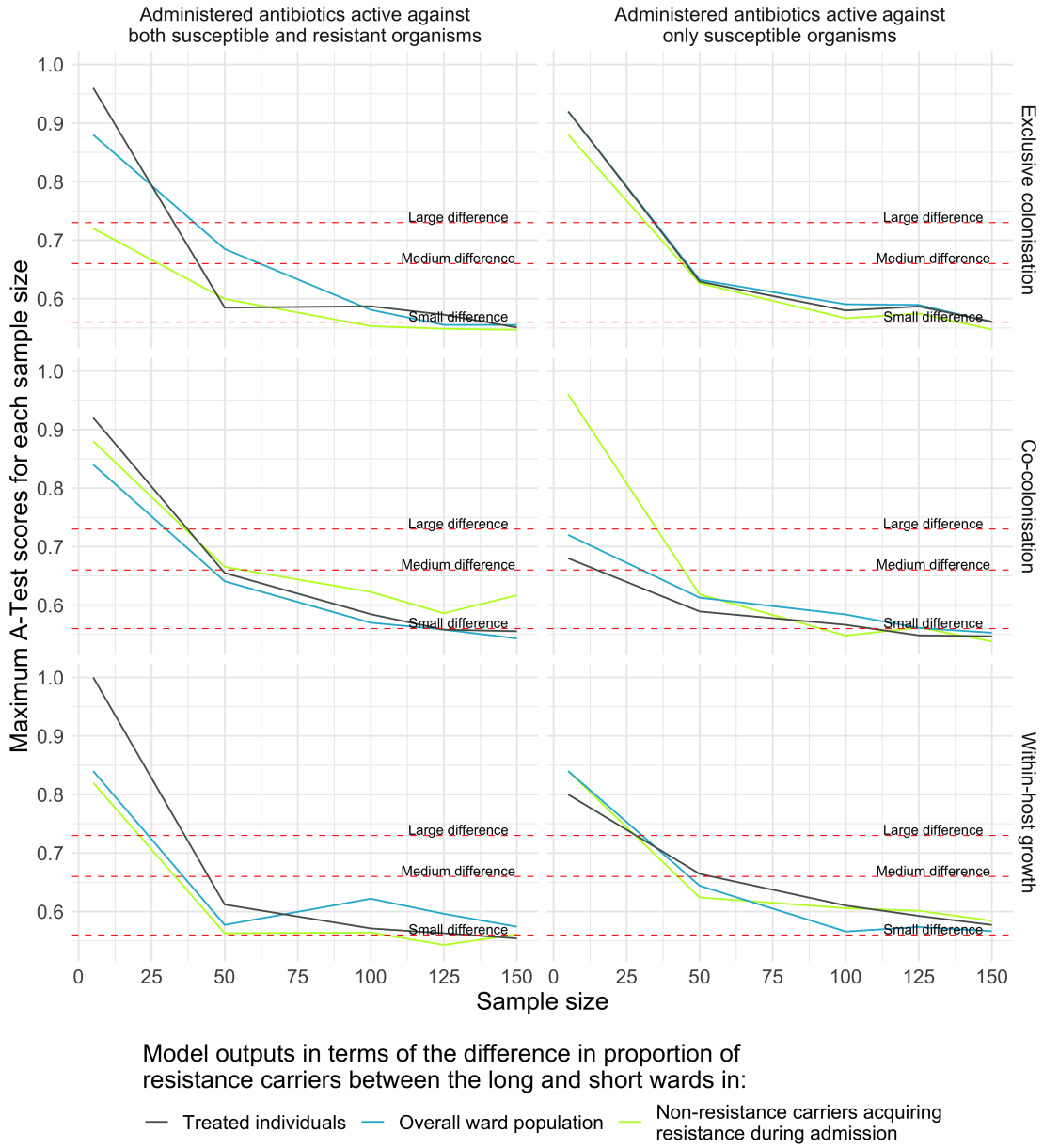


Figure C.2: Determining optimal number of iterations for each model. Consistency analysis was performed with the Vargha-Delaney A-Test to find the optimal number of iterations for each set of simulations ran with the same parameter value combinations. The horizontal dashed lines represent arbitrary A-Test score thresholds for small, medium and large differences at 0.56, 0.66 and 0.73 respectively. Left-side panels show results from the scenario where administered antibiotics were active against both resistant and susceptible organisms; right-side panels are obtained from the scenario where administered antibiotics were only active against susceptible organisms. Since there were small differences detected beyond 100 iterations across all models, 100 iterations per simulation with a single set of parameters value combination were ran.

C.1.2 Correlation measures to ensure monotonicity

Scatter plots

The following scatter plots display the model outputs as a function of each parameter. The parameter values are sampled from the hypercubes formed within each parameter space. The following scatter plots show the parameter values against with the output from the three models. Each dot represents the model output, either in terms of proportion of resistance carriers in the overall ward population (blue), amongst those who were admitted as non-resistance carriers but gained resistance colonisation during admission (orange), and in the treated individuals (grey). The lines are regression lines obtained from locally estimated scatterplot smoothing (LOESS). Details of each parameter can be found in main text Table 4.1.

i) Exclusive colonisation model

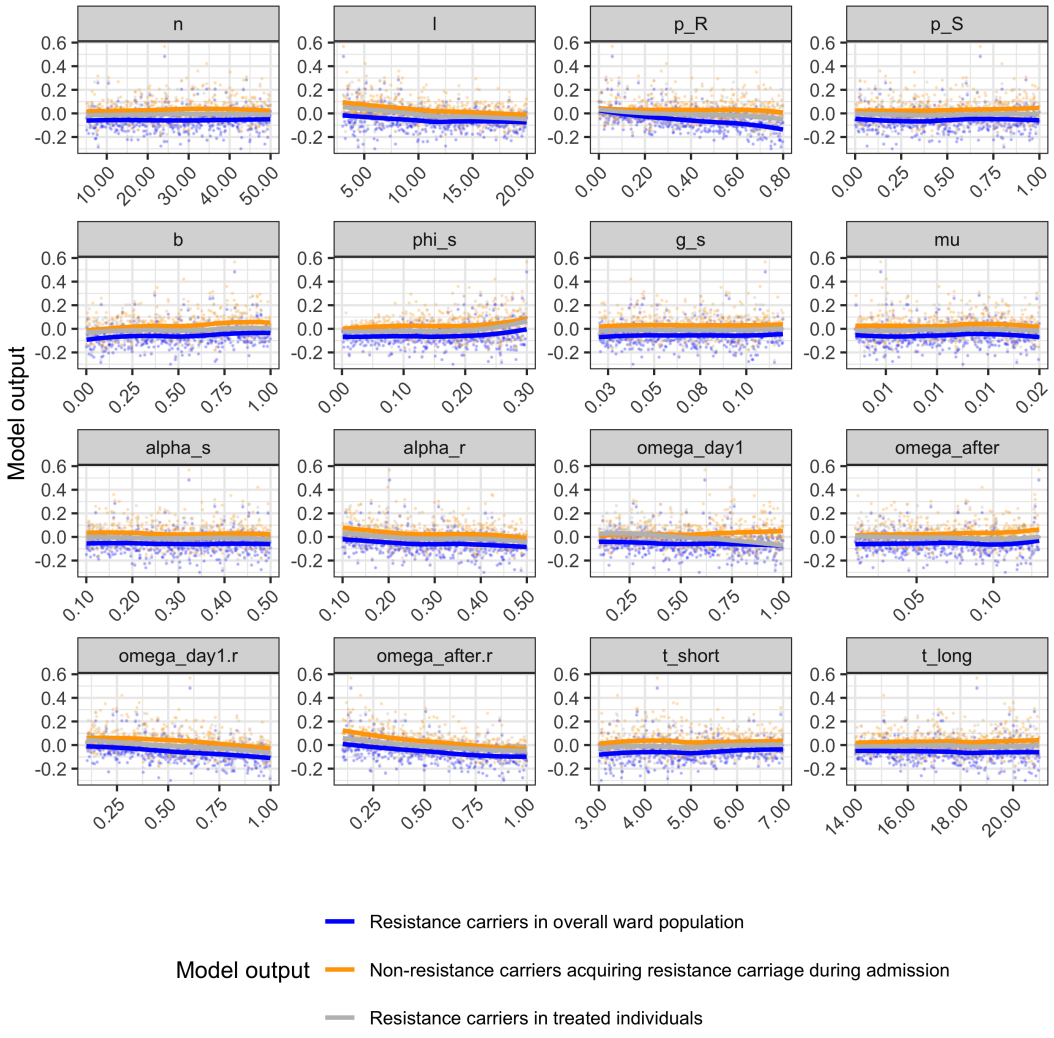


Figure C.3: Scatter plot of each parameter against the model outputs - Exclusive colonisation model.

ii) Co-colonisation model

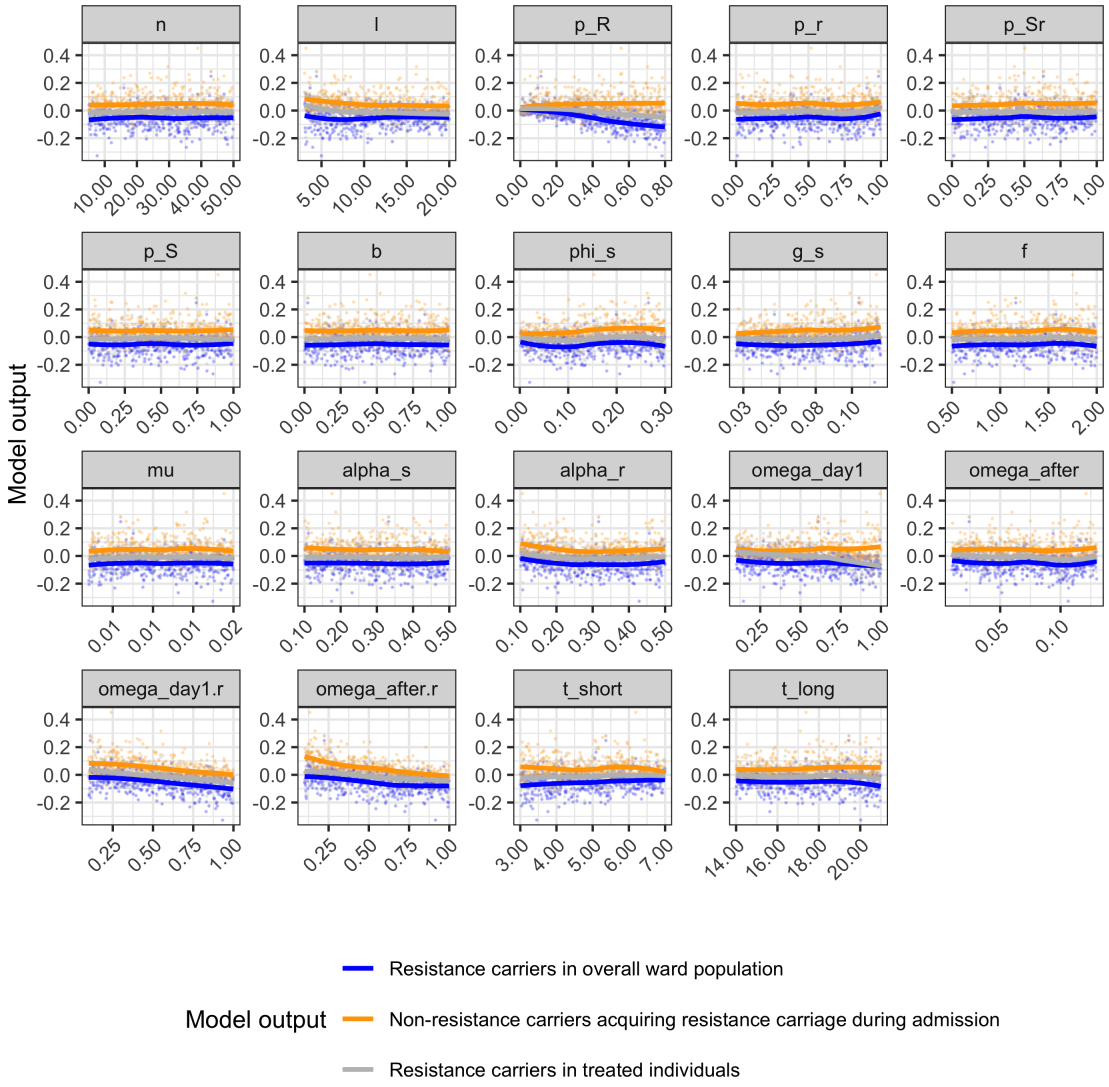


Figure C.4: Scatter plot of each parameter against the model outputs - Co-colonisation model.

iii) Within-host growth model

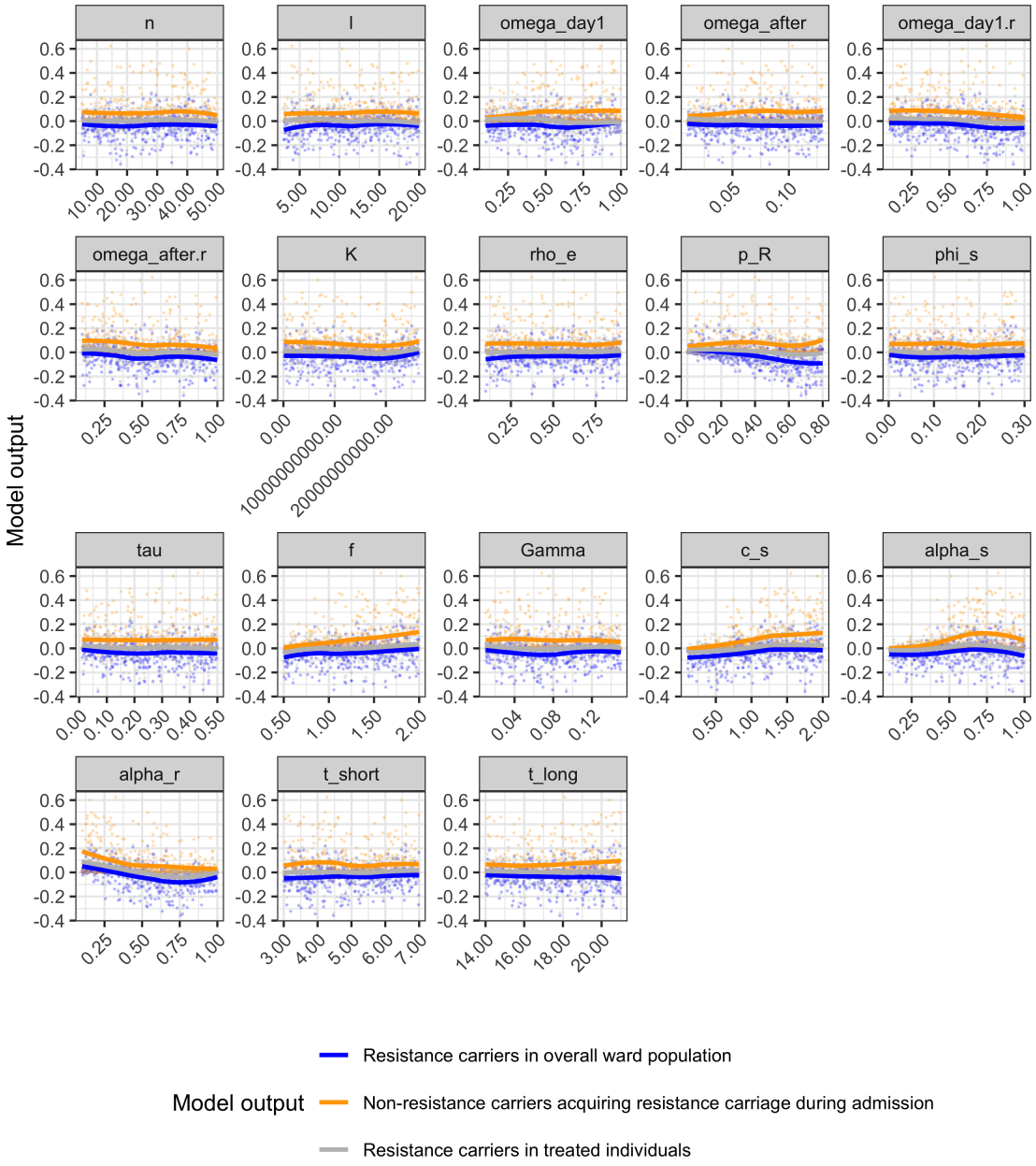


Figure C.5: Scatter plot of each parameter against the model outputs - Within-host growth model.

Hoeffding's D measure and Spearman's rank correlation measures
Spearman's rank correlation indicated the strength and direction of a relationship between two variables and falls between 1 and -1. Spearman's correlation can be applied to a non-monotonic relationship to determine if there is a monotonic component to the association.

Hoeffding's D measure tests the independence of the data sets by calculating the distance between the product of the marginal distributions under the null hypothesis and the empirical bivariate distribution. Unlike the Pearson or Spearman measures, it detects non-linear relationships. Hoeffding's D lies on the interval [-0.5, 1] if there are no tied ranks, with larger values indicating a stronger relationship between the variables.

		Simple 3-state model		Co-carriage 5-state model		Population growth model	
Parameter	**Outcome**	**HD**	**SPM**	**HD**	**SPM**	**HD**	**SPM**
n	R carriers per day	0 (0.14)	0.08 (0.10)	0 (0.70)	0.02 (0.67)	0 (0.29)	0.03 (0.49)
	New R acquisitions per admission	0 (0.57)	0.02 (0.70)	0 (0.62)	0.02 (0.62)	0 (0.59)	0.01 (0.80)
l	R carriers per day	0.01 (<0.01)	0.14 (0.01)	0 (0.73)	0.01 (0.91)	0 (0.50)	-0.01 (0.77)
	New R acquisitions per admission	0.01 (<0.01)	-0.22 (<0.01)	0 (0.09)	-0.10 (0.04)	0 (0.14)	0.07 (0.15)
p_R	R carriers per day	0.25 (<0.01)	0.76 (<0.01)	0.46 (<0.01)	0.91 (<0.01)	0.55 (<0.01)	0.93 (<0.01)
	New R acquisitions per admission	0 (0.29)	0 (0.97)	0 (0.05)	0.10 (0.04)	0 (0.23)	0.01 (0.92)
p_S	R carriers per day	0 (0.21)	-0.06 (0.24)	0 (0.21)	-0.03 (0.54)	-	-
	New R acquisitions per admission	0 (0.09)	0.09 (0.05)	0 (0.48)	0.02 (0.73)	-	-
b	R carriers per day	0.01 (<0.01)	-0.18 (<0.01)	0 (0.76)	-0.01 (0.83)	-	-
	New R acquisitions per admission	0.01 (<0.01)	0.17 (<0.01)	0 (0.44)	0.02 (0.70)	-	-
φ	R carriers per day	0.06 (<0.01)	0.42 (<0.01)	0 (0.55)	0.03 (0.53)	0 (0.55)	0.01 (0.81)
	New R acquisitions per admission	0.03 (<0.01)	0.12 (0.01)	0.01 (<0.01)	0.17 (<0.01)	0 (0.08)	-0.05 (0.31)
g_s	R carriers per day	0 (0.36)	-0.05 (0.34)	0 (0.05)	0.11 (0.03)	-	-
	New R acquisitions per admission	0 (0.46)	0 (0.99)	0.01 (<0.01)	0.19 (<0.01)	-	-
μ	R carriers per day	0 (0.03)	-0.09 (0.07)	0 (0.11)	-0.07 (0.17)	-	-
	New R acquisitions per admission	0 (0.27)	0.07 (0.14)	0 (0.65)	0.02 (0.67)	-	-
α_s	R carriers per day	0 (0.34)	0.01 (0.80)	0 (0.47)	0.03 (0.50)	0 (0.10)	0.08 (0.09)
	New R acquisitions per admission	0 (0.49)	-0.04 (0.47)	0 (0.12)	-0.06 (0.20)	0.06 (<0.01)	0.36 (<0.01)
α_r	R carriers per day	0 (0.03)	-0.11 (0.03)	0 (0.05)	-0.11 (0.02)	0.01 (<0.01)	-0.18 (<0.01)
	New R acquisitions per admission	0.01 (<0.01)	-0.17 (<0.01)	0 (0.15)	-0.09 (0.07)	0 (0.01)	-0.13 (0.01)
ω_{day1}	R carriers per day	0 (0.17)	-0.09 (0.07)	0 (0.08)	-0.08 (0.09)	0 (0.22)	-0.04 (0.37)
	New R acquisitions per admission	0.01 (<0.01)	0.11 (0.02)	0.03 (<0.01)	0.23 (<0.01)	0.01 (<0.01)	0.22 (<0.01)
ω_{after}	R carriers per day	0 (0.14)	0.08 (0.11)	0 (0.24)	0.03 (0.56)	0 (0.47)	0.01 (0.91)
	New R acquisitions per admission	0 (0.25)	-0.07 (0.14)	0 (0.36)	-0.06 (0.24)	0 (0.41)	-0.02 (0.66)
$\omega_{\text{day1.r}}$	R carriers per day	0 (0.02)	-0.13 (0.01)	0.01 (<0.01)	-0.17 (<0.01)	0 (0.02)	-0.10 (0.05)
	New R acquisitions per admission	0.22 (<0.01)	-0.75 (<0.01)	0.25 (<0.01)	-0.77 (<0.01)	0.03 (<0.01)	-0.34 (<0.01)
$\omega_{\text{after.r}}$	R carriers per day	0 (0.80)	-0.01 (0.88)	0 (0.63)	0 (0.94)	0 (0.45)	-0.01 (0.86)
	New R acquisitions per admission	0 (0.57)	-0.01 (0.82)	0 (0.09)	-0.07 (0.13)	0 (0.50)	-0.03 (0.49)
t_{short}	R carriers per day	0 (0.51)	-0.01 (0.82)	0 (0.52)	-0.02 (0.66)	0 (0.17)	-0.01 (0.86)
	New R acquisitions per admission	0 (0.47)	-0.04 (0.44)	0 (0.16)	-0.03 (0.52)	0 (0.12)	-0.02 (0.68)

Table C.1 continued from previous page

		Simple 3-state model		Co-carriage 5-state model		Population growth model	
t_{long}	R carriers per day	0 (0.65)	-0.01 (0.78)	0 (0.50)	0.01 (0.91)	0 (0.54)	-0.01 (0.92)
	New R acquisitions per admission	0 (0.05)	0.08 (0.09)	0 (0.05)	0.08 (0.08)	0 (0.14)	0.10 (0.04)
p_r	R carriers per day	-	-	0 (0.04)	0.11 (0.03)	-	-
	New R acquisitions per admission	-	-	0 (0.01)	0.12 (0.01)	-	-
p_{Sr}	R carriers per day	-	-	0 (0.04)	-0.09 (0.06)	-	-
	New R acquisitions per admission	-	-	0 (0.17)	0.06 (0.24)	-	-
f	R carriers per day	-	-	0 (0.04)	0.10 (0.05)	0 (0.49)	0.03 (0.51)
	New R acquisitions per admission	-	-	0 (0.05)	0.10 (0.04)	0.03 (<0.01)	0.31 (<0.01)
K	R carriers per day	-	-	-	-	0 (0.53)	-0.01 (0.84)
	New R acquisitions per admission	-	-	-	-	0 (0.13)	-0.08 (0.12)
ρ_e	R carriers per day	-	-	-	-	0 (0.41)	-0.06 (0.23)
	New R acquisitions per admission	-	-	-	-	0 (0.39)	-0.06 (0.26)
Γ	R carriers per day	-	-	-	-	0 (0.10)	-0.08 (0.10)
	New R acquisitions per admission	-	-	-	-	0 (0.18)	-0.08 (0.10)
τ	R carriers per day	-	-	-	-	0 (0.45)	0.04 (0.47)
	New R acquisitions per admission	-	-	-	-	0 (0.30)	0.07 (0.16)
c_s	R carriers per day	-	-	-	-	0 (0.53)	0.03 (0.52)
	New R acquisitions per admission	-	-	-	-	0.07 (<0.01)	0.44 (<0.01)

Table C.1: Hoeffding's D measure and Spearman's rank correlation measure. The numbers in the brackets refer to the p-values calculated with 0.05 type 1 error. HD: Hoeffding's D measure; SPM: Spearman's rank correlation

C.2 Meta-analysis of antibiotic duration effect on resistant bacterial carriage

C.2.1 Model assessment

Prior distributions were selected to be weakly informative normal distributions. We assessed the models using measures of Markov chain convergence including effective sample sizes and $\hat{R}$ which indicate if the chains had run for long enough and had mixed well.

In all the models, the $\hat{R}$ values were about 1 and the minimum effective sample size was at least 1000 across all parameters. The chains' mixing from the chosen model is shown below (Figure C.6).

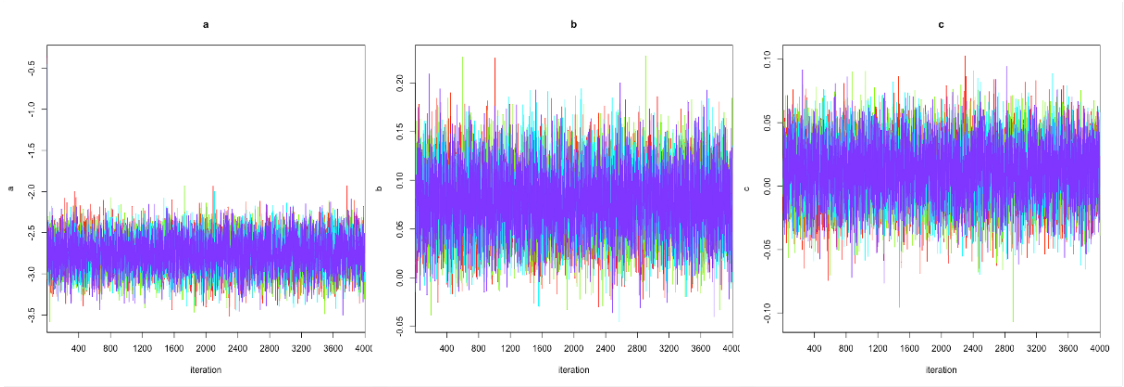


Figure C.6: MCMC chains' mixing for the selected meta-analysis model. The plots show iterations vs. sampled values for model parameters in the MCMC chains. The four different chains are plotted using different colours.

C.2.2 Model comparison

The three models were compared with the best fit to data by deviance information criterion (DIC) and Watanabe–Akaike information criterion (WAIC). The chosen model, model 2 described in main text Box 4.2, was the one which has an intercept, representing the baseline colonisation risk, and slopes, representing colonisation risk associated with one additional day of antibiotic treatment and follow-up period respectively. In this model, both the intercepts and slopes varied by trials (random effect given to the trials).

No.	Model	WAIC	DIC
1	Trials as random effect	-2.80	128.33
2	Trials as random effect, exclusion of healthcare setting as an independent variable	-2.81	128.15
3	No random effect	-2.78	192.60

Table C.2: Comparison of widely applicable information criterion (WAIC) and deviance information criterion (DIC) between the three models.

C.2.3 Sensitivity analysis

		Odds ratio for being colonised with resistant bacteria per additional day of antibiotic treatment (80% credible intervals)
1	**Main analysis**	1.05 (0.90 to 1.23)
2	**30-day cut-off for surveillance cultures**	1.10 (0.86 to 1.38)
3	**Using priors normal(0, 7) instead of normal(0, 3)**	1.05 (0.94 to 1.17)

Table C.3: Sensitivity analyses were performed with a 30-day cut-off for surveillance cultures and different priors.

D
Conclusion

Contents

D.1 Abstracts for other published works

D.1.1 Transmission of community- and hospital-acquired SARS-CoV-2 in hospital settings in the UK: a cohort study

Mo Y, Eyre DW, Lumley SF, Walker TM, Shaw RH, O'Donnell D, Butcher L, Jeffery K, Donnelly CA; Oxford COVID infection review team, Cooper BS. Transmission of community- and hospital-acquired SARS-CoV-2 in hospital settings in the UK: A cohort study. PLoS Med. 2021 Oct 12;18(10):e1003816.

Background: Nosocomial spread of SARS-CoV-2 has been widely reported, but the transmission pathways amongst patients and healthcare workers are unclear. Identifying the risk factors and drivers for these nosocomial transmissions are critical for infection prevention and control interventions. The main aim of our study was to quantify the relative importance of different transmission pathways of SARS-CoV-2 in the hospital setting.

Methods and findings: This is an observational cohort study using data from four teaching hospitals in Oxfordshire, UK, from January to October 2020. Associations between infectious SARS-CoV-2 individuals and infection risk were quantified using logistic, generalised additive and linear mixed models. Cases were classified as community- or hospital-acquired using likely incubation periods of three to seven days. Nine-hundred and twenty of 66184 patients who were hospitalised during the study period had a positive SARS-CoV-2 PCR test within the same period (1.4%). The mean age was 67.9 (±20.7) years, 49.2% were females and 68.5% were from the white ethnic group. Out of these, 571 patients had their first positive PCR tests while hospitalised (62.1%), and 97 of these occurred at least seven days after admission (10.5%). Amongst the 5596 healthcare workers, 615 (11.0%) tested positive during the study period using PCR or serological tests. The mean age was 39.5 (±11.1) years, 78.9% were females and 49.8% were nurses. For susceptible patients, one day in the same ward with another patient with hospital-acquired SARS-CoV-2 was associated with an additional 7.5 infections per 1000 susceptible patients (95% Credible Interval (CrI) 5.5 to 9.5/1000 susceptible patients/day) per day. Exposure to an infectious patient with community-acquired COVID-19 or to an infectious healthcare worker was associated with substantially lower infection risks (2.0/1000 susceptible patients/day, 95%CrI 1.6 to 2.2). As for healthcare worker infections, exposure to an infectious patient with hospital-acquired SARS-CoV-2 or to an infectious healthcare worker were both associated with an additional 0.8 infection per 1000 susceptible healthcare workers per day (95%CrI 0.3 to 1.6 and 0.6 to 1.0 respectively). Exposure to an infectious patient with community-acquired SARS-CoV-2 was associated with less than half this risk (0.2/1000 susceptible healthcare workers/day, 95%CrI 0.2 to 0.2). These assumptions were tested in sensitivity analysis which showed broadly similar results. The main limitations were

that the symptom onset dates and healthcare worker absence days were not available.

Conclusions: In this study, we observed that exposure to patients with hospital-acquired SARS-CoV-2 is associated with a substantial infection risk to both healthcare workers and other hospitalised patients. Infection control measures to limit nosocomial transmission must be optimised to protect both staff and patients from SARS-CoV-2 infection.

D.1.2 Duration of carbapenemase-producing Enterobacteriaceae carriage in hospital patients

Mo Y, Hernandez-Koutoucheva A, Musicha P, Bertrand D, Lye D, Ng O, et al. Duration of Carbapenemase-Producing Enterobacteriaceae Carriage in Hospital Patients. Emerg Infect Dis. 2020;26(9):2182-2185.

Carriage duration of carbapenemase-producing Enterobacteriaceae (CPE) is uncertain. We followed 21 CPE carriers over one year. Mean carriage duration was 86 (95%CrI 60 to 122) days, with 98.5% (95%CrI 95.0 to 99.8) probability of decolonization in one year, suggesting that CPE-carriers' status can be reviewed yearly. Antibiotic consumption was associated with prolonged carriage.

References

[1] Eili Y Klein et al. "Assessment of WHO antibiotic consumption and access targets in 76 countries, 2000–15: an analysis of pharmaceutical sales data". In: *The Lancet Infectious Diseases* 21.1 (2021), pp. 107–115.

[2] World Health Organization(WHO). *No time to wait: securing the future from drug-resistant infections report to the secretary-general of the United Nation.* Tech. rep. 2019. URL: https://www.who.int/docs/default-source/documents/no-time-to-wait-securing-the-future-from-drug-resistant-infections-en.pdf?sfvrsn=5b424d7%7B%5C_%7D6.

[3] Alison H Holmes et al. "Understanding the mechanisms and drivers of antimicrobial resistance". In: *The Lancet* 387.10014 (2016), pp. 176–187.

[4] OJ Dyar et al. "What is antimicrobial stewardship?" In: *Clinical Microbiology and Infection* 23.11 (2017), pp. 793–798.

[5] Public Health England. *Antimicrobial Resistance Resource Handbook.* Tech. rep. 2016. URL: https://www.gov.uk/government/publications/antimicrobial-resistance-resource-handbook.

[6] World Health Organization(WHO). *Mapping educational opportunities and resources for health-care workers to learn about antimicrobial resistance and stewardship.* Tech. rep. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/259362/9789241512787-eng.pdf.

[7] Tamar F Barlam et al. "Implementing an antibiotic stewardship program: guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America". In: *Clinical Infectious Diseases* 62.10 (2016), e51–e77.

[8] V.A. Schweitzer et al. "The quality of studies evaluating antimicrobial stewardship interventions: a systematic review". In: *Clin. Microbiol. Infect.* 25.5 (May 2019), pp. 555–561.

[9] Louis B Rice. "The Maxwell Finland Lecture: For the duration—rational antibiotic administration in an era of antimicrobial resistance and Clostridium difficile". In: *Clinical Infectious Diseases* 46.4 (2008), pp. 491–496.

[10] Martin J Llewelyn et al. "The antibiotic course has had its day". In: *BMJ* 358 (2017).

[11] Edward P Abraham et al. "Further observations on penicillin". In: *The Lancet* 238.6155 (1941), pp. 177–189.

[12] Chester S Keefer et al. "Penicillin in the treatment of infections: a report of 500 cases". In: *Journal of the American Medical Association* 122.18 (1943), pp. 1217–1224.

[13] William S Tillett, Margaret J Cambier, and James E McCormack. “The treatment of lobar pneumonia and pneumococcal empyema with penicillin”. In: *Bulletin of the New York Academy of Medicine* 20.3 (1944), p. 142.

[14] Martin Henry Dawson and Gladys L Hobby. “The clinical use of penicillin: observations in one hundred cases”. In: *Journal of the American Medical Association* 124.10 (1944), pp. 611–622.

[15] Manson Meads et al. “Treatment of pneumococcal pneumonia with penicillin”. In: *New England Journal of Medicine* 232.26 (1945), pp. 747–755.

[16] Michael S Niederman et al. “Guidelines for the management of adults with community-acquired pneumonia: diagnosis, assessment of severity, antimicrobial therapy, and prevention”. In: *American journal of respiratory and critical care medicine* 163.7 (2001), pp. 1730–1754.

[17] M Woodhead et al. “Guidelines for the management of adult lower respiratory tract infections-full version”. In: *Clinical microbiology and infection* 17 (2011), E1–E59.

[18] Wei Shen Lim et al. “BTS guidelines for the management of community acquired pneumonia in adults: update 2009”. In: *Thorax* 64.Suppl 3 (2009), pp. iii1–iii55.

[19] Alexander Fleming. “On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae”. In: *British Journal of experimental pathology* 10.3 (1929), p. 226.

[20] Inga Odenholt et al. “Suboptimal antibiotic dosage as a risk factor for selection of penicillin-resistant Streptococcus pneumoniae: in vitro kinetic model”. In: *Antimicrobial Agents and Chemotherapy* 47.2 (2003), pp. 518–523.

[21] Ivan Surya Pradipta et al. “Risk factors of multidrug-resistant tuberculosis: A global systematic review and meta-analysis”. In: *Journal of Infection* 77.6 (2018), pp. 469–478.

[22] David Nora et al. “Biomarker-guided antibiotic therapy—strengths and limitations”. In: *Annals of translational medicine* 5.10 (2017).

[23] Matthew J Page et al. “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”. In: *BMJ* 372 (2021).

[24] Wichor M Bramer et al. “Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study”. In: *Systematic reviews* 6.1 (2017), pp. 1–12.

[25] Jonathan AC Sterne et al. “RoB 2: a revised tool for assessing risk of bias in randomised trials”. In: *BMJ* 366 (2019).

[26] Maxime Vallée and Franck Bruyère. “Effect of 5-Day Nitrofurantoin vs Single-dose Fosfomycin on Clinical Resolution of Uncomplicated Lower Urinary Tract Infection in Women: A Randomized Clinical Trial. JAMA 2018; 319: 1781-9”. In: *European urology* 74.5 (2018), e124.

[27] Joshua P Metlay et al. “Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America”. In: *American Journal of Respiratory and Critical Care Medicine* 200.7 (2019), e45–e67.

[28] The National Institute for Health and Care Excellence (NICE). *Urinary tract infection (lower): antimicrobial prescribing NICE guideline*. Tech. rep. 2018. URL: www.nice.org.uk/guidance/ng109.

[29] Stijn W de Jonge et al. "Effect of postoperative continuation of antibiotic prophylaxis on the incidence of surgical site infection: a systematic review and meta-analysis". In: *The Lancet Infectious Diseases* 20.10 (2020), pp. 1182–1192.

[30] Gunilla Skoog Ståhlgren et al. "Penicillin V four times daily for five days versus three times daily for 10 days in patients with pharyngotonsillitis caused by group A streptococci: randomised controlled, open label, non-inferiority study". In: *BMJ* 367 (2019).

[31] Matthew E Falagas et al. "Effectiveness and safety of short-course vs long-course antibiotic therapy for group A β-hemolytic streptococcal tonsillopharyngitis: A meta-analysis of randomized trials". In: *Mayo Clinic Proceedings*. Vol. 83. 8. Elsevier. 2008, pp. 880–889.

[32] Jonathan R Carapetis et al. "The global burden of group A streptococcal diseases". In: *The Lancet Infectious Diseases* 5.11 (2005), pp. 685–694.

[33] Matthew E Falagas et al. "Effectiveness and safety of short vs. long duration of antibiotic therapy for acute bacterial sinusitis: a meta-analysis of randomized trials". In: *British Journal of Clinical Pharmacology* 67.2 (2009), pp. 161–171.

[34] Anthony W Chow et al. "IDSA clinical practice guideline for acute bacterial rhinosinusitis in children and adults". In: *Clinical Infectious Diseases* 54.8 (2012), e72–e112.

[35] The National Institute for Health and Care Excellence (NICE). *Sinusitis (acute): antimicrobial prescribing NICE guideline*. Tech. rep. 2017. URL: www.nice.org.uk/guidance/ng79.

[36] DE Karageorgopoulos et al. "Short versus long duration of antibiotic therapy for bacterial meningitis: a meta-analysis of randomised controlled trials in children". In: *Archives of Disease in Childhood* 94.8 (2009), pp. 607–614.

[37] Elizabeth Molyneux et al. "5 versus 10 days of treatment with ceftriaxone for bacterial meningitis in children: a double-blind randomised equivalence study". In: *The Lancet* 377.9780 (2011), pp. 1837–1845.

[38] Matthew D Snape and Dominic F Kelly. "Fine with five? Shorter antibiotic courses for childhood meningitis." In: *Lancet (London, England)* 377.9780 (2011), pp. 1809–1810.

[39] G Ralph Corey. "Staphylococcus aureus bloodstream infections: definitions and treatment". In: *Clinical Infectious Diseases* 48.Supplement_4 (2009), S254–S259.

[40] Andre C. Kalil et al. "Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society". In: *Clinical Infectious Diseases* 63.5 (Sept. 1, 2016), e61–e111.

[41] Thomas L Holland et al. "Effect of algorithm-based therapy vs usual care on clinical success and serious adverse events in patients with staphylococcal bacteremia: a randomized clinical trial". In: *JAMA* 320.12 (2018), pp. 1249–1258.

[42] Michael W Dunne et al. “A randomized clinical trial of single-dose versus weekly dalbavancin for treatment of acute bacterial skin and skin structure infection”. In: *Clinical Infectious Diseases* 62.5 (2016), pp. 545–551.

[43] Cédric Daubin et al. “Procalcitonin algorithm to guide initial antibiotic therapy in acute exacerbations of COPD admitted to the ICU: a randomized multicenter study”. In: *Intensive care medicine* 44.4 (2018), pp. 428–437.

[44] Amy-Sarah Ginsburg et al. “Amoxicillin for 3 or 5 days for chest-indrawing pneumonia in Malawian children”. In: *New England Journal of Medicine* 383.1 (2020), pp. 13–23.

[45] Jean Chastre et al. “Comparison of 8 vs 15 Days of Antibiotic Therapy for Ventilator-Associated Pneumonia in Adults”. In: *JAMA* 290.19 (Nov. 19, 2003), p. 2588.

[46] Adriano Arguedas et al. “Single-dose extended-release azithromycin versus a 10-day regimen of amoxicillin/clavulanate for the treatment of children with acute otitis media”. In: *International Journal of Infectious Diseases* 15.4 (2011), e240–e248.

[47] Alejandro Hoberman et al. “Shortened antimicrobial treatment for acute otitis media in young children”. In: *New England Journal of Medicine* 375.25 (2016), pp. 2446–2456.

[48] Mohd Z Mazlan et al. “Efficacy and safety of the point-of-care procalcitonin test for determining the antibiotic treatment duration in patients with ventilator-associated pneumonia in the intensive care unit: a randomised controlled trial”. In: *Anaesthesiology Intensive Therapy* 53.1 (2021).

[49] Manuela Aguilar-Guisado et al. “Optimisation of empirical antimicrobial therapy in patients with haematological malignancies and febrile neutropenia (How Long study): an open-label, randomised, controlled phase 4 trial”. In: *The Lancet Haematology* 4.12 (2017), e573–e583.

[50] Gilles Capellier et al. “Early-Onset Ventilator-Associated Pneumonia in Adults Randomized Clinical Trial: Comparison of 8 versus 15 Days of Antibiotic Treatment”. In: *PLoS ONE* 7.8 (Aug. 31, 2012). Ed. by Brad Spellberg, e41290.

[51] Gilda Piaggio et al. “Reporting of Noninferiority and Equivalence Randomized Trials”. In: *JAMA* 308.24 (Dec. 26, 2012), p. 2594.

[52] Jennifer Schumi and Janet T Wittes. “Through the looking glass: understanding non-inferiority”. In: *Trials* 12.1 (2011), pp. 1–12.

[53] Dennis A Revicki and Lori Frank. “Pharmacoeconomic evaluation in the real world”. In: *Pharmacoeconomics* 15.5 (1999), pp. 423–434.

[54] U.S. Department of Health et al. *Non-Inferiority Clinical Trials to Establish Effectiveness Guidance for Industry*. Tech. rep. 2016.

[55] *The ethics of research related to healthcare in developing countries*. Apr. 2002. URL: https://www.nuffieldbioethics.org/publications/research-in-developing-countries.

[56] Infectious Diseases Society of America. “Position paper: recommended design features of future clinical trials of antibacterial agents for community-acquired pneumonia”. In: *Clinical Infectious Diseases* 47.Supplement_3 (2008), S249–S265.

[57] Sanjay Kaul, George A Diamond, and William S Weintraub. "Trials and tribulations of non-inferiority: the ximelagatran experience". In: *Journal of the American College of Cardiology* 46.11 (2005), pp. 1986–1995.

[58] Pranita D Tamma et al. "Association of adverse events with antibiotic use in hospitalized patients". In: *JAMA Internal Medicine* 177.9 (2017), pp. 1308–1315.

[59] Anthony Jourdan et al. "Antibiotic hypersensitivity and adverse reactions: management and implications in clinical practice". In: *Allergy, Asthma & Clinical Immunology* 16.1 (2020), pp. 1–7.

[60] Helena Seppälä et al. "The effect of changes in the consumption of macrolide antibiotics on erythromycin resistance in group A streptococci in Finland". In: *New England Journal of Medicine* 337.7 (1997), pp. 441–446.

[61] David C Bean et al. "Resistance among Escherichia coli to sulphonamides and other antimicrobials now little used in man". In: *Journal of Antimicrobial Chemotherapy* 56.5 (2005), pp. 962–964.

[62] Sunil Pokharel, Shristi Raut, and Bipin Adhikari. "Tackling antimicrobial resistance in low-income and middle-income countries". In: *BMJ Global Health* 4.6 (2019). eprint: https://gh.bmj.com/content/4/6/e002104.full.pdf. URL: https://gh.bmj.com/content/4/6/e002104.

[63] M El Setouhy et al. "Ethics: Fair benefits for research in developing countries: Participants in the 2001 conference on ethical aspects of research in developing countries". In: *Science* 298.5601 (2002), pp. 2133–2134.

[64] Somphou Sayasone et al. "Efficacy and safety of tribendimidine versus praziquantel against Opisthorchis viverrini in Laos: an open-label, randomised, non-inferiority, phase 2 trial". In: *The Lancet Infectious Diseases* 18.2 (2018), pp. 155–161.

[65] Daniela Munteanu et al. "Efficacy and safety of sequential versus quadruple therapy as second-line treatment for helicobacter pylori infection—A randomized controlled trial". In: *PloS one* 12.9 (2017), e0183302.

[66] The BREATHER (PENTA 16) Trial Group. "Weekends-off efavirenz-based antiretroviral therapy in HIV-infected children, adolescents, and young adults (BREATHER): a randomised, open-label, non-inferiority, phase 2/3 trial". In: *The lancet HIV* 3.9 (2016), e421–e430.

[67] Chloe Orkin et al. "Doravirine/lamivudine/tenofovir disoproxil fumarate is non-inferior to efavirenz/emtricitabine/tenofovir disoproxil fumarate in treatment-naive adults with human immunodeficiency virus–1 infection: week 48 results of the DRIVE-AHEAD trial". In: *Clinical Infectious Diseases* 68.4 (2019), pp. 535–544.

[68] M Matilde Sanchez and Xun Chen. "Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat". In: *Statistics in Medicine* 25.7 (2006), pp. 1169–1181.

[69] Miguel A Hernán, James M Robins, et al. "Per-protocol analyses of pragmatic trials". In: *New England Journal of Medicine* 377.14 (2017), pp. 1391–1398.

[70] Ying Wu et al. “Correcting for non-compliance in randomized non-inferiority trials with active and placebo control using structural models”. In: *Statistics in Medicine* 34.6 (2015), pp. 950–965.

[71] Sameer Parpia et al. “Treatment crossovers in time-to-event non-inferiority randomised trials of radiotherapy in patients with breast cancer”. In: *BMJ Open* 4.10 (2014), e006531.

[72] Dan Sheng and Mimi Y Kim. “The effects of non-compliance on intent-to-treat analysis of equivalence trials”. In: *Statistics in Medicine* 25.7 (2006), pp. 1183–1199.

[73] Committee for Proprietary Medicinal Products (CPMP et al. “Points to consider on switching between superiority and non-inferiority”. In: *British Journal of Clinical Pharmacology* 52.3 (2001), p. 223.

[74] Mimi Y Kim. “Using the instrumental variables estimator to analyze noninferiority trials with noncompliance”. In: *Journal of Biopharmaceutical Statistics* 20.4 (2010), pp. 745–758.

[75] Michael O Harhay et al. “Outcomes and statistical power in adult critical care randomized trials”. In: *American Journal of Respiratory and Critical Care Medicine* 189.12 (2014), pp. 1469–1478.

[76] Rasha Al-Lamee et al. “Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomised controlled trial”. In: *The Lancet* 391.10115 (2018), pp. 31–40.

[77] Kasia Stepniewska and Nicholas J White. “Some considerations in the design and interpretation of antimalarial drug trials in uncomplicated falciparum malaria”. In: *Malaria Journal* 5.1 (2006), pp. 1–15.

[78] Kasia Stepniewska et al. “In vivo assessment of drug efficacy against Plasmodium falciparum malaria: duration of follow-up”. In: *Antimicrobial Agents and Chemotherapy* 48.11 (2004), pp. 4271–4280.

[79] Noortje van Herwaarden et al. “Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis: open label, randomised controlled, non-inferiority trial”. In: *BMJ* 350 (2015).

[80] Sunita Rehal et al. “Non-inferiority trials: are they inferior? A systematic review of reporting in major medical Journals”. In: *BMJ Open* 6.10 (2016), e012594.

[81] Committee for Proprietary Medicinal Products. *The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs*. Tech. rep. European Medicines Agency, 2005.

[82] Iosief Abraha et al. “Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study”. In: *BMJ* 350 (2015).

[83] Alessandro Montedori et al. “Modified versus standard intention-to-treat reporting: are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study”. In: *Trials* 12.1 (2011), pp. 1–9.

[84] Ane Uranga et al. “Duration of antibiotic treatment in community-acquired pneumonia: a multicenter randomized clinical trial”. In: *JAMA Internal Medicine* 176.9 (2016), pp. 1257–1265.

[85] William C Blackwelder. ““Proving the null hypothesis” in clinical trials”. In: *Controlled Clinical Trials* 3.4 (1982), pp. 345–353.

[86] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2020. URL: https://www.R-project.org/.

[87] MA Hernán and JM Robins. *Causal Inference: What If*. Boca Raton: Chapman and Hall/CRC, 2020.

[88] Miguel A Hernán, John Hsu, and Brian Healy. “A second chance to get causal inference right: a classification of data science tasks”. In: *Chance* 32.1 (2019), pp. 42–49.

[89] Olli S Miettinen. “Stratification by a multivariate confounder score”. In: *American Journal of Epidemiology* 104.6 (1976), pp. 609–620.

[90] Mohammad Ali Mansournia and Douglas G. Altman. “Inverse probability weighting”. In: *BMJ (Online)* 352 (Jan. 15, 2016).

[91] Stanley Xu et al. “Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals”. In: *Value in Health* 13.2 (2010), pp. 273–277.

[92] Joshua D Angrist, Guido W Imbens, and Donald B Rubin. “Identification of causal effects using instrumental variables”. In: *Journal of the American Statistical Association* 91.434 (1996), pp. 444–455.

[93] James Robins and Andrea Rotnitzky. “Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models”. In: *Biometrika* 91.4 (2004), pp. 763–783.

[94] Paul S Clarke and Frank Windmeijer. “Instrumental variable estimators for binary outcomes”. In: *Journal of the American Statistical Association* 107.500 (2012), pp. 1638–1652.

[95] SD Walter et al. “A new preference-based analysis for randomized trials can estimate treatment acceptability and effect in compliant patients”. In: *Journal of Clinical Epidemiology* 59.7 (2006), pp. 685–696.

[96] Kara E Rudolph and Elizabeth A Stuart. “Using sensitivity analyses for unobserved confounding to address covariate measurement error in propensity score methods”. In: *American Journal of Epidemiology* 187.3 (2018), pp. 604–613.

[97] Til Stürmer et al. “Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration”. In: *American Journal of Epidemiology* 162.3 (2005), pp. 279–289.

[98] Stijn Vansteelandt and Vanessa Didelez. “Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators”. In: *Scandinavian Journal of Statistics* 45.4 (2018), pp. 941–961.

[99] Miguel A Hernán et al. “Comparison of dynamic treatment regimes via inverse probability weighting”. In: *Basic & Clinical Pharmacology & Toxicology* 98.3 (2006), pp. 237–242.

[100] Shaun R Seaman and Ian R White. "Review of inverse probability weighting for dealing with missing data". In: *Statistical Methods in Medical Research* 22.3 (2013), pp. 278–295.

[101] Heejung Bang and James M Robins. "Doubly robust estimation in missing data and causal inference models". In: *Biometrics* 61.4 (2005), pp. 962–973.

[102] Scott K Aberegg, Andrew M Hersh, and Matthew H Samore. "Do non-inferiority trials of reduced intensity therapies show reduced effects? A descriptive analysis". In: *BMJ Open* 8.3 (2018), e019494.

[103] Caroline Kistin and Michael Silverstein. "Pilot studies: a critical but potentially misused component of interventional research". In: *JAMA* 314.15 (2015), pp. 1561–1562.

[104] Amanda Jane Blatch-Jones et al. "Role of feasibility and pilot studies in randomised controlled trials: a cross-sectional study". In: *BMJ Open* 8.9 (2018), e022233.

[105] Jared K Lunceford and Marie Davidian. "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study". In: *Statistics in Medicine* 23.19 (2004), pp. 2937–2960.

[106] James M Robins and Dianne M Finkelstein. "Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests". In: *Biometrics* 56.3 (2000), pp. 779–788.

[107] Miguel A Hernán et al. "Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology". In: *American Journal of Epidemiology* 155.2 (2002), pp. 176–184.

[108] Tyler J VanderWeele. "Principles of confounder selection". In: *European Journal of Epidemiology* 34.3 (2019), pp. 211–219.

[109] Daniel O Scharfstein, Andrea Rotnitzky, and James M Robins. "Adjusting for nonignorable drop-out using semiparametric nonresponse models". In: *Journal of the American Statistical Association* 94.448 (1999), pp. 1096–1120.

[110] Yin Mo et al. "Reducing antibiotic treatment duration for ventilator-associated pneumonia (REGARD-VAP): a trial protocol for a randomised clinical trial". In: *BMJ open* 11.5 (2021), e050105.

[111] Axel Grothey et al. "Duration of adjuvant chemotherapy for stage III colon cancer". In: *New England Journal of Medicine* 378.13 (2018), pp. 1177–1188.

[112] Jean-Louis Vincent et al. "International Study of the Prevalence and Outcomes of Infection in Intensive Care Units". In: *JAMA* 302.21 (Dec. 2, 2009), p. 2323.

[113] A E Stebbings, T Y Ti, and W C Tan. "Hospital acquired pneumonia in the medical intensive care unit–a prospective study." In: *Singapore Medical Journal* 40.8 (Aug. 1999), pp. 508–12.

[114] Yaseen Arabi et al. "Ventilator-associated pneumonia in adults in developing countries: a systematic review". In: *International Journal of Infectious Diseases* 12.5 (Sept. 2008), pp. 505–512.

[115] N Fàbregas et al. "Clinical diagnosis of ventilator associated pneumonia revisited: comparative validation using immediate post-mortem lung biopsies." In: *Thorax* 54.10 (Oct. 1999), pp. 867–73.

[116] J D Hill et al. “Pulmonary pathology in acute respiratory insufficiency: lung biopsy as a diagnostic tool.” In: *The Journal of Thoracic and Cardiovascular Surgery* 71.1 (Jan. 1976), pp. 64–71.

[117] Nina Singh et al. “Short-course Empiric Antibiotic Therapy for Patients with Pulmonary Infiltrates in the Intensive Care Unit”. In: *American Journal of Respiratory and Critical Care Medicine* 162.2 (Aug. 2000), pp. 505–511.

[118] Scott T Micek et al. “A randomized controlled trial of an antibiotic discontinuation policy for clinically suspected ventilator-associated pneumonia.” In: *Chest* 125.5 (May 2004), pp. 1791–9.

[119] Christian Brun-Buisson et al. “Contribution of Blinded, Protected Quantitative Specimens to the Diagnostic and Therapeutic Management of Ventilator-Associated Pneumonia”. In: *Chest* 128.2 (Aug. 2005), pp. 533–544.

[120] Kirthana Raman et al. “Early Antibiotic Discontinuation in Patients With Clinically Suspected Ventilator-Associated Pneumonia and Negative Quantitative Bronchoscopy Cultures*”. In: *Critical Care Medicine* 41.7 (July 2013), pp. 1656–1663.

[121] Ronald N. Jones. “Microbial Etiologies of Hospital-Acquired Bacterial Pneumonia and Ventilator-Associated Bacterial Pneumonia”. In: *Clinical Infectious Diseases* 51.S1 (Aug. 1, 2010), S81–S87.

[122] P. Rattanaumpawan et al. “Randomized controlled trial of nebulized colistimethate sodium as adjunctive therapy of ventilator-associated pneumonia caused by Gram-negative bacteria”. In: *Journal of Antimicrobial Chemotherapy* 65.12 (Dec. 1, 2010), pp. 2645–2649.

[123] Juthamas Inchai et al. “Ventilator-Associated Pneumonia: Epidemiology and Prognostic Indicators of 30-Day Mortality”. In: *Japanese Journal of Infectious Diseases* 68.3 (2015), pp. 181–186.

[124] U.S. Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN). *Pneumonia (Ventilator-associated [VAP] and nonventilator-associated Pneumonia [PNEU]) Event*. Tech. rep. 2018.

[125] Alan E Jones, Stephen Trzeciak, and Jeffrey A Kline. “The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation.” In: *Critical Care Medicine* 37.5 (May 2009), pp. 1649–54.

[126] Maarten Bekaert et al. “Attributable mortality of ventilator-associated pneumonia: a reappraisal using causal analysis.” In: *American Journal of Respiratory and Critical Care Medicine* 184.10 (Nov. 15, 2011), pp. 1133–9.

[127] The Canadian Critical Care Trials Group. “A Randomized Trial of Diagnostic Techniques for Ventilator-Associated Pneumonia”. In: *New England Journal of Medicine* 355.25 (Dec. 21, 2006), pp. 2619–2630.

[128] Danilo Cortozi Berton, Andre C Kalil, and Paulo José Zimermann Teixeira. “Quantitative versus qualitative cultures of respiratory secretions for clinical outcomes in patients with ventilator-associated pneumonia”. In: *Cochrane Database of Systematic Reviews* 10 (Oct. 30, 2014), p. CD006482.

[129] European Society of Clinical Microbiology and Infectious Diseases. *The European Committee on Antimicrobial Susceptibility Testing. Routine and extended internal quality control for MIC determination and disk diffusion as recommended by EUCAST. Version 8.0.* Tech. rep. 2018.

[130] *M100: Antimicrobial Susceptibility Testing Standards.* [Online; accessed 2019-10-31]. 0. URL: https://clsi.org/standards/products/microbiology/documents/m100/.

[131] Alain Combes et al. "Factors predicting ventilator-associated pneumonia recurrence". In: *Critical Care Medicine* 31.4 (Apr. 1, 2003), pp. 1102–1107.

[132] J R Zahar et al. "Is methicillin resistance associated with a worse prognosis in Staphylococcus aureus ventilator-associated pneumonia?" In: *Clinical Infectious Diseases* 41.1537-6591 (Electronic) (Nov. 1, 2005), pp. 1224–1231.

[133] Alain Combes et al. "Early predictors for infection recurrence and death in patients with ventilator-associated pneumonia". In: *Critical Care Medicine* 35.1 (Jan. 2007), pp. 146–154.

[134] Wendy L. Wahl et al. "Does bronchoalveolar lavage enhance our ability to treat ventilator-associated pneumonia in a trauma-burn intensive care unit?" In: *Journal of Trauma* 54.4 (2003), pp. 633–639.

[135] François Stéphan et al. "Ventilator-associated pneumonia leading to acute lung injury after trauma: Importance of Haemophilus influenzae". In: *Anesthesiology* 104.2 (Feb. 2006), pp. 235–241.

[136] Alain Combes et al. "Incidence and outcome of polymicrobial ventilator-associated pneumonia". In: *Chest* 121.5 (2002), pp. 1618–1623.

[137] Charles Edouard Luyt et al. "Procalcitonin kinetics as a prognostic marker of ventilator-associated pneumonia". In: *American Journal of Respiratory and Critical Care Medicine* 171.1 (Jan. 1, 2005), pp. 48–53.

[138] Moi Lin Ling, Anucha Apisarnthanarak, and Gilbert Madriaga. "The burden of healthcare-associated infections in southeast Asia: A systematic literature review and meta-analysis". In: *Clinical Infectious Diseases* 60.11 (2015), pp. 1690–1699.

[139] Ilias I. Siempos, Zoe Athanassa, and Matthew E. Falagas. "Frequency and predictors of ventilator-associated pneumonia recurrence: A meta-analysis". In: *Shock* 30.5 (Nov. 2008), pp. 487–495.

[140] Center for Biologics Evaluation, Center for Drug Evaluation Research (CBER), and Research (CDER). *Guidance for Industry Non-Inferiority Clinical Trials.* Tech. rep. 2010.

[141] Mark D. Rothmann, Brian L. Wiens, and Ivan S.F. Chan. *Design and Analysis of Non-Inferiority Trials.* Chapman and Hall/CRC, 2011.

[142] Sylvain Durrleman and Richard Simon. "Planning and Monitoring of Equivalence Studies". In: *Biometrics* 46.2 (June 1990), p. 329.

[143] Yin Mo et al. "Statistical considerations in the design and analysis of non-inferiority trials with binary endpoints in the presence of non-adherence: a simulation study". In: *Wellcome Open Research* 4 (Dec. 18, 2019), p. 207.

[144] ELSEVIER. "MACRO Electronic Data Capture". In: (2019).

[145] Christopher Jennison and Bruce W. Turnbull. *Group sequential methods with applications to clinical trials*. Chapman and Hall/CRC, 2000.

[146] Stuart J. Pocock. "Group Sequential Methods in the Design and Analysis of Clinical Trials". In: *Biometrika* 64.2 (Aug. 1977), p. 191.

[147] Yin Mo et al. "Non-adherence in non-inferiority trials: Pitfalls and recommendations". In: *The BMJ* 370 (July 1, 2020).

[148] D. Stolz et al. "Procalcitonin for reduced antibiotic exposure in ventilator-associated pneumonia: A randomised study". In: *European Respiratory Journal* 34.6 (Dec. 2009), pp. 1364–1375.

[149] Sandra M Eldridge et al. "CONSORT 2010 statement : extension to randomised pilot and feasibility trials The Consolidated Standards of Reporting Trials (CONSORT) statement reporting of randomised controlled an extension to that statement for". In: *BMJ* 355 (2016), pp. 1–29.

[150] An-Wen Chan et al. "SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials." In: *BMJ (Clinical research ed.)* 346 (Jan. 8, 2013), e7586.

[151] U Galbraith et al. "Antibiotic prophylaxis in cardiovascular surgery: a prospective randomized comparative trial of one day cefazolin versus single dose cefuroxime." In: *Drugs under experimental and clinical research* 19.5 (1993), pp. 229–234.

[152] Hung-Chi Lue et al. "Three-versus four-week administration of benzathine penicillin G: effects on incidence of streptococcal infections and recurrences of rheumatic fever". In: *Pediatrics* 97.6 (1996), pp. 984–988.

[153] Evelien de Jong et al. "Efficacy and safety of procalcitonin guidance in reducing the duration of antibiotic treatment in critically ill patients: a randomised, controlled, open-label trial". In: *The Lancet Infectious Diseases* 16.7 (2016), pp. 819–827.

[154] Theodore Gouliouris et al. "Duration of exposure to multiple antibiotics is associated with increased risk of VRE bacteraemia: a nested case-control study". In: *Journal of Antimicrobial Chemotherapy* 73.6 (2018), pp. 1692–1699.

[155] Tjeerd Pieter Van Staa et al. "The effectiveness of frequent antibiotic use in reducing the risk of infection-related hospital admissions: results from two large population-based cohorts". In: *BMC medicine* 18.1 (2020), pp. 1–11.

[156] Danping Zheng, Timur Liwinski, and Eran Elinav. "Interaction between microbiota and immunity in health and disease". In: *Cell research* 30.6 (2020), pp. 492–506.

[157] Infectious Diseases Society of America (IDSA). "White paper: recommendations on the conduct of superiority and organism-specific clinical trials of antibacterial agents for the treatment of infections caused by drug-resistant bacterial pathogens". In: *Clinical Infectious Diseases* 55.8 (2012), pp. 1031–1046.

[158] Richard G Wunderink et al. "Linezolid in methicillin-resistant Staphylococcus aureus nosocomial pneumonia: a randomized, controlled study". In: *Clinical Infectious Diseases* 54.5 (2012), pp. 621–629.

[159] Center for Drug Evaluation and Research. *Guidance for Industry Hospital-Acquired Bacterial Pneumonia and Ventilator-Associated Bacterial Pneumonia: Developing Drugs for Treatment.* May 2014. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/hospital-acquired-bacterial-pneumonia-and-ventilator-associated-bacterial-pneumonia-developing-drugs.

[160] Brad Spellberg and George Talbot. "Recommended design features of future clinical trials of antibacterial agents for hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia". In: *Clinical Infectious Diseases* 51.Supplement_1 (2010), S150–S170.

[161] Åsa Sullivan, Charlotta Edlund, and Carl Erik Nord. "Effect of antimicrobial agents on the ecological balance of human microflora". In: *The Lancet infectious diseases* 1.2 (2001), pp. 101–114.

[162] Christine Tedijanto et al. "Estimating the proportion of bystander selection for antibiotic resistance among potentially pathogenic bacterial flora". In: *Proceedings of the National Academy of Sciences* 115.51 (2018), E11988–E11995.

[163] Gilles Troché et al. "Detection and treatment of antibiotic-resistant bacterial carriage in a surgical intensive care unit: a 6-year prospective survey". In: *Infection Control & Hospital Epidemiology* 26.2 (2005), pp. 161–165.

[164] Biana Dubinsky-Pertzov et al. "Carriage of extended-spectrum beta-lactamase–producing Enterobacteriaceae and the risk of surgical site infection after colorectal surgery: a prospective cohort study". In: *Clinical Infectious Diseases* 68.10 (2019), pp. 1699–1704.

[165] Pravani Reddy et al. "Screening for extended-spectrum β-lactamase-producing Enterobacteriaceae among high-risk patients and rates of subsequent bacteremia". In: *Clinical Infectious Diseases* 45.7 (2007), pp. 846–852.

[166] Osman Balci. "Validation, verification, and testing techniques throughout the life cycle of a simulation study". In: *Annals of operations research* 53.1 (1994), pp. 121–173.

[167] Evelina Tacconelli et al. "Antibiotic usage and risk of colonization and infection with antibiotic-resistant bacteria: a hospital population-based study". In: *Antimicrobial Agents and Chemotherapy* 53.10 (2009), pp. 4264–4269.

[168] Yin Mo et al. "Duration of carbapenemase-producing Enterobacteriaceae carriage in hospital patients". In: *Emerging Infectious Diseases* 26.9 (2020), p. 2182.

[169] Sohn Kim, April Covington, and Eric G Pamer. "The intestinal microbiota: antibiotics, colonization resistance, and enteric pathogens". In: *Immunological reviews* 279.1 (2017), pp. 90–105.

[170] Rene Niehus et al. "Quantifying antibiotic impact on within-patient dynamics of extended-spectrum beta-lactamase resistance". In: *Elife* 9 (2020), e49206.

[171] Jon Iredell, Jeremy Brown, and Kaitlin Tagg. "Antibiotic resistance in Enterobacteriaceae: mechanisms and clinical implications". In: *BMJ* 352 (2016).

[172] Christopher Jackson. "Multi-state modelling with R: the msm package". In: *Cambridge, UK* (2007), pp. 1–53.

[173] A Chalom, CY Mandai, and PIKL Prado. "R Package" pse": parameter space exploration with Latin Hypercubes". In: *Comprehensive R Archive Network-CRAN* (2017).

[174] Kieran Alden et al. "Spartan: a comprehensive tool for understanding uncertainty in simulations of biological systems". In: *PLoS computational biology* 9.2 (2013), e1002916.

[175] Maris S Arcilla et al. "Import and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae by international travellers (COMBAT study): a prospective, multicentre cohort study". In: *The Lancet Infectious Diseases* 17.1 (2017), pp. 78–85.

[176] Sandra P van Tongeren et al. "Fecal microbiota composition and frailty". In: *Applied and environmental microbiology* 71.10 (2005), pp. 6438–6442.

[177] Emanuele Rinninella et al. "What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases". In: *Microorganisms* 7.1 (2019), p. 14.

[178] Albert Palleja et al. "Recovery of gut microbiota of healthy adults following antibiotic exposure". In: *Nature microbiology* 3.11 (2018), pp. 1255–1265.

[179] Liam P Shaw et al. "Modelling microbiome recovery after antibiotics using a stability landscape framework". In: *The ISME Journal* 13.7 (2019), pp. 1845–1856.

[180] Egija Zaura et al. "Same exposure but two radically different responses to antibiotics: resilience of the salivary microbiome versus long-term microbial shifts in feces". In: *MBio* 6.6 (2015), e01693–15.

[181] Beth Gibson et al. "The distribution of bacterial doubling times in the wild". In: *Proceedings of the Royal Society B* 285.1880 (2018), p. 20180789.

[182] Thomas Tängdén et al. "Foreign travel is a major risk factor for colonization with Escherichia coli producing CTX-M-type extended-spectrum β-lactamases: a prospective study with Swedish volunteers". In: *Antimicrobial Agents and Chemotherapy* 54.9 (2010), pp. 3564–3568.

[183] Markus Hilty et al. "Transmission dynamics of extended-spectrum β-lactamase–producing Enterobacteriaceae in the tertiary care hospital and the household setting". In: *Clinical Infectious Diseases* 55.7 (2012), pp. 967–975.

[184] Dien M Tran et al. "High prevalence of colonisation with carbapenem-resistant Enterobacteriaceae among patients admitted to Vietnamese hospitals: Risk factors and burden of disease". In: *Journal of Infection* 79.2 (2019), pp. 115–122.

[185] Haggai Bar-Yoseph et al. "Natural history and decolonization strategies for ESBL/carbapenem-resistant Enterobacteriaceae carriage: systematic review and meta-analysis". In: *Journal of Antimicrobial Chemotherapy* 71.10 (2016), pp. 2729–2739.

[186] LARS Kager et al. "Imipenem concentrations in colorectal surgery and impact on the colonic microflora". In: *Antimicrobial Agents and Chemotherapy* 33.2 (1989), pp. 204–208.

[187] K Bächer et al. "Multiple dose pharmacokinetics, safety, and effects on faecal microflora, of cefepime in healthy volunteers". In: *Journal of Antimicrobial Chemotherapy* 30.3 (1992), pp. 365–375.

[188] CE Nord et al. “Effect of piperacillin/tazobactam treatment on human bowel microflora”. In: *Journal of Antimicrobial Chemotherapy* 31.suppl_A (1993), pp. 61–65.

[189] Sumio Watanabe and Manfred Opper. “Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory.” In: *Journal of machine learning research* 11.12 (2010).

[190] Martyn Plummer et al. “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling”. In: *Proceedings of the 3rd international workshop on distributed statistical computing*. Vol. 124. 125.10. Vienna, Austria. 2003, pp. 1–10.

[191] Irja Lutsar et al. “Meropenem vs standard of care for treatment of neonatal late onset sepsis (NeoMero1): A randomised controlled trial”. In: *PloS one* 15.3 (2020), e0229380.

[192] Nurgul Ceran et al. “A randomized comparative study of single-dose fosfomycin and 5-day ciprofloxacin in female patients with uncomplicated lower urinary tract infections”. In: *Journal of Infection and Chemotherapy* 16.6 (2010), pp. 424–430.

[193] Tiny Van Merode et al. “Acute uncomplicated lower urinary tract infections in general practice: clinical and microbiological cure rates after three-versus five-day treatment with trimethoprim”. In: *The European journal of general practice* 11.2 (2005), pp. 55–58.

[194] Gordon Dow et al. “A prospective, randomized trial of 3 or 14 days of ciprofloxacin treatment for acute urinary tract infection in patients with spinal cord injury”. In: *Clinical Infectious Diseases* 39.5 (2004), pp. 658–664.

[195] Juan Carlos Palomino and Anandi Martin. “Drug resistance mechanisms in Mycobacterium tuberculosis”. In: *Antibiotics* 3.3 (2014), pp. 317–340.

[196] P Howard et al. “An international cross-sectional survey of antimicrobial stewardship programmes in hospitals”. In: *Journal of Antimicrobial Chemotherapy* 70.4 (2015), pp. 1245–1255.

[197] Poojan Shrestha et al. “Enumerating the economic cost of antimicrobial resistance per antibiotic consumed to inform the evaluation of interventions affecting their use”. In: *Antimicrobial Resistance & Infection Control* 7.1 (2018), pp. 1–9.

[198] Bastiaan H Wittekamp et al. “Decontamination strategies and bloodstream infections with antibiotic-resistant microorganisms in ventilated patients: a randomized clinical trial”. In: *JAMA* 320.20 (2018), pp. 2087–2098.

[199] Nele Brusselaers, Dirk Vogelaers, and Stijn Blot. “The rising problem of antimicrobial resistance in the intensive care unit”. In: *Annals of Intensive Care* 1.1 (2011), pp. 1–7.

[200] Marin H Kollef, Cindy W Hamilton, and Frank R Ernst. “Economic impact of ventilator-associated pneumonia in a large matched cohort”. In: *Infection Control & Hospital Epidemiology* 33.3 (2012), pp. 250–256.

[201] John G Muscedere, Andrew Day, and Daren K Heyland1. “Mortality, attributable mortality, and clinical events as end points for clinical trials of ventilator-associated pneumonia and hospital-acquired pneumonia”. In: *Clinical Infectious Diseases* 51.Supplement_1 (2010), S120–S125.

[202] Nguyen Thi Khanh Nhu et al. "Emergence of carbapenem-resistant Acinetobacter baumannii as the major cause of ventilator-associated pneumonia in intensive care unit patients at an infectious disease hospital in southern Vietnam". In: *Journal of Medical Microbiology* 63.Pt 10 (2014), p. 1386.

[203] Anupama Vasudevan et al. "Inappropriate empirical antimicrobial therapy for multidrug-resistant organisms in critically ill patients with pneumonia is not an independent risk factor for mortality: results of a prospective observational study of 758 patients". In: *Journal of Global Antimicrobial Resistance* 1.3 (2013), pp. 123–130.

[204] Wipa Reechaipichitkul et al. "Causative agents and resistance among hospital-acquired and ventilator-associated pneumonia patients at Srinagarind Hospital, northeastern Thailand". In: *Southeast Asian J Trop Med Public Health* 44.3 (2013), pp. 490–502.

[205] Pattarachai Kiratisin et al. "Comparative in vitro activity of carbapenems against major Gram-negative pathogens: results of Asia-Pacific surveillance from the COMPACT II study". In: *International Journal of Antimicrobial Agents* 39.4 (2012), pp. 311–316.

[206] A Clinton White Jr et al. "Effects of requiring prior authorization for selected antimicrobials: expenditures, susceptibilities, and clinical outcomes". In: *Clinical Infectious Diseases* 25.2 (1997), pp. 230–239.

[207] Amy L Pakyz, Michael Oinonen, and Ronald E Polk. "Relationship of carbapenem restriction in 22 university teaching hospitals to carbapenem use and carbapenem-resistant Pseudomonas aeruginosa". In: *Antimicrobial Agents and Chemotherapy* 53.5 (2009), pp. 1983–1986.

[208] G Jonathan Lewis et al. "Decreased resistance of Pseudomonas aeruginosa with restriction of ciprofloxacin in a large teaching hospital's intensive care and intermediate care units". In: *Infection Control & Hospital Epidemiology* 33.4 (2012), pp. 368–373.

[209] Jessina C McGregor and Jon P Furuno. "Optimizing research methods used for the evaluation of antimicrobial stewardship programs". In: *Clinical Infectious Diseases* 59.suppl_3 (2014), S185–S192.

[210] Emelie C Schuts et al. "Current evidence on hospital antimicrobial stewardship objectives: a systematic review and meta-analysis". In: *The Lancet Infectious Diseases* 16.7 (2016), pp. 847–856.

[211] Lidia Kardaś-Słoma et al. "Antibiotic reduction campaigns do not necessarily decrease bacterial resistance: the example of methicillin-resistant Staphylococcus aureus". In: *Antimicrobial Agents and Chemotherapy* 57.9 (2013), pp. 4410–4416.

[212] Lester Caudill and Joanna R Wares. "The role of mathematical modeling in designing and evaluating antimicrobial stewardship programs". In: *Current Treatment Options in Infectious Diseases* 8.2 (2016), pp. 124–138.

[213] Lionel A Mandell et al. "Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults". In: *Clinical Infectious Diseases* 44.Supplement_2 (2007), S27–S72.

[214] Michael D McKay, Richard J Beckman, and William J Conover. "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code". In: *Technometrics* 21.2 (1979), pp. 239–245.

[215] Andrea Saltelli and J Marivoet. "Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques". In: *Reliability Engineering & System Safety* 28.2 (1990), pp. 229–253.

[216] Suzana de Siqueira Santos et al. "A comparative study of statistical methods used to identify dependencies between gene expression signals". In: *Briefings in bioinformatics* 15.6 (2013), pp. 906–918.

[217] Simeone Marino et al. "A methodology for performing global uncertainty and sensitivity analysis in systems biology". In: *Journal of theoretical biology* 254.1 (2008), pp. 178–196.

[218] Christian Genest and Jean-François Plante. "On Blest's measure of rank correlation". In: *Canadian Journal of Statistics* 31.1 (2003), pp. 35–52.

[219] Mark Read et al. "Techniques for grounding agent-based simulations in the real domain: a case study in experimental autoimmune encephalomyelitis". In: *Mathematical and Computer Modelling of Dynamical Systems* 18.1 (2012), pp. 67–86.

[220] András Vargha and Harold D Delaney. "A critique and improvement of the CL common language effect size statistics of McGraw and Wong". In: *Journal of Educational and Behavioral Statistics* 25.2 (2000), pp. 101–132.