Measuring the impact of reduced antibiotic use in hospital settings using electronic health records

Eric Peter Budgell

Nuffield Department of Medicine

Linacre College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity 2023

Table of Contents

0.1 Abstract

Hospital inpatient antibiotic use has increased year-on-year in the NHS despite considerable investments in stewardship activity. Although hospitals have financial incentives to reduce overuse, current antibiotic reduction targets are modest and hospital stewardship leads have expressed significant concerns about the safety of achieving antimicrobial reduction targets, due to fears of under-treatment leading to patient harm. To reassure practitioners and the public that substantial reductions in antibiotic use could be safe, this thesis begins by exploring associations between variation in hospital-level antibiotic prescribing and patient-level clinical outcomes in 36 million general medicine admissions in England. I use this analysis to estimate what system-wide reductions might, at least in theory, be safely achieved if higher-using hospitals could replicate the prescribing practices of low-using hospitals. I then use patient-level antibiotic and outcome data to investigate the causal effect of using shorter versus longer antibiotic courses by treating time-varying antibiotic use as sequence of non-randomised nested "trials". Although observational analyses cannot prove shorter antibiotic courses are safe due to the potential for residual confounding, finding no evidence of harm can help to justify the evaluation of antibiotic reduction strategies in randomised experiments for unbiased inference. To inform its adoption in other acute care settings, the thesis then addresses two key questions emerging from the recently completed ARK-Hospital cluster-randomised trial, which tested a multifaceted behaviour change intervention to reduce antibiotic overuse in general medicine. Since the intervention effect varied widely by site, I examined the role of 26 contextual factors as potential mediators of impact, to better understand sources of this between-site variation. Finally, although the intervention had a modest impact on antibiotic prescribing trends overall, trends in non-ARK sites are unknown, and I therefore compared ward-level prescribing trends in ARK vs non-ARK sites to understand whether the intervention had a greater (or lesser) impact than suggested by the trial-site specific analysis alone. The thesis discusses the analytic challenges of using routine electronic databases to compare antibiotic treatment durations and evaluate antimicrobial stewardship initiatives, and highlights the implications of

specific data quality and accessibility issues for investigators and policy makers. Overall, the thesis demonstrates the need for high-quality health-record-linked e-prescribing systems to support robust assessments of the impact of antibiotic reduction strategies.

0.2 Acknowledgements

I am immensely grateful to my primary supervisor A. Sarah Walker for her encouragement, guidance, and support throughout this project. Always generous with her time – which is in desperately short supply – I owe a significant debt of gratitude to her for believing in me and lending her considerable expertise to help troubleshoot the myriad academic and personal challenges that have cropped up over the past four years. Getting this thesis off the ground and across the finish line would not have been possible without her commitment to students like me, and I consider myself extremely fortunate to have benefitted from her mentorship.

I would also like to thank my co-supervisors, Tim Peto and Martin Llewelyn, whose insights and constructive feedback have helped me through many stages of the DPhil. More broadly, I wish to extend my appreciation to the support and friendship of many colleagues, notably Phuong Quan, Kevin Chau, Jeremy Swann, Chang Ho Yoon, and Tim Davies. Outside my immediate sphere of colleagues are countless friends in Oxford who have enriched my time as a student and weathered the DPhil journey with me, including Matthew Downer, Tim Donnison, and many other members of the Oxford University Ice Hockey Club, including my cousin Chris Byrne.

On the home front, I am eternally indebted to the unwavering support and patience of my wife Ritta Shine and our children Akio and Mai, who have brought joy to my life through the highs and lows of the DPhil. Finally, to my parents Lynn and Peter Budgell, whose sacrifices and continued support have made this moment possible for me.

0.3 Funding

This DPhil was generously supported by Antibiotic Research UK (grant ANTSRG 02/2018). The ARK-Hospital trial was funded by the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research Programme (reference number RP-PG-0514–20015).

0.4 Statement of contribution to this work

I, Eric Peter Budgell, designed and conducted all analyses, wrote all code and text, and produced all tables and figures presented in this thesis. I conceived of the research questions with input from my supervisors, who also advised on the design and methodology of each chapter, and reviewed its contents for scientific merit, with a level of support considered appropriate for a DPhil thesis.

I am grateful to all those who provided support with data acquisition for this thesis, including at NHS Digital (Chapters 2, 5), the UK Health Security Agency (UKHSA) and IQVIA (Chapter 5), in the informatics and pharmacy departments of the University Hospitals Birmingham NHS Foundation Trust (Chapter 3), and at all NHS Trusts participating in the ARK-Hospital trial (Chapters 4, 5).

0.5 Peer reviewed publications arising directly from this work

Budgell EP, Davies TJ, Donker T, Hopkins S, Wyllie DH, Peto TEA, Gill MJ, Llewelyn MJ, Walker AS. Impact of antibiotic use on patient-level risk of death in 36 million hospital admissions in England. *Journal of Infection*. 2022 Mar 1;84(3):311–320.

0.6 Other peer reviewed publications arising in the course of this work

Chau KK, Barker L, **Budgell EP**, Vihta KD, Sims N, Kasprzyk-Hordern B, Harriss E, Crook DW, Read DS, Walker AS, Stoesser N. Systematic review of wastewater surveillance of antimicrobial resistance in human populations. *Environment International*. 2022 Apr;162:107171.

Yokoyama M, Peto L, **Budgell EP**, Jones N, Sheridan E, Liu J, Walker AS, Stoesser N, Gweon HS, Llewelyn MJ. Microbial diversity and antimicrobial resistance in faecal samples from acute medical patients assessed through metagenomic sequencing. *PLoS One*. 2023 Mar 16;18(3):e0282584.

Walker AS, **Budgell EP**, Laskawiec-Szkonter M, Sivyer K, Wordsworth S, Quaddy J, Santillo M, Krusche A, Roope LSJ, Bright N, Mowbray F, Jones N, Hand K, Rahman N, Dobson M, Hedley E, Crook D, Sharland M, Roseveare C, Hobbs FDR, Butler C, Vaughan L, Hopkins S, Yardley L, Peto TEA, Llewelyn MJ, ARK trial team. Antibiotic Review Kit for Hospitals (ARK-Hospital): study protocol for a stepped-wedge cluster-randomised controlled trial. *Trials*. 2019 Jul 11;20(1):421.

Llewelyn M*, **Budgell EP***, Laskawiec-Szkonter M, Cross E, Alexander R, Bond S, Coles P, Conlon-Bingham G, Dymond S, Evans M, Fok R, Frost KJ, Garcia-Arias V, Glass S, Gormley C, Gray K, Hamson C, Harvey D, Hills T, Iyer S, Johnson A, Jones N, Kang P, Kiapi G, Mack D, Makanga C, Mawer D, McCullagh B, Mirfenderesky M, McEwen R, Nag S, Nagar A, Northfield J, O'Driscoll J, Pegden A, Porter R, Powell N, Price D, Sheridan E, Slatter M, Stewart B, Watson C, Weichert I, Sivyer K, Wordsworth S, Quaddy J, Santillo M, Krusche A, Roope LSJ, Mowbray F, Hand KS, Dobson M, Crook DW, Vaughan L, Hopkins S, Yardley L, Peto TEA, Walker AS. Antibiotic review kit for hospitals (ARK-Hospital): a stepped-wedge cluster-randomised controlled trial. *Lancet Infectious Diseases*. 2023 Feb;23(2):207–221.

** Joint first authors*

0.7 Abbreviations

A&E	Accidents and Emergency
AIC	Akaike information criterion
ALT	Alanine transaminase
AMR	Antimicrobial resistance
AMS	Antimicrobial stewardship
ATC	Anatomical Therapeutic Chemical
AWaRe	Access, Watch, and Reserve antibiotics
BIC	Bayesian information criterion
CCS	Clinical Classifications Software
CI	Confidence interval
CPRD	Clinical Practice Research Datalink
CQUIN	Commissioning for Quality and Innovation
CRP	C-reactive protein
CRT	Cluster randomised trial
DARS	Data Access Request Service
DDD	Defined-daily-dose
df	Degrees of freedom
DOT	Days of therapy
EHR	Electronic health record
ESPAUR	English Surveillance Programme for Antimicrobial Utilisation and Resistance
FDP	False discovery proportion
FDR	False discovery rate
FWER	Familywise error rate
GDH	Glutamase dehydrogenate
GP	General practitioner
Hb	Haemoglobin
HES	Hospital Episode Statistics
HES APC	Hospital Episode Statistics Admitted Patient Care data
HESID	Hospital Episode Statistics patient identifier
HR	Hazard ratio
ICD-10	International Classification of Diseases and Related Health Problems - 10th Revision
ICU	Intensive care unit
IMD	Index of Multiple Deprivation
IPCW	Inverse probability of censoring weighting
IPTW	Inverse probability of treatment weighting
IQR	Interquartile range
IRR	Incidence rate ratio
ITT	Intention-to-treat
IGARD	Independent Group Advising on the Release of Data (part of NHS Digital)
LFC	Lactose-fermenting colony

LOT	Length of therapy
MIC	Minimum inhibitory concentration
MRSA	Methicillin-resistant *Staphylococcus aureus*
MSM	Marginal structural model
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
ONS	Office for National Statistics
OR	Odds ratio
PCR	Polymerase chain reaction
PHE	Public Health England
PS	Propensity score
RCT	Randomised controlled trial
SSTF	Start Smart, then Focus
TARGET	Treat Antibiotics Responsibly, Guidance Education Tools toolkit
UHB	University Hospitals Birmingham
UK	United Kingdom
UKHSA	UK Health Security Agency (formerly Public Health England)
WHO	World Health Organization

0.8 List of Figures

0.9 List of Tables

Chapter 1: Introduction

1.1 Antibiotic overuse as a driver of antimicrobial resistance

The selection and transmission of bacterial antimicrobial resistance (AMR) is a leading public health threat(1–5) driven by antimicrobial exposure(6,7), including use in humans(8–15). Although this threat was presaged by Sir Alexander Fleming at the dawn of the antibiotic era(16), his concern was about resistance selection in microbes targeted by antibiotic treatment, motivated by evidence of resistance developing rapidly *in vitro* and accounts of subtherapeutic drug concentrations leading to primary treatment failure and relapse in *Streptococcus pyogenes*(17). This form of resistance selection can arise from systematic underdosing (e.g. due to poor absorption, treatment noncompliance/interruptions) or monotherapy, as illustrated by the failure of early trials into antimicrobial monotherapy for *Mycobacterium tuberculosis*(18,19), prompting follow-up studies that quickly revealed the benefits of more efficacious combination treatment(20). However, whilst target selected resistance may develop in highly clonal species that acquire *de novo* resistance mutations on treatment (e.g. *M tuberculosis*, *Salmonella enterica* serotype *typhi*, and *Neisseria gonorrhoeae*), the majority of AMR-attributable morbidity and mortality today results instead from the selection of resistance in non-target microbial populations in our environment and commensal microbiome(21).

Although normally harmless, these commensal microflora reservoirs harbour a remarkable diversity and density of bacterial species, some of which will have intrinsic insensitivity (e.g. if the microbe naturally lacks a given drug-target pathway) or low-frequency resistant mutations that can be inadvertently selected for during antibiotic treatment of a completely separate infection. Resistance genes may then proliferate vertically through clonal expansion or horizontally through the transfer of mobile genetic elements such as plasmids and transposons to sensitive strains, including pathogenic species(22–25). If the mobile elements contain genes conferring resistance to multiple antibiotic classes, use of one agent may select for the acquisition of multidrug resistance, and

established AMR clones may accumulate additional resistance mechanisms through horizontal gene transfer, particularly in the presence of escalating antibiotic treatment(25). These bacteria may then persist in the human host environment for years, with resistant strains being transmitted between asymptomatic carriers who are missed by traditional syndromic surveillance, acting as sources of opportunistic infection(26,27). Whilst this poses an infection risk to the carriers themselves (e.g. if they experience immunosuppression, injury, surgery, or use indwelling medical devices)(28), it can also lead to wider dissemination, seeding prolonged outbreaks through community(29,30) or nosocomial transmission(31,32), resulting in hard-to-treat surgical site contamination, endocarditis, bacteraemia, and urinary tract infections that drive up costs of care and result in longer hospital admissions and poorer clinical outcomes(5). Of these opportunistic pathogens, the greatest threat is posed by *Escherichia coli* and *Staphylococcus aureus*, which account for half of the fatal AMR burden in high-income countries, followed by *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Enterococcus faecium*, and *Enterobacter spp*, which together accounted for over three-quarters of global AMR-attributable deaths in 2019(2). To detect and limit transmission from asymptomatic carriers, UK hospitals routinely screen for select pathogens in high risk patients, including admissions to intensive care or oncology, and those previously colonised or infected with a target organism; however, this routine screening is only focused on select organisms (e.g. carbapenemase-producing *Enterobacteriaceae*, and methicillin-resistant *S. aureus* (MRSA)(33)) and is therefore inherently limited in its ability to prevent hospital and community transmission of resistant opportunistic pathogens.

In England, surveillance of antibiotic use and AMR trends by the UK Health Security Agency (UKHSA) has shown that more than 80% of resistant bacteraemias are caused by Enterobacterales (predominantly *E. coli*), and rates of resistant bloodstream infections have been on the rise, increasing by 12.4% between 2017-2019, with recent declines (-14.7% in 2019-2021) likely due to fewer community-onset cases during the COVID-19 pandemic(34). Also, in contrast to primary care, hospital inpatient antibiotic prescribing is increasing, with total defined-daily-doses (DDDs) per 1,000

admissions climbing by 3.0% between 2017-2021. Finally, whilst most (80.5%) antibiotic consumption occurred in primary care in 2021, most broad-spectrum use – with the greatest potential to drive non-target ("collateral") resistance selection – occurred in hospitals. Although acquired resistance is not always reversible after the removal of selective pressure(35–37), the extent of resistance selection is strongly correlated with antimicrobial exposure at the community-level(7,10–12) and in individuals(13–15). Reducing antibiotic misuse and overuse is therefore essential, if not sufficient on its own(38), to minimise selective pressure for resistance and curb the transmission of AMR strains between asymptomatic carriers.

1.2 Addressing antimicrobial overuse in hospitals in England

To address the threat posed by AMR, 26/30 countries in Europe and North America have established or planned targets to reduce antimicrobial use in humans(39), including the United Kingdom (UK) which aims to reduce total antimicrobial use by 15% by 2024 (vs 2014) and the number of drug-resistant infections by 10% by 2025 (vs 2018)(40). To achieve this and other quality of care targets, the UK's National Health Service (NHS) established the Commissioning for Quality and Innovation (CQUIN) payments framework(41), which from 2016/17-2018/19 contained financial incentives for acute secondary care organisations (NHS Trusts) to:

- reduce carbapenem, piperacillin-tazobactam, and total antibiotic DDDs/1,000 admissions by <1% per year, with the exact target varying depending on previous progress,
- increase their share of antibiotics used within the "Access" group (i.e. antibiotics deemed to have the lowest toxicity and resistance potential) of the World Health Organization's (WHO) Access, Watch, and Reserve (AWaRe) categories, which were established to help monitor and optimise usage to preserve the effectiveness of "last resort" agents, with local adaptations made by UKHSA (formerly Public Health England (PHE))(42), and
- report the percentage of inpatient antibiotic prescriptions documented and reviewed within 72 hours of treatment initiation

The CQUIN framework made a fraction of each hospital's income conditional on achieving these indicators (together worth 0.25% of acute Trust income), and hospitals were required to share their antibiotic consumption and prescription audit data with PHE (now UKHSA), which was then made publicly available on the "Fingertips" AMR data portal(43). Although these specific AMR-CQUIN indicators were stopped in 2019/20, antibiotic reduction targets are still included in the NHS Standard Contract with Trusts, which is mandated by NHS England and updated yearly. The latest contract requires acute Trusts to cut their per-patient usage of "Watch" and "Reserve" antibiotics by 10% by March 2024 (vs 2017 baseline)(44), equivalent to 1.7% per year, which is slightly greater than the 0.5%-1.3% per year reduction targets for broad-spectrum or total usage in previous contracts(45). Although these reduction targets are all relatively modest, a 2017 survey covering 116/155 (75%) acute Trusts in England found most (81%, 94/116) hospital antimicrobial stewardship leads were concerned about the safety of achieving CQUIN antimicrobial reduction targets(46), likely due to fears of under-treatment leading to patient harm. This is perhaps unsurprising, given that assessments of antimicrobial reduction strategies often suffer from study design weaknesses and fail to report clinical or microbiological outcomes(47).

How Trusts achieve these targets is also not suggested or fixed, however, all Trusts are required by the Health and Social Care Act 2008 to implement antimicrobial stewardship (AMS) programmes to optimise antibiotic use(48), with support from relevant guidelines (e.g. National Institute for Health and Care Excellence (NICE) prescribing guidelines), monitoring and audit tools, and policies such as "Start Smart, then Focus" (SSTF) in secondary care(49). AMS also forms a key component of the UK's five-year plan to tackle AMR(40), with most Trusts coordinating AMS activity through a multidisciplinary team that reports to the executive board, often led by an antimicrobial pharmacist or infection specialist (e.g. clinical microbiologist). Controlling antibiotic overuse in hospitals is challenging though, as clinical urgency and diagnostic uncertainty mean antibiotics often need to be prescribed rapidly and empirically, before diagnostic results are available. To prevent excessively prolonged or broad-spectrum treatment, SSTF guidance recommends that prescriptions should be

reviewed and finalised (stopped or de-escalated if appropriate) within 24-72h of treatment initiation(49). However, opportunities to stop antibiotics at review are often missed due to a combination of organisational and individual factors (e.g. specialty variation in prescribing norms, risk tolerance, and professional hierarchy) that make it hard to change the prescribing decisions made by colleagues(50–53), creating a culture of "non-interference" during prescription reviews(54). As a result, prescription stop rates are relatively low (<10% in 2017), despite high review rates (>90%)(55). Given at least some evidence that higher stop rates (20-30%) can be safely achieved in acute hospital settings(56), low stop rates likely contribute to reports of excessively prolonged treatment(57) and increases in hospital use of Access, Watch, and Reserve DDDs/1,000 admissions, which climbed by 6.1%, 4.7%, 15.8%, respectively, between 2017-2020 in England(34). Although this trend reversed in 2020-2021 (with declines between 9.8-11.3%)(34), this coincided with COVID-related changes in hospital admissions and case-mix, and as more recent national surveillance data is not yet available, it is unclear if these declines have been sustained beyond 2021.

1.3 Measuring the impact of reduced antibiotic use

Antibiotic consumption varies widely among acute hospitals, likely due to variations in hospital-specific characteristics affecting case-mix and prescribing behaviour(58,59). For example, geographic region, hospital size, intensity of antimicrobial stewardship programmes, and hospital type (e.g. private/public or teaching/non-teaching) have all been identified as being associated with hospital antibiotic consumption rates(58). However, comparisons between hospitals are complicated by variation in measures of antibiotic use(59), data collection methods (e.g. whether analyses included paediatric patients or low-consumption departments), and data sources (e.g. dispensing vs usage vs purchasing data), suggesting standardised methods of monitoring consumption are needed to benchmark usage and identify sources of heterogeneity in consumption rates. Where metrics of antibiotic use have been standardised across hospitals, it is unclear whether hospitals that routinely use fewer antibiotics have higher rates of adverse outcomes (after adjusting for case-mix), or

whether more ambitious reduction targets can be safely achieved. The NHS Standard Contract suggests current reduction targets can be safely achieved through relative increases in the use of "Access" antibiotics and shorter antibiotic courses(44), but antimicrobial pharmacists report wide variation in recommended antibiotic prescribing duration for key indications(60) and there remains widespread uncertainty regarding the optimal treatment duration for many bacterial infections(61). Whilst randomised controlled trials (RCTs) could provide unbiased evidence regarding optimal treatment strategies and the impact of shorter versus longer antibiotic prescriptions(62), in practice, the size and cost of trials means very few treatment durations have been compared, and those that have often lack a clear scientific rationale (e.g. comparing 7 versus 14 days)(61). Since the optimal course length may not be identified in a conventional two-arm trial, alternative fixed and adaptive multi-arm designs have been proposed(63,64), but these remain costly, time-consuming, and often have small homogeneous samples, meaning subgroup analyses, detection of rare outcomes, and estimation of complete duration-response curves may be infeasible.

By contrast, observational comparisons of the impact of shorter vs longer antibiotic prescriptions offer better affordability, timeliness, longer follow-up, and more flexible inclusion criteria, but are vulnerable to confounding bias(65). This might occur because confounding factors are measured with error or imperfectly modelled, leading to residual bias, or because relevant confounders are omitted or unavailable, leading to unmeasured confounding. Ordinarily, the effect of this bias is unknown, either masking or exaggerating differences in treatment effects. However, in this context patients receiving prolonged antibiotic courses are likely to be more medically complex and acutely ill and at higher risk of adverse outcomes, resulting in confounding by indication(66). As a result, residual biases are likely to mask any harms from shorter antibiotic therapy and unbiased inferences about the safety of shorter antibiotic course cannot be guaranteed from observational analyses. Nevertheless, observational studies may be of considerable value, even when randomised studies are available, as they can provide evidence of adverse events that cannot be detected with sufficient precision in RCTs(67). They also enable comparison of a wide range of duration-response

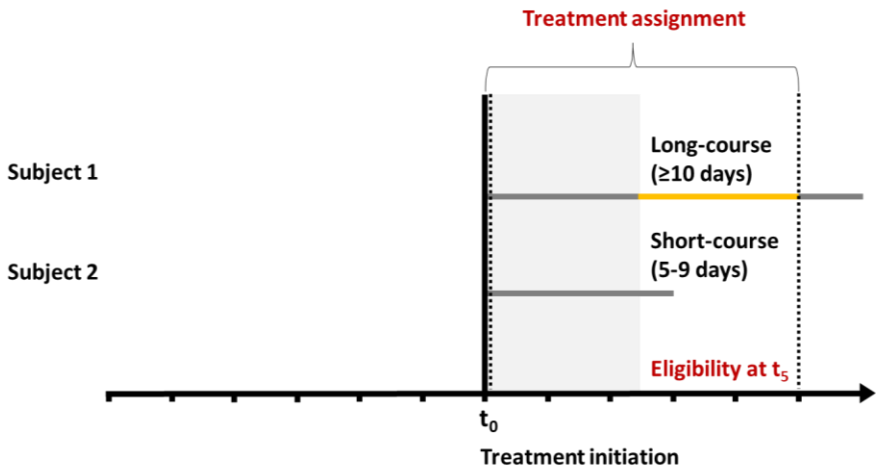
relationships and decision rules that are representative of real-world practice(68), including those that might be unethical to randomly allocate, such as delays in hospital antibiotic treatment(69). Since high-quality observational studies may provide approximately unbiased inferences when a target trial protocol is adequately emulated(70–73), failure to detect harm can help justify randomised experiments and contribute to the wider evidence-base needed to inform prescribing guidelines and reassure practitioners and the public that reducing antibiotic use could be safe.

The growing availability of large routine databases linking electronic health records (EHR) and antibiotic prescribing data presents an opportunity to investigate the possible harms and benefits associated with reductions in antibiotic use, though the use of these data can present challenges for investigators. For example, hospital or ward-level antibiotic use may be measured using DDDs normalised for admissions or bed days, but the numerator (DDDs) and denominator (admissions or bed days) are likely to be estimated from separate routine data sources, and this can result in consumption measures that are highly sensitive to the estimation method and data source(58), particularly if these methods change over time. Where patient-level antibiotic data is available, prescribing registers may only capture treatment start and end dates, rather than individual administered doses, and there may be temporal overlaps between repeat prescriptions, making it difficult to reconstruct treatment episodes or assess compliance with shorter vs longer antibiotic courses.

Other key challenges concern how the observational data are analysed. For example, investigators using patient-level prescribing data to compare shorter vs longer antibiotic courses must be careful to align assessment of eligibility and start of follow-up with exposure assignment(74) to avoid bias resulting from misclassified(75–77) or left-censored follow-up (**Figure 1.1**)(78,79). In RCTs these 3 all occur at the same time, but misalignment remains a common methodological issue in observational comparative effectiveness research(80,81). Longitudinal studies of drug effects are particularly susceptible(82), as time-varying treatments are often modelled as time-fixed variables with exposure

assignment relying on data collected after the start of follow-up, resulting in substantial bias(83). Also, investigators need to account for baseline and time-varying confounding factors, including those affected by previous treatment, because otherwise the comparability (exchangeability) of patients will be unlikely with non-random treatment assignment(84,85). For example, prognostic markers of infection severity such as C-reactive protein (CRP) and procalcitonin(86) may be used to guide antibiotic treatment initiation and duration but are also affected by past treatment(87–90), leading to time-dependent confounding (**Figure 1.2**)(91). Causal statistical methods can address this issue(92–94) but applications remain concentrated in select disciplines (particularly applications in HIV infection) and published reviews of their use (up to 2016) have not identified any studies addressing questions of antibiotic treatment duration(95,96).

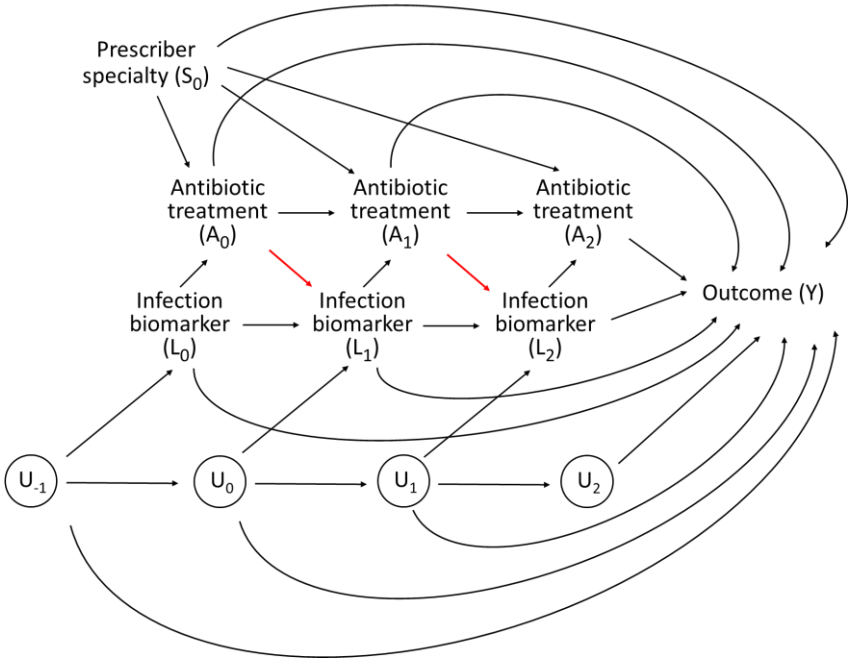
Figure 1.1 Systematic error arising from the failure to align eligibility, follow-up, and treatment assignment in observational comparisons of antibiotic treatment duration



Immortal time bias: If treatment duration is treated as a time-fixed (e.g. "short-course" vs "long-course") and exposure assignment relies on dispensing data collected after treatment initiation (t_0), then follow-up between t_0 and exposure assignment will be misclassified as exposed, rather than unexposed. The result is a span of "immortal" follow-up during which the outcome cannot occur, as represented by the yellow portion of the treatment bar for Subject 1. Since patients must survive long enough to be classified as receiving "long-course" treatment, outcomes on days 5-9 will be credited to Subject 2 by definition and Subject 1 will experience a survival advantage if compared Subject 2. This bias can be substantial and may yield opposite inferences regarding treatment effects(97). **Solutions**: Immortal time can be avoided by modelling antibiotic use as a time-varying exposure and incorporating methods for time-dependent confounding adjustment(92). Alternatively, to contrast alternate treatment strategies subjects can be cloned and assigned to each strategy in an expanded dataset and censored when they deviate from the assigned strategy. Since this censoring is informative (i.e. patients with greater illness/infection severity are likely to receive antibiotic treatment longer or might die early, making them more likely to be censored), inverse probability of censoring weighting (IPCW) can be used to eliminate time-varying selection bias induced by the artificial censoring(76), as illustrated in recent comparisons of short vs long-course antibiotic treatment strategies(98,99).

Selection bias from the use of post-treatment eligibility criteria (immortal time excluded): If eligibility criteria are imposed after the start of treatment, the resulting left-truncation of follow-up (shaded area) will mean outcomes are only ascertained for those who remained alive and event-free at the time of eligibility. The inclusion of such patients (prevalent users) will induce bias if risk varies with time and outcomes early on treatment are not ascertained(74,100). If attrition or treatment adherence before eligibility were differential across exposure groups, this could further enhance bias. **Solution**: Align eligibility with the start of treatment and follow-up (at t_0) to avoid selection bias from the use of post-treatment eligibility criteria(74).

Figure 1.2 A directed acyclic graph demonstrating baseline and time-dependent confounding in the causal relationship between antibiotic treatment and a single outcome



In comparative effectiveness research, investigators are often interested in contrasting clinical outcomes between dynamic treatment strategies that rely on time-varying covariates for sequential decision making. For example, decisions concerning antibiotic treatment often depend on a patient's evolving clinical status, monitored using markers of infection severity ($L_0 \rightarrow L_1 \rightarrow L_2$). When faced with similar patient histories, variation in risk-benefit judgements among prescribers may lead to variation in treatment, including decisions to initiate, modify (e.g. escalate/de-escalate), discontinue, or re-start antibiotics over time ($A_0 \rightarrow A_1 \rightarrow A_2$)(83)(56). This relationship is reflected in the directed acyclic graph above(101), which shows physicians are more likely to prescribe and prolong antibiotics when infection biomarkers such as CRP are elevated (or vice-versa), but those biomarkers are themselves affected by past antibiotic use ($A_{t-1} \rightarrow L_t$)(88).

Standard regression methods can adjust for time-fixed covariates (e.g. prescriber specialty S_0(56)) and time-varying covariates that are not affected by previous treatment, but produce biased estimates in the presence of time-dependent confounding (i.e. where confounder L_t is affected by previous treatment A_{t-1}). This occurs for two reasons. First, since future biomarker measurements (L_t) are intermediates (mediators) that lie on the causal pathway from past antibiotic treatment to the outcome ($A_{t-1} \rightarrow L_t \rightarrow Y$), conditioning on them will block the component of the treatment effect running via L_t, which prevents the total (direct and indirect) causal treatment effect from being consistently estimated—a phenomenon known as over-adjustment(102). Second, in the presence of a time-varying unmeasured common cause of L_t and Y (U_{t-1}), L_t will act as a collider (i.e. share common causes with the outcome: $A_{t-1} \rightarrow L_t \leftarrow U_{t-1}$) and thus conditioning on L_t may open a spurious treatment-outcome path ($A_{t-1} \rightarrow U_{t-1} \rightarrow Y$) known as collider-stratification bias(100,103). Causal methods are required to address these issues(92–94).

Some arrows have been omitted to improve readability (e.g. from $A_{t-2} \rightarrow A_t$, and $L_{t-2} \rightarrow L_t$, and U_{t-2} to U_t), and this example assumes the prescriber S_0 does not change during treatment.

1.4 Thesis outline

To reassure practitioners and the public that substantial reductions in antibiotic use could be safe, Chapter 2 explores associations between variation in hospital-level antibiotic prescribing and patient-level clinical outcomes in 36 million general medicine admissions in England. I use this analysis to estimate what system-wide reductions might, at least in theory, be safely achieved if higher-using hospitals could replicate the prescribing practices of low-using hospitals. Chapter 3 builds on this work by using patient-level antibiotic and outcome data to investigate the causal effect of reducing antibiotic use by treating time-varying antibiotic use as sequence of non-randomised nested "trials", using an approach proposed by Miguel Hernán *et al*(104). By design, this analytic strategy avoids the bias that can arise from misclassified or left-censored follow-up (**Figure 1.1**), and also avoids time-dependent confounding from infection biomarkers used to guide inpatient antibiotic treatment (**Figure 1.2**).

To inform its adoption in other acute care settings, Chapters 4-5 then address two key questions emerging from the recently completed ARK-Hospital cluster-randomised trial, which tested a multifaceted behaviour change intervention to reduce antibiotic overuse in general medicine(105). The underlying rationale, design, and findings of this trial are summarised in the introduction of Chapter 4. Since the effectiveness of the ARK intervention varied widely among the trial sites, Chapter 4 examines the role of 26 contextual factors as potential mediators of impact, to better understand sources of this between-site variation. Lastly, in Chapter 5 I used ward-level prescribing data from all acute Trusts in England, spanning a longer time period than was available in the trial, to compare antibiotic use trends in ARK vs non-ARK sites, with the latter sites serving as a counterfactual to estimate what may have occurred in ARK sites had they not implemented. This analysis was conducted to shed light on the longer-term impact of ARK intervention and to understand whether the ARK intervention had a greater (or lesser) impact than suggested by the trial-site specific analysis alone, given national background trends in the acute hospital setting.

I conclude by discussing the analytic challenges of estimating prescribing trends and the impact of reduced antibiotic use with routinely collected data. I also discuss the implications of this work for antimicrobial stewardship activity aiming to reduce antibiotic overuse.

Chapter 2: Impact of antibiotic use on patient-level risk of death in 36 million hospital admissions in England

Note: This chapter was published in December 2021 and I was the first and corresponding author(106). I authored the text and figures below, which are reproduced here largely unaltered from the published work, with minor amendments. Input from supervisors was commensurate with the amount of advice that would be considered appropriate for a DPhil thesis. All authors have agreed to the inclusion of this published work in my thesis.

2.1 Introduction

Antibiotic overuse puts individual patients(13) and whole populations(11,12) at risk of antimicrobial resistance (AMR). AMR infections cause higher mortality, longer hospital admissions, and increased costs of care(5). Reducing unnecessary antibiotic use is therefore essential to reduce the selective pressure for resistance(8,9) and avoid other harms including adverse drug events(107,108), toxicity(109–111), changes to the gut microbiome(24–27), and increased susceptibility to infections such as *Clostridioides difficile*(112–115).

In England's NHS, primary care antibiotic stewardship initiatives focusing on decisions to start antibiotics have been successful in reducing antibiotic overuse(34), but in hospital practice, prescribers must balance the risks of harm posed by under-treating an infection and those of prolonged or excessively-broad spectrum antibiotic use(116). For patients with life-threatening infections, even modest treatment delays can increase mortality risk(117), and diagnostic uncertainty means it is often necessary to administer broad-spectrum antibiotics empirically while waiting for additional clinical, microbiological, or radiographic information.

For these reasons, controlling antibiotic overuse in hospitals depends on early antibiotic initiation followed by the review of prescriptions after 24-72 hours to re-evaluate whether continued therapy is appropriate(118). Different healthcare systems operationalise this approach in different ways,

including through “antibiotic timeouts” in the United States(119) and “Start Smart then Focus” in England(118). Although 20-30% of prescriptions may be safely stopped at review(56), stop and review dates are poorly documented(120) and opportunities to stop early or “de-escalate” – i.e. switch from parenteral to oral antibiotics, or to agents with a narrower spectrum of activity – are often missed(46,121–123). In NHS hospitals inpatient antibiotic consumption has continued to increase year-on-year(34) despite the introduction of financial incentives(41) and targets to reduce overuse(45).

The issues of clinical urgency and diagnostic uncertainty which make limiting antibiotic overuse in hospitals challenging for clinicians also make attempts to define inappropriate use inherently subjective(124). Another way to approach this problem is through “benchmarking”, whereby low-prescribing organisations are used to drive improvements(125–127), on the assumption that high-prescribing organisations could reduce use to these levels without harm. Previous studies have reported wide variation in both recommended antibiotic prescribing duration(60) and total antibiotic consumption among acute hospitals(58,59). However, simple comparisons of hospital-level consumption data largely fail to account for case-mix, and in hospitals with more acute patients, systematic under-treatment might be expected to harm patients. Antibiotic stewardship leads at acute hospitals in England have also expressed concern about the safety of a target-driven antibiotic reduction strategy(46). As a result, many hospitals are benchmarked against their own historical performance rather than externally, and only relatively small reduction targets are sought, such as the 1% year-on-year reduction target used in the 2021/22 NHS Standard Contract with hospital Trusts(128). Examination of the possibility that substantially driving down antibiotic use could compromise clinical outcomes is needed to reassure practitioners and the public that substantially reducing antibiotic use could be safe.

This chapter therefore aimed to determine the extent to which variation in antibiotic prescribing was associated with risk of death and hospital readmission in general medicine inpatients in England.

Whilst observational analysis can never prove safety, were reduced antibiotic treatment associated with harm, this should be identifiable from case-mix adjusted observational data.

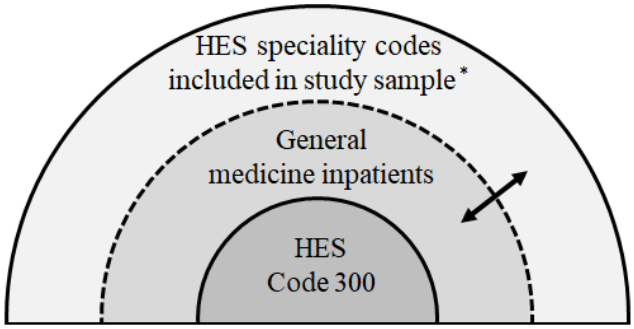
2.2 Methods

2.2.1 Data sources

Hospital Episode Statistics (HES) Admitted Patient Care (APC) data was obtained through NHS Digital's Data Access Request Service (DARS). The HES APC data captures all emergency, planned, and day-case admissions requiring an NHS hospital bed in England, but excludes outpatient visits and Accidents and Emergency (A&E) attendances, which are stored in separate HES databases not accessed in this analysis(129). The configuration of acute medical services varies widely across hospitals and general medical patients may be treated under several different adult specialties, with coding practices varying by hospital(130). The study population was therefore defined using multiple consultant specialty codes, chosen to reflect the specialties most commonly used to admit adult general medicine inpatients (**Figure 2.1**). A binary measure of death within 14 and 30 days of admission (all-cause, in or out of hospital) was obtained through linkage with data from the Office for National Statistics (ONS), performed in advance by NHS Digital. A list of collected HES data fields is provided in the appendix.

To calculate hospital exposure, all inpatient spells and episodes were requested for patients with any eligible admission. A provider "spell" reflects a period of care in a single hospital, starting at the time of admission and ending at the time of discharge, death, or transfer out. Within a spell an "episode" reflects a defined period of care under a consultant(129). Thus, a spell comprises one or more episodes, and episodes which share a patient identification number (HESID), admission date, provider code, and hospital provider spell number together compose a spell(131).

Figure 2.1 Study sample



*Spells were included if the specialty under which attending consultants were contracted (HES field: mainspef) or worked (HES field: tretspef) included any of the following codes in the first or second episode: general medicine (300), gastroenterology (301), endocrinology (302), haematology (303), diabetic medicine (307), cardiology (320), acute internal medicine (326), respiratory/thoracic medicine (340), infectious diseases (350), neurology (400), rheumatology (410), and geriatric medicine (430). Given the variation in coding practices among hospitals, admission codes in general medicine (300) are likely to be only a subset of general medicine inpatients (i.e. perfect specificity, but less than perfect sensitivity). By contrast, the wider set of admission codes used to define this study sample is likely to have very high sensitivity but at the cost of reduced specificity. The extent of the reduction in specificity will vary by hospital depending on local coding practices.

Information on NHS hospital-level antibiotic consumption was obtained on 05/January/2020 from pharmacy dispensing records provided by IQVIA (formerly Quintiles/IMS Health)(132) through UKHSA (formerly PHE). Antibiotic use was measured in DDDs, which enable standardised comparisons of antibiotic use and are defined by WHO as the average maintenance dose per day for a drug used for its main indication in adults(133). DDDs reflected the 2020 WHO Anatomical Therapeutic Chemical (ATC)/DDD index and were disaggregated by drug, route of administration, quarter (April/2014-March/2017), and inpatient and outpatient prescribing at the hospital-level (not by specialty). For each acute hospital Trust, the annual number of occupied overnight bed-days was collected from UKHSA(134) (which compiled the data based on NHS England's KH03 reports(135)), and yearly admissions were obtained from NHS Digital(136), spanning April/2014-March/2017.

2.2.2 HES data cleaning

The raw HES data extract included 123,079,071 consultant episodes. Episodes were dropped if they were:

- Identified as perfect duplicates (n=14,672,193) or duplicates across all fields except episode identifier number (epikey) (n=22,196), or
- Admitted before 01/April/2009 (n=139,179), or
- Missing an admission date (n=5,804), or
- Missing a HESID (n=70,619), or
- Missing a hospital provider spell number (n=340)

Among the remaining 108,168,740 episodes, there were 88,718,419 spells from 15,708,476 patients admitted between 01/April/2009-31/March/2017. The remaining cleaning steps are outlined in **Figure 2.2**.

A "Trust" is an organisational unit within the NHS that may provide hospital services, community services, or act as a commissioner when sub-contracting care to other health services providers. Occasionally, Trusts may merge or split over time and this was accounted for by updating provider codes so they were current at March/2017 (appendix **Table 2.2**)(131).

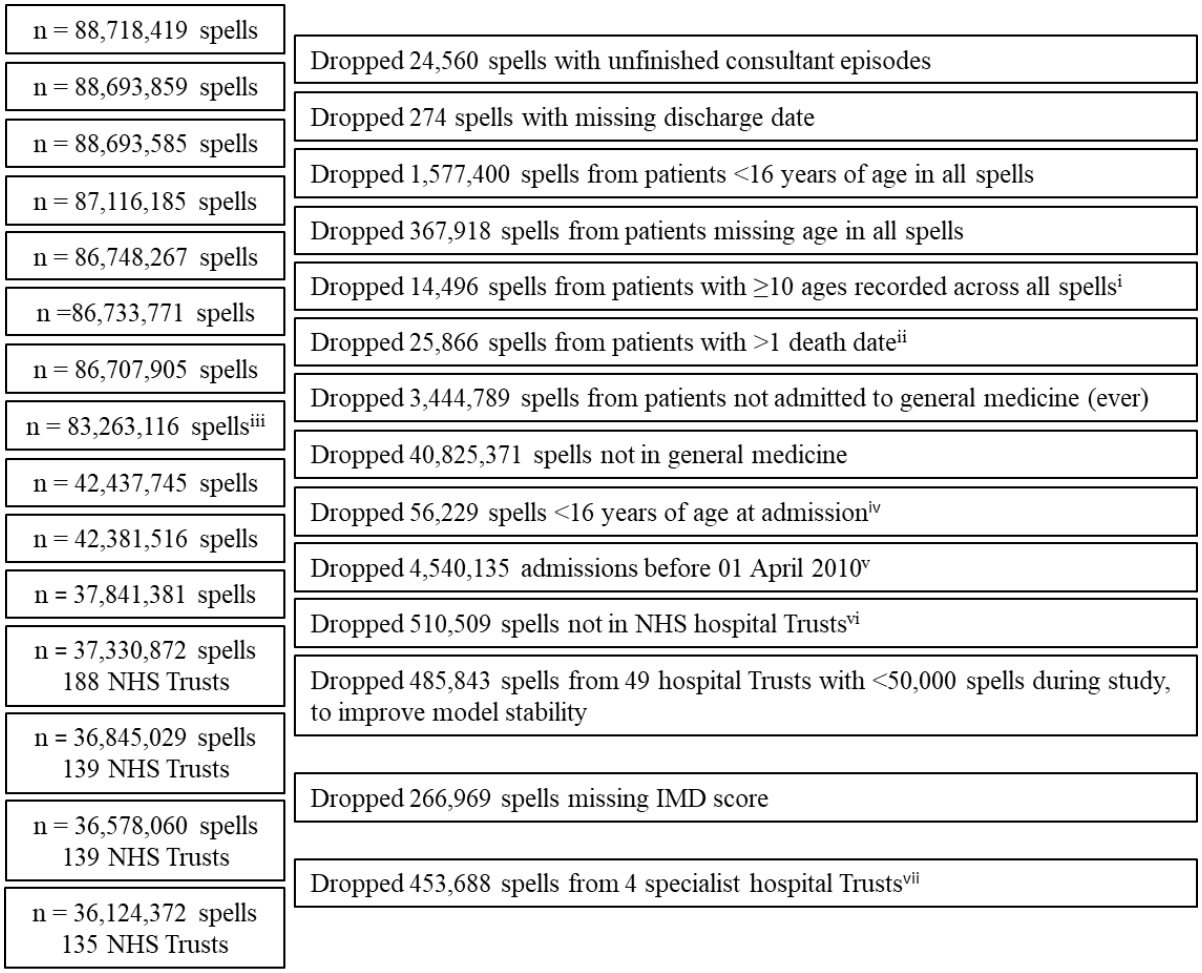
All data processing was carried out in Stata/MP 16 with data held on NHS servers located at the Oxford University Hospitals NHS Foundation Trust. I followed the RECORD statement(137) for routinely collected observational health data (the location of each checklist item is presented in appendix Table S2 in(106)).

2.2.3 Antibiotic data cleaning

I restricted the antibiotics to systemic agents; thus eye drops and other topical agents (gels, creams) were excluded, as were pessaries, beads, bone cement, and non-absorbed antibiotics such as neomycin. A small number of other agents were excluded for miscellaneous reasons, including methenamine (not an antibiotic, but is metabolised to formalin to suppress urinary tract infections), demeclocycline (used primarily to treat hyponatremia), and colistin (ATC code: A07AA10) which was

rarely used. Rifabutin was excluded due to its probable use in tuberculosis treatment; other antibiotics used primarily in the treatment of tuberculosis were excluded in advance by UKHSA.

Figure 2.2 Data cleaning steps followed to derive analytic cohort



[i] Patients admitted annually between 2009-2017 should not have ≥10 different ages recorded and were therefore dropped, as age was judged to be unreliable.

[ii] If a patient was recorded as having died in >1 spell, the earliest date of death was retained and subsequent spells were dropped.

[iii] At this point, previous hospital exposure and emergency re-admission after discharge were estimated, as all patients had at least 1 spell meeting the inclusion criteria in **Figure 2.1**.

[iv] These spells belonged to patients with a subsequent admission in which they were aged ≥16 years.

[v] Admissions between 01/April/2009-31/March/2010 were used only to estimate previous hospital exposure for spells beginning from 01/April/2010.

[vi] By restricting the dataset to spells with a provider code of treatment beginning with "R", this step excluded primary care Trusts, independent providers, and NHS treatment centres(138).

[viii] Includes one hospital Trust where most (84.9%) admissions were to clinical haematology (The Royal Marsden NHS Foundation Trust) and three hospital Trusts where all admissions were to either cardiology or respiratory medicine (Royal Brompton and Harefield NHS Foundation Trust, Papworth Hospital NHS Foundation Trust, and Liverpool Heart and Chest NHS Foundation Trust).

2.2.4 Primary analysis

The primary ecological analysis used meta-regression(139) on hospital-level summary data to compare outcomes in acute/general medicine inpatients with hospital-level antibiotic use.

As a first step, I derived confounding-adjusted relative risks of death (all-cause, in or out of hospital) within 30 days of admission to acute/general medicine (as defined above) using Poisson regression. To account for patients with repeated hospital admissions (observations which are not independent), a clustered version of the sandwich estimator was used, which affects the variance–covariance matrix of the estimators (including standard errors) but has no impact on estimated coefficients, thus yielding wider confidence bounds around point estimates than would be observed if all observations were assumed to be independent(140). Models included admissions between 01/April/2010-31/March/2017, with prior data used only to calculate previous hospital exposure. A separate model was fit to each hospital (135 models) to allow each potential confounder to have a different impact on the outcome in each hospital, including the same factors in each model regardless of statistical significance (i.e. this was done to ensure the best possible adjustment for case-mix, rather than to assess the relative effect of factors within models). Multivariate models included an *a priori* list of 16 admission factors used to adjust for case-mix in previous mortality analyses (**Table 2.1**), including nine interaction terms that improved model fit by lowering the Bayesian information criterion (BIC) of a single multivariate model applied to all hospitals (with hospital as a main effect)(86,141). Continuous factors were truncated at the 2.5th, 95th, or 99th percentiles to reduce the influence of outliers, and natural cubic splines were used (Stata command: mkspline, cubic) to account for non-linear effects if they lowered the BIC of a model containing hospital as a factor; the number of knots was chosen based on BIC (appendix **Figure 2.7**). Further detail regarding model selection is provided in the appendix.

Table 2.1 Admission characteristics of 36,124,372 acute or general medical admissions in 135 acute NHS hospital Trusts, England, April 2010 to March 2017

Characteristic		Admissions (N=36,124,372)
Age (years)[i]	Median (IQR)	66 (51-78)
Charlson Comorbidity Index[i]	Median (IQR)	0 (0-7)
Index of multiple deprivation (IMD)[ii]	Median (IQR)	18.5 (10.5-32.2)
Overnight admissions in past year[i]	Median (IQR)	0 (0-1)
	0	20,684,824 (57.3%)
	1-3	12,626,901 (35.0%)
	4-6	2,032,847 (5.6%)
	7+	779,800 (2.2%)
Any complex overnight admission in past year (>1 consultant episode, excluding episodes in A&E or rehabilitation)	No[iii]	27,299,336 (75.6%)
	Yes	8,825,036 (24.4%)
Sex	Female[iii]	18,197,325 (50.4%)
	Male	17,927,047 (49.6%)
Ethnic category	White[iii]	30,270,021 (83.8%)
	Asian	1,728,243 (4.8%)
	Black	874,178 (2.4%)
	Mixed/other	640,043 (1.8%)
	Unknown	2,611,887 (7.2%)
Patient classification	Ordinary admission[iii]	20,663,007 (57.2%)
	Day-case admission	13,559,311 (37.5%)
	Regular day attender	1,902,054 (5.3%)
Admission source	Usual/other residence[iii]	34,680,673 (96.0%)
	NHS general ward / other care provider	1,443,699 (4.0%)
Admission method	Accident and emergency (A&E) [iii]	14,477,420 (40.1%)
	Elective / non-emergency	17,218,343 (47.7%)
	Emergency via GP or other	4,428,609 (12.3%)
Immunosuppression	No[iii]	35,119,372 (97.2%)
	Yes	1,005,000 (2.8%)
Admission time of week	Weekday[iii]	30,719,593 (85.0%)
	Weekend	5,404,779 (15.0%)
Admission specialty	General medicine[iii]	19,144,018 (53.0%)
	Gastroenterology	4,991,951 (13.8%)
	Clinical Haematology	4,441,171 (12.3%)
	Cardiology	2,456,490 (6.8%)
	Geriatric Medicine	1,878,069 (5.2%)
	Other	3,212,673 (8.9%)

Characteristic		Admissions (N=36,124,372)
Admission financial year (April-March)	2010	4,532,650 (12.5%)
	2011	4,712,543 (13.0%)
	2012	4,939,621 (13.7%)
	2013	5,124,951 (14.2%)
	2014	5,383,865 (14.9%)
	2015	5,634,144 (15.6%)
	2016[iii]	5,796,598 (16.0%)
Admission month	January[iii]	3,137,513 (8.7%)
	February	2,922,611 (8.1%)
	March	3,170,887 (8.8%)
	April	2,890,364 (8.0%)
	May	2,975,721 (8.2%)
	June	2,951,925 (8.2%)
	July	3,043,793 (8.4%)
	August	2,937,379 (8.1%)
	September	2,962,424 (8.2%)
	October	3,088,645 (8.6%)
	November	3,040,081 (8.4%)
	December	3,003,029 (8.3%)
5 most prevalent Clinical Classifications Software (CCS) diagnosis groups (abbreviated, out of 29)	Nonspecific chest pain[iii]	4,649,241 (12.9%)
	Cancer related	3,848,695 (10.7%)
	Headache / other nervous system disorders	2,520,624 (7.0%)
	Regional enteritis and ulcerative colitis	2,288,317 (6.3%)
	Other gastrointestinal disorders	2,062,927 (5.7%)
Five largest hospital Trusts	Barts Health NHS Trust	655,728 (1.8%)
	Heart of England NHS Trust	600,781 (1.7%)
	University Hospitals of Leicester	592,378 (1.6%)
	King's College Hospital NHS Foundation Trust	577,624 (1.6%)
	Pennine Acute Hospitals NHS Trust	563,816 (1.6%)
Five smallest hospital Trusts	Yeovil District Hospital NHS Foundation Trust	103,543 (0.3%)
	George Eliot Hospital NHS Trust	94,417 (0.3%)
	East Cheshire NHS Trust	93,900 (0.3%)
	Weston Area Health NHS Trust	89,741 (0.2%)
	Wye Valley NHS Trust	86,567 (0.2%)

[i] Truncated at the 99th percentile to improve model stability. Median age was used as reference in regression models. Overnight admissions in past year was included in adjusted models as a continuous covariate.

[ii] Higher IMD scores indicate greater deprivation (range: 0.53-87.8), and median IMD score was used as the reference in regression models.

[iii] Reference category in regression models.

Second, I derived marginal effects to understand the absolute probability of the outcome (30-day mortality) at the reference level of all model covariates(142), estimated as $p(x) = e(\beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + \beta_n x_n)$, where β_0 (intercept) reflects the log-rate given the parameter estimates ($\beta_1 x_1 \ldots \beta_n x_n$). Marginal effects were derived in Stata using "nlcom" for nonlinear combination of estimators, and results were compared with output from Stata's "margins" command, which was computationally slow but provided a reference check on all estimates and their standard errors. Time-to-event models could not be used as they required knowledge of date of death, which is available from ONS but considered identifiable by NHS Digital; only a binary indicator of death pre-calculated by NHS Digital was therefore available for analysis. For the primary mortality analysis, marginal effects were also derived using logistic regression(142,143), with the probability of the outcome estimated as: $p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2 \ldots+\beta_n x_n)}}$. However, these were very highly correlated (Spearman's ρ: 0.9992, p<0.0001) and similar in magnitude to estimates derived from Poisson models, so were not considered further.

Finally, random-effects meta-regression was used (Stata command: metareg(144)) to investigate whether heterogeneity in the adjusted probability of death (one estimate per hospital) was associated with hospital-level antibiotic use, measured in DDDs/1,000 bed-days. This method is an extension of standard random-effects meta-analysis(139,145–147) and is often performed on study-level summary data, allowing the underlying sources of statistical heterogeneity in marginal effects across hospitals to be explored. The use of random-effects models accounts for both the variance of hospital-level effects and residual (unexplained) sources of between-hospital heterogeneity, assuming hospital-level effects follow a normal distribution around the linear predictor. I carefully reviewed spaghetti plots of quarterly changes in antibiotic use for all 135 Trusts and found antibiotic use was largely stable over time. DDD estimates were therefore calculated as the annual mean from April/2014-March/2017. Multiple other metrics of antibiotic use were considered, including inpatient and outpatient DDDs, narrow-spectrum and broad-spectrum DDDs, parenteral and oral

DDDs, DDDs for piperacillin/tazobactam and meropenem, and DDDs for UKHSA-interpretations of WHO AWaRe antibiotics(42) (appendix **Table 2.3**). The minor UKHSA amendments to AWaRe classifications reflected which antibiotics NHS hospitals should be prioritising for human use. Meta-regression models adjusted for hospital size measured in terciles of either bed-days(134) or admissions(136), depending on whether DDDs was measured per bed-days or per admissions. An interaction between DDDs and hospital size was assessed and retained where the heterogeneity $p<0.01$.

2.2.5 Sensitivity analyses

In sensitivity analyses, a more narrow definition of general medicine was used, where all spells had a HES main specialty code or treatment specialty code of 300 in the first or second consultant episode. As outlined in **Figure 2.1**, this definition is likely to have very high specificity but varying sensitivity among hospitals due to local variations in coding practices. Since antibiotic data was only available from April/2014, a separate sensitivity test restricted the study sample to admissions between 01/April/2014-31/March/2017.

2.2.6 Secondary analyses

Secondary outcomes included death within 14 days of admission (all-cause, in or out of hospital) and non-elective re-admission to hospital within 30 days of discharge (regardless of re-admission speciality). Analyses of re-admission were restricted to patients discharged alive more than 30 days before 31/March/2017. Without death date, deaths outside hospital could not be treated as competing events for re-admission, and therefore were effectively ignored. Poisson models for each binary outcome adjusted for the same admission characteristics and interaction terms and this was followed by random-effects meta-regression as previously described.

2.2.7 Scope for achieving antibiotic prescribing reductions

I considered the impact of reducing hospital antibiotic use to the 25th, 10th and 5th percentiles among hospitals, by examining the proportion of DDDs that could be safely avoided if high-using hospitals were able to replicate the prescribing practices of lower using hospitals.

2.2.8 Ethics

Patient consent was waived as the study protocol was approved by the Health and Social Care Information Centre (now NHS Digital), whose guidance at the time was that the use of non-identifiable data did not require separate Research Ethics Committee review.

2.3 Results

2.3.1 Primary analysis

The final analytic cohort contained 36,124,372 acute/general medicine admissions from 12,320,069 patients between April/2010 and March/2017 inclusive, with a median of two admissions (IQR: 1-3) per patient. Admissions increased year-on-year (**Table 2.1**). The largest and smallest hospitals received 655,728 admissions and 86,567 admissions respectively (appendix **Table 2.4**). Median age was 66 years (IQR: 51-78), the most prevalent admission characteristics were female sex (50.4%), ethnically white (83.8%), admitted on a weekday (85.0%), a low median Charlson Comorbidity Index (0, IQR: 0-7), low median IMD score (18.5, IQR: 10.5-32.2), and a Clinical Classifications Software (CCS) diagnosis group indicating non-specific chest pain (12.9%) (appendix **Table 2.5**).

In primary analyses, which adjusted for all the factors in **Table 2.1**, the adjusted probability of 30-day mortality varied three-fold by hospital (0.58-1.75%; median: 1.12%, IQR: 0.96-1.28%). Antibiotic consumption measured as mean total DDDs/1,000 bed-days (2014-2016) varied 15-fold across hospitals (median: 1,814; IQR: 1,624-2,080; range: 266-4,006), with a 13-fold difference when expressed as DDDs/1,000 admissions (median: 4,132; IQR: 3,604-4,766; range: 584-7,494). Wide

variation was also observed for other antibiotic metrics (**Figure 2.3**). Ten antibiotics accounted for the majority prescribed (63.36-85.10%) at every acute hospital, but there was also considerable variation in the use of individual agents (appendix **Figure 2.8**).

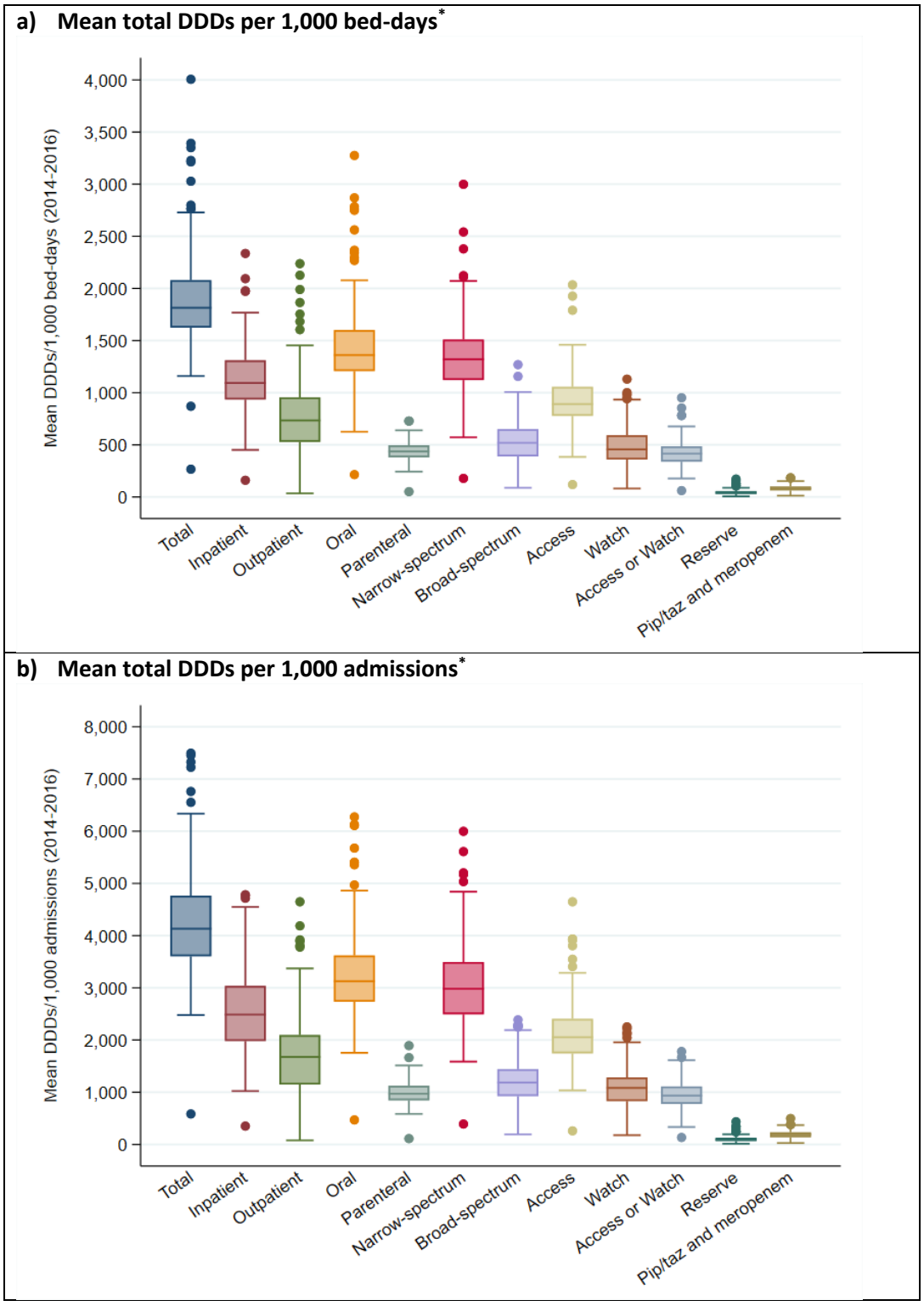
Meta-regression estimates are displayed in **Figure 2.4** and appendix **Table 2.6**, with associations between selected antibiotics and 30-day mortality in **Figure 2.5** and appendix **Figure 2.9**. In 22/24 meta-regression models I found evidence of no association between the adjusted probability of death and hospital-level antibiotic use; the two models with some evidence of association identified effects in opposite directions, with slightly lower mortality (-0.028%, 95% CI: -0.053, -0.003; p=0.028) for each 500 oral DDDs/1,000 admissions higher, and higher mortality (+0.284, 95% CI: +0.031, +0.538; p=0.028) for each 500 parenteral DDDs/1,000 bed-days higher.

2.3.2 Sensitivity analyses

Using a narrower definition of general medicine resulted in an adjusted probability of 30-day mortality that varied between 0.59-2.10% across hospitals (median: 1.11%, IQR: 0.96-1.33%). In 23/24 meta-regression models there was no evidence of association between the adjusted probability of death and hospital-level antibiotic use (**Figure 2.4** and appendix **Table 2.7**). In one model there was a slightly higher adjusted probability of death of +0.373% (95% CI: +0.082,+0.663; p=0.012) for 500 parenteral DDDs/1,000 bed-days higher.

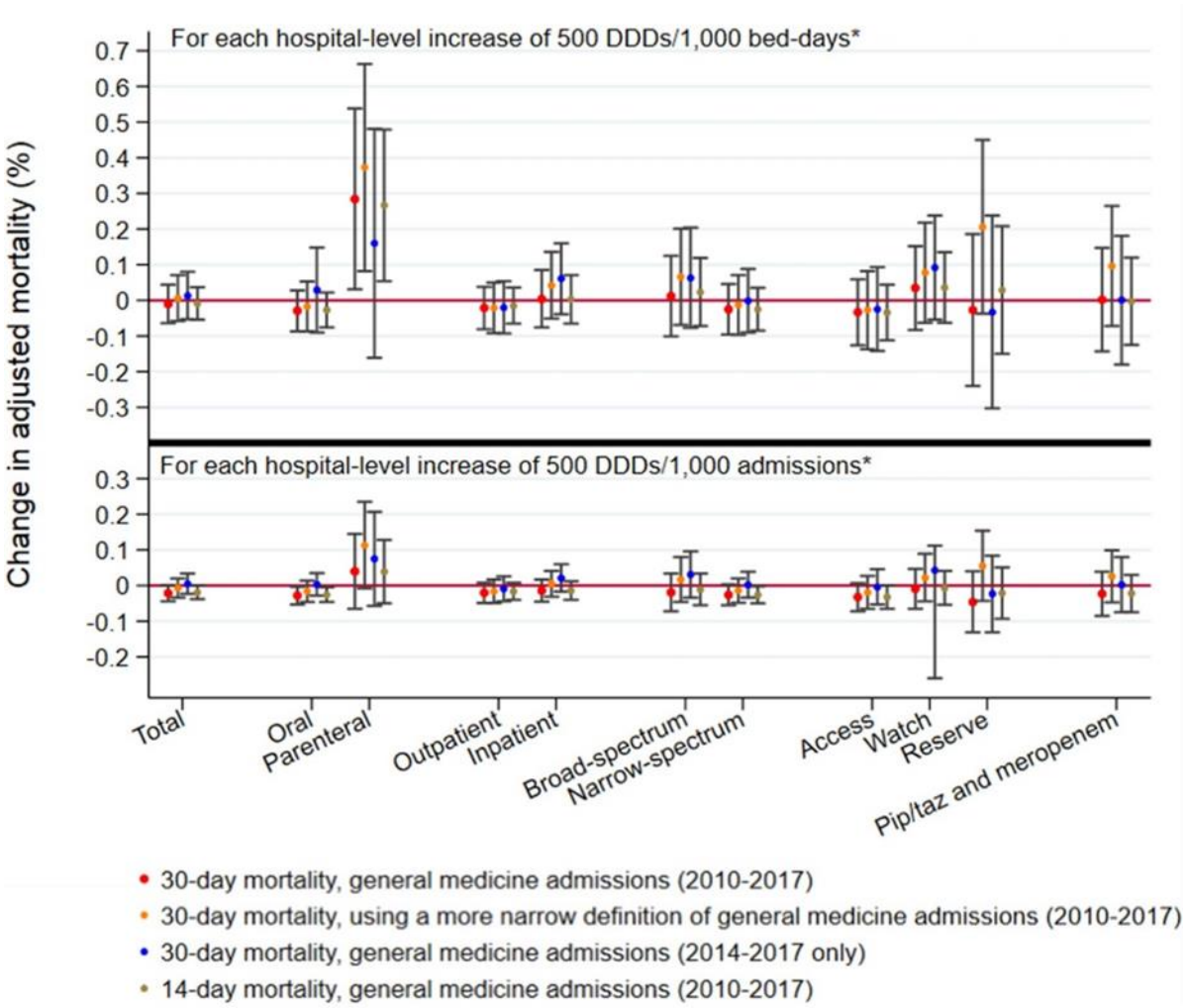
Restricting the study sample to admissions between April/2014-March/2017 inclusive (i.e. the period with overlapping antibiotic data) yielded adjusted probability of death estimates that were highly correlated (Spearman's ρ: 0.851, p<0.0001) and similar in magnitude to those in the primary analysis and showed no evidence of association with hospital-level antibiotic use, regardless of how antibiotic use was measured (**Figure 2.4** and appendix **Table 2.8**).

Figure 2.3 Mean antibiotic defined-daily-doses consumed among 135 NHS acute hospital Trusts (April 2014 – March 2017)



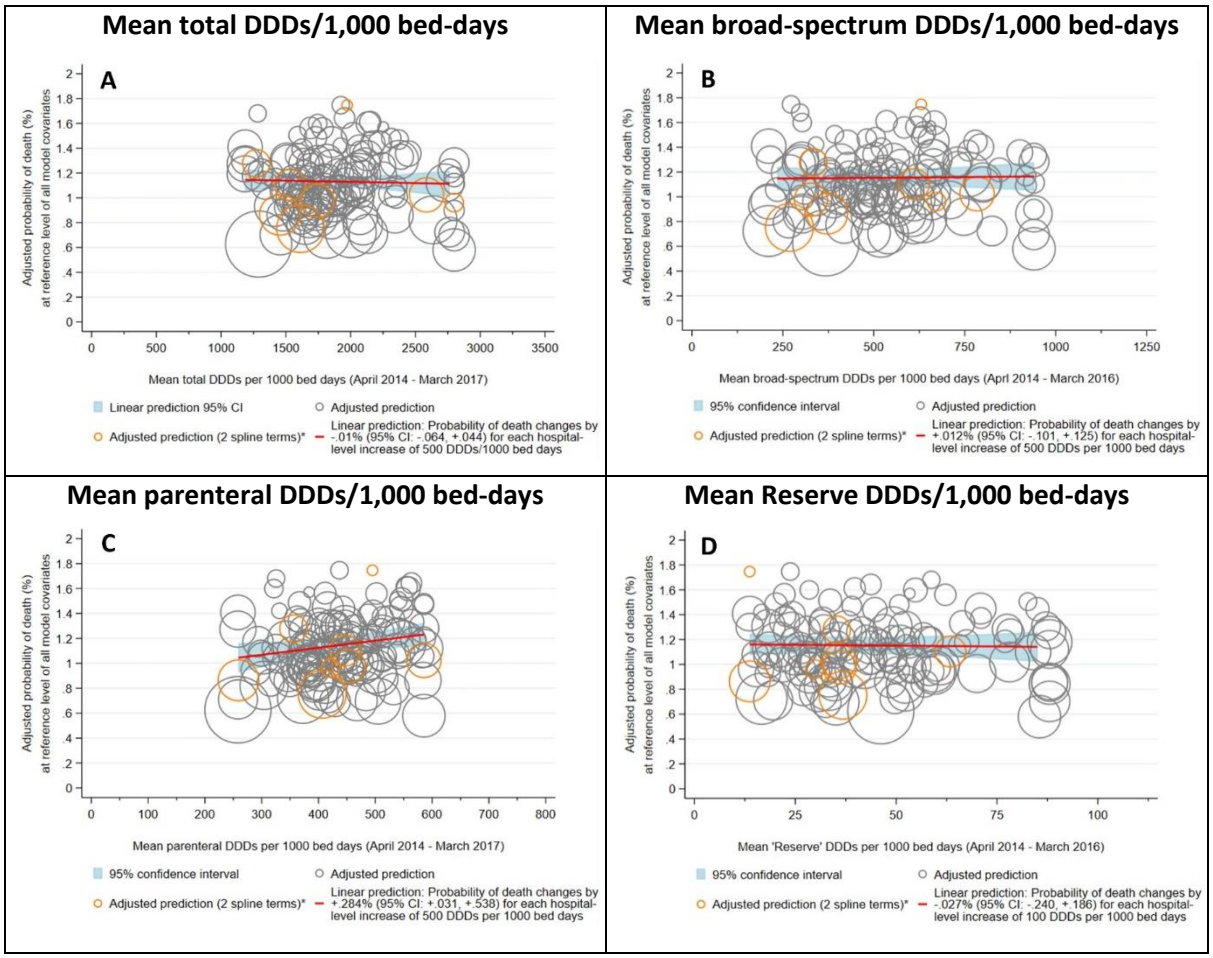
* Each box represents the interquartile range of the distribution and is subdivided by a horizontal line representing the median. The ends of the whiskers display the most extreme DDD estimate within 1.5 IQR of the nearest quartile, while even more extreme outliers are displayed as isolated points. "Pip/taz" refers to piperacillin/tazobactam. Some antibiotics may be considered Access or Watch depending on indication, so they have been included as their own separate category.

Figure 2.4 Random-effects meta-regression of the association between the adjusted probability of death (in or out of Hospital) and different metrics of hospital-level antibiotic use



* Point estimates above the null (red line) suggest higher hospital-level antibiotic use is associated with harm (higher mortality risk), or conversely that lower hospital-level antibiotic use is associated with clinical benefit (lower mortality risk). Estimates below the null suggest higher antibiotic use is associated with clinical benefit, or conversely that lower antibiotic use is associated with clinical harm. The associations displayed for Reserve antibiotics and piperacillin/tazobactam ("pip/taz") and meropenem are for a unit of 100 DDDs higher rather than 500 DDDs. Some antibiotics (not shown) may be considered either Access or Watch (depending on indication) and have been analysed as their own separate category in appendix **Table 2.6–Table 2.10** alongside the categories shown here.

Figure 2.5 Random-effects meta-regression of the association between the adjusted probability of death within 30 days of admission (in or out of hospital) and hospital-level antibiotic use



[1] Marginal effects were derived from 135 separate (hospital-specific) models, represented here as circles sized according to the precision of the estimate (inverse of the within-hospital variance). Most (127/135) hospital-specific multivariate models included four spline terms (i.e. for age, Charlson Comorbidity Index, IMD score, and overnight admissions in the past year), however models for eight hospitals would only converge with two spline terms (age and overnight admissions in the past year). Antibiotic use was truncated below the 2.5th percentile and above the 95th percentile. Probability estimates were truncated above the 99th percentile. Associations displayed here can be found in appendix **Table 2.6**.

2.3.3 Secondary analyses

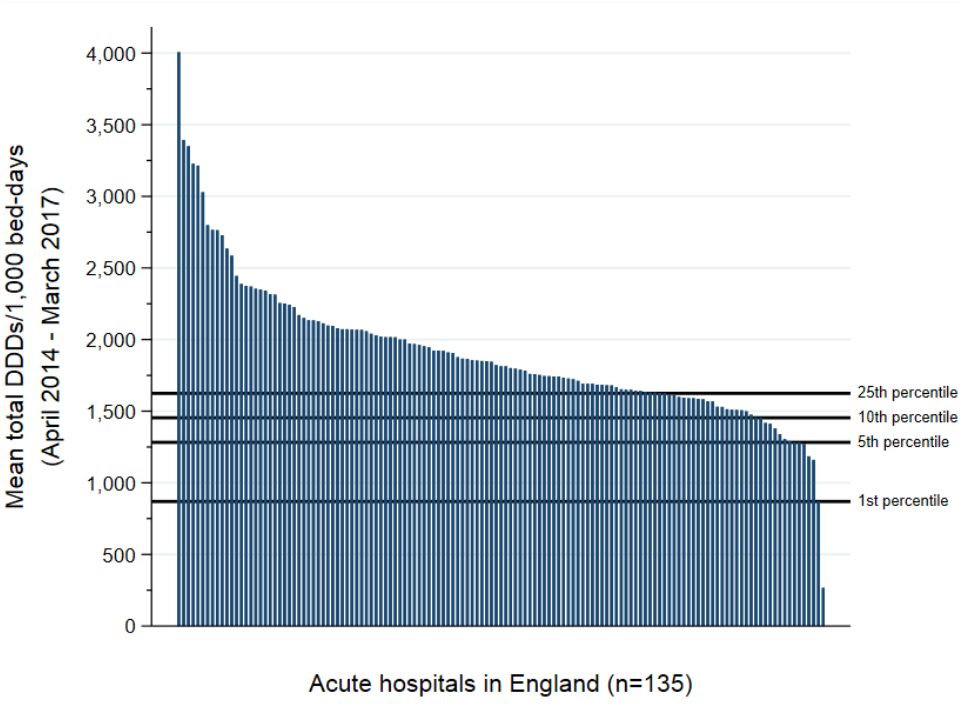
The adjusted probability of 14-day mortality (in/out of hospital) varied from 0.38-1.40% across hospitals (median: 0.79%, IQR: 0.65-0.93%) at the reference level of all model factors. Most (20/24) metrics of antibiotic use showed no evidence of association with 14-day mortality, while the remaining estimates showed both positive and negative associations (**Figure 2.4** and appendix **Table 2.9**). The adjusted probability of non-elective re-admission to hospital within 30 days of discharge

varied between 7.07-13.59% across hospitals (median: 10.16%, IQR: 9.34-10.86%) at the reference level of all model factors. Some (14/24) metrics of antibiotic use suggested re-admission risk was higher the greater the hospital-level antibiotic use, while the remainder (10/24) showed no evidence of association with re-admission risk (appendix **Table 2.10**).

2.3.4 Scope for achieving antibiotic prescribing reductions

If hospitals with antibiotic use above the 25th percentile (1,624 DDDs/1,000 bed-days; 101 hospitals), 10th percentile (1,454 DDDs/1,000 bed-days; 121 hospitals), or 5th percentile (1,282 DDDs/1,000 bed-days; 128 hospitals) reduced their consumption to this level, total DDD use would decline by 21.6% (from 51,732,671 to 40,558,491 DDDs), 27.0% (from 58,838,197 to 42,953,040 DDDs), or 34.4% (from 61,250,673 to 40,153,594 DDDs), respectively (**Figure 2.6**). With antibiotic use measured in DDDs/1,000 admissions, reducing consumption to the 25th percentile (3,604 DDDs/1,000 admissions), 10th percentile (3,198 DDDs/1,000 admissions), or 5th percentile (2,992 DDDs/1,000 admissions) would drop total DDD use by 23.2%, 28.4%, and 31.8%, respectively.

Figure 2.6 Distribution of antibiotic use by NHS acute hospital Trust



2.4 Discussion

I found very wide variation in the quantity of antibiotics being consumed across NHS acute hospitals, which is consistent with previous reports(59), including a systematic review of antibiotic consumption in 3130 (primarily European) hospitals which found a 40-fold difference among studies(58). By calculating the confounding-adjusted probability of death among general medicine inpatients in each hospital I was able to control extensively for case-mix and found no evidence that variation in antibiotic use is associated with 14 or 30-day mortality. This finding was consistent across different measures of antibiotic consumption and in multiple sensitivity analyses focusing on sub-populations. These results suggest further opportunities for antibiotic reduction at hospitals may potentially be possible without negatively impacting patient outcomes, although this is based on an ecological analysis and therefore does need to be interpreted carefully. Rather than set an arbitrary threshold for what reduction in use could be achieved, I considered the impact of reducing hospital antibiotic use to the 25th, 10th and 5th percentiles among hospitals. Depending on the threshold, the models indicate that system-wide reductions of up to one-third of DDDs could be achieved safely if high-using hospitals could replicate prescribing practices of lower using hospitals.

This chapter has not addressed how some Trusts achieve similar patient outcomes with much lower antibiotic consumption. This needs further exploration but it is likely that the quality of antibiotic prescription reviews is a major factor. Safe control of antibiotic overuse in hospital depends on balancing the need to initiate prompt effective therapy when bacterial infection is present with early review and revision of antibiotic prescriptions in the light of clinical and diagnostic data(118,119). This is challenging in clinical practice, though hospitals which do this well have lower rates of *C. difficile* infection(60) and the ARK-Hospital trial has shown interventions to increase stop rates at review can be safe and effective at reducing antibiotic use in hospital settings(105,121). It is to be expected that at hospitals where antibiotic prescription reviews are done well, fewer antibiotics would be used without compromising patient outcomes. Data from medical inpatients in Oxford

showed substantial reductions in antibiotic use (20-30%) could be achieved without adverse clinical outcomes when patients were admitted under an infectious diseases specialist versus other clinical teams(56). This is in keeping with a growing body of evidence that reducing antibiotic treatment duration across a wide-range of clinical scenarios is a safe and effective way of controlling antibiotic overuse(21,61,148–153). This chapter suggests the magnitude of reductions that might be safely achieved (subject to the caveats above) dwarf the 1% year-on-year reductions previously required of NHS hospitals(128), or the 6.5% reduction in broad-spectrum antibiotics now required over 6 years from 2018-2023(154).

This chapter has important limitations, notably the fact that it is an ecological comparison at the hospital-level, albeit of outcomes adjusted for patient-level characteristics. Patient-level factors significantly associated with mortality risk such as baseline haematology and biochemistry test results and admission time of day are not available in HES(86). Any residual bias arising because hospitals with more acutely unwell patients are likely to use more antibiotics and to have patients at greater risk of death could mean I failed to detect harm associated with low antibiotic use. Some of the analyses indicated greater use of parenteral antibiotics (per 1,000 bed-days) was associated with higher risk of death, which is consistent with residual confounding after adjustment. Nevertheless, the magnitude of differences in antibiotic use observed and the marginal impact on mortality means residual confounders would have to be exerting a very large effect to meaningfully change my inferences.

Although I found no consistent evidence that lower antibiotic use was associated with higher risk of readmission to hospital, this result should be interpreted with caution. More than half of my analyses using different measures of antibiotic consumption indicated higher antibiotic use to be associated with greater risk of non-elective readmission. It may be that some hospitals which use more antibiotics also discharge more quickly, for example discharging patients on oral antibiotics as a safety net, and there is evidence from the United States indicating patients discharged against

medical advice have higher readmission rates(155). Alternatively this may be because the model factors, selected *a priori* based on previous analyses to adjust for case-mix in mortality models(86), are not sufficient to control for confounding of the relationship between antibiotic use and readmission. While this does not undermine my fundamental observation that lower antibiotic use is not associated with case-mix adjusted mortality, it is a reminder that other, albeit less impactful harms, could be associated with undertreatment of infection and should be investigated.

I did not restrict the analysis to sub-populations of general medical inpatients with specific diagnoses for several reasons. The diagnosis (ICD-10) codes used to derive CCS group, Charlson comorbidity index, and immunosuppression status, are captured in HES by clinical coding departments using discharge summaries. This process is both standardised and audited but primarily serves administrative and reimbursement purposes and describes the main condition managed in an episode, which may be ruled out at a later date or be unrelated to the clinical indication for antibiotics. Diagnostic coding practices also vary by hospital, with coding depth and accuracy improving over time(129). As a result, reliably identifying sub-populations based on these codes is challenging, with varied and potentially unknown sensitivity and specificity.

Another important limitation of this chapter is that marginal effects were derived from outcomes in acute/general medicine admissions between 2010-2017, yet antibiotic data was not available at the specialty-level during this entire period or at all before April/2014. As a result, the share of DDDs attributable to general medical inpatients likely varies by hospital, though this drawback may be partially off-set by the broad definition of general medicine used (**Figure 2.1**) and the fact that general medical inpatients are the largest consumers of non-prophylactic antibiotics in hospitals(156). Also, sensitivity analyses restricting to April/2014-March/2017 yielded marginal effects that were very similar to those in the primary analysis and similarly showed no evidence of association between 30-day mortality and hospital-level antibiotic use. The antibiotic data for 2014-

2017 showed most hospitals' use was fairly stable over this period, supporting my comparison of average effects.

While the use of DDDs enables standardised comparisons of hospital antibiotic use without the need for patient-level data, it also has drawbacks(157). For example, to reduce the selective pressure for resistance, antibiotic policies may recommend combinations of narrow-spectrum agents rather than one broad-spectrum agent. This would increase DDDs for treatment of the same indication, producing an apparent increase in antibiotic use. Local differences in the prevalence of AMR may also require some hospitals to use multiple agents to provide sufficient empiric coverage. Differences between prescribed doses and WHO's DDD reference values could also lead to overestimates or underestimates of antibiotic use, depending on local treatment guidelines. Unfortunately, linkage of clinical data with electronic prescribing data is not yet widely available in England.

Despite these limitations, analyses of electronic health records such as this allow the impact of antibiotic therapy on patient outcomes to be assessed on a scale that cannot be achieved with other approaches. Where richer patient-level antibiotic data are available, investigations could analyse inappropriate prescribing and the use of alternate measures of antibiotic use that mitigate the limitations of DDDs, such as length of therapy (LOT) (i.e. days between first and last administered antibiotic inclusive) or days of therapy (DOT) (i.e. the sum of days between first and last administered antibiotic inclusive, with each antibiotic counted separately)(157).

Future observational work could also expand outcome measures to include other potential markers of harm, including hospital length of stay or admission to intensive care. By using HES data I have not been able to assess potential benefits of lower antibiotic prescribing, such as declines in AMR or *C. difficile* infection. Such data would need to either be obtained from individual hospitals, or only analysed at a hospital-level (rather than within individual general medicine inpatients) by using mandatory reported surveillance data.

2.4.1 Conclusion

I found no evidence that the wide variation in hospital antibiotic prescribing was associated with mortality risk in medical inpatients. Accordingly, risk-adjusted benchmarking of antibiotic use in hospitals could, at least in theory, be used to drive safe and substantial reductions in antibiotic consumption. Further investigations should consider how some hospitals achieve low levels of antibiotic use. Understanding what explains these differences could facilitate the design of interventions that can be evaluated in randomised trials for unbiased inference. Although trials are costly and difficult to implement, the results of this chapter provide evidence to justify this further work.

2.5 Appendix

2.5.1 Collected HES data fields

The HES data contained a pseudonymised patient ID that is unique across hospital Trusts and over time(158), patient information (sex, age at admission, ethnicity), clinical information (diagnoses), geographic information (admitting hospital, IMD score), and administrative information such as admission and discharge dates, method and source of admission, method and destination of discharge, patient classification (day-case or ordinary admission), the specialties under which attending consultants were contracted or worked, and the start and end dates of consultant care episodes. The IMD score was based on the 2010 English index of deprivation associated with the Super Output Area Level of each patient's residence at the time of admission (i.e. postal codes were not collected); more recent 2019 and 2015 English deprivation indices are available, but HES was still using the 2010 index when the data was provided by NHS Digital(138).

2.5.2 Preparation of HES data and model selection

In the final analytic cohort (n=36,124,372 spells), the nearest preceding or subsequent spells from the same patient were used to impute missing Index of Multiple Deprivation (IMD) score (n=83,642). Missing sex (n=1,547) was imputed to the mode (female), as was missing admission method (n=3,832) (mode: elective and other non-emergency), and missing admission source (n=36,848) (mode: usual/other place of residence).

Following NHS Digital guidance(159), the primary diagnosis code was used to derive 142 CCS diagnosis groups, while secondary diagnosis fields were used to derive a Charlson Comorbidity Index. Following an approach reported in previous analyses(86), the CCS groups were aggregated into 43 subgroups based on clinical feedback, and categories with low (<0.5%) observed mortality were collapsed into a single 'low risk' group. To improve model stability, the smallest remaining categories (cumulatively accounting for 6.11% of all spells) were then collapsed into two 'other'

categories with 30-day mortality risk above and below the overall mortality risk, respectively. This process yielded a total of 29 CCS categories shown in appendix **Table 2.5**. A binary indicator of immunosuppression was defined from secondary diagnosis code indicating metastatic cancer, severe liver disease, or HIV.

The categories used to define primary admission specialty were based on the descending prevalence of consultant specialties. First, spells with a main specialty code (HES: mainspef) or treatment specialty code (HES: tretspef) of 300 in the first or second consultant episode were defined as acute/general medicine. If this code was not present then the next most common admission specialty code was used. This logic was then repeated for each of the specialties included in the sample population (**Figure 2.1**). To improve model stability, individual admission specialties comprising <1 million spells were grouped as "other", including endocrinology, diabetic medicine, acute internal medicine, respiratory medicine, infectious diseases, neurology, and rheumatology. In the primary analysis (n=36,124,372 spells), Royal Surrey NHS Foundation Trust had very few admissions (n=182) in the reference group of admission specialty (general medicine) and yielded the highest adjusted probability estimate for 30-day mortality (2.96%, 95% CI: 0.84,5.07). In meta-regression models this point estimate was therefore truncated down to the 99th percentile.

Ethnic categories (HES: ethnos) were defined as white (ethnos="British (White)", "Irish (White)", "Any other White background"), Asian (ethnos="Indian (Asian or Asian British)", "Pakistani (Asian or Asian British)", "Bangladeshi (Asian or Asian British)", "Any other Asian background", "Chinese (other ethnic group)"), black (ethnos="Caribbean (Black or Black British)", "African (Black or Black British)", "Any other Black background"), other (ethnos="White and Black Caribbean (Mixed)", "White and Black African (Mixed)", "White and Asian (Mixed)", "Any other Mixed background", "Any other ethnic group"), and unknown (ethnos="Not stated", "Not known"). Admission method categories (HES: admimeth) were defined as A&E (admimeth="2A", "21"), elective and other non-emergency (admimeth="11", "12", "13", "25", "31", "32", "81", "82", "83"), or emergency via GP or other source

("22", "23", "24", "28", "2B", "2D"). Admission source categories (HES admisorc) were defined as usual/other place of residence (admisorc="19", "29", "39", "66", "79") or NHS general ward/other care provider (admisorc="49", "51", "52", "53", "54", "65", "85", "87", "88"). The definition of each admission method and source category can be found in the HES APC data dictionary(138). Intended management was collected from HES but not included in regression models as it was unknown in 53.8% (n=19,417,403) of spells.

Patient classification categories were collapsed, with "Regular night attender" (n=1,113) being combined with "Regular day attender", "Mothers and babies using only delivery facilities" (n=168) being combined with "Ordinary admission", and "Day-case admission" being combined with "Ordinary admission" if length of stay was >1 day (n=92).

If no outcomes were observed in one of the categories of a model factor (for example, no deaths among "regular day attenders"), the category was combined with the mode for that hospital Trust only. One exception was the "low risk" category of CCS group which was combined with the "Other" CCS group where mortality risk was below the overall risk (appendix **Table 2.5**). In the primary analysis of 30-day mortality, such changes were made to 13/135 hospital Trusts and affected only one or two variables where categories concerned had very few admissions (often <0.1% of total admissions within the hospital).

Substantial collinearity was observed between alternate measures of previous hospital exposure, including the number of overnight admissions in the past year versus ever (Spearman's ρ: 0.73, p<0.0001) and the number of complex overnight admissions in the past year versus ever (Spearman's ρ: 0.75, p<0.0001). Each measure was assessed in a single model adjusting for all hospital Trusts, as was a binary indicator for ever having had a complex admission in the past year. Factors were selected for inclusion in multivariate models based on minimizing the BIC in these single models, and included overnight admissions in the past year and any complex admission in the past year.

To improve model stability, age at admission (years), Charlson Comorbidity Index, and overnight admissions in the past year were truncated at the 99th percentile. To account for non-linear effects of continuous variables, age, Charlson score, IMD score, and overnight admissions in the past year were fit as natural cubic splines, as they lowered BIC of models adjusting for all hospital Trusts. The use of natural cubic splines provides similar prediction performance to fractional polynomials(160) and fits a linear function before the first knot and after the last knot, and a piecewise cubic polynomial between the remaining knots. The number of knots (up to six) was chosen based on BIC, with five Harrell knots(161) chosen for age, six Harrell knots chosen for IMD score, six knots chosen for Charlson score (with knots placed at the at the 10th, 35th, 50th, 65th, 80th, and 90th percentile of the non-zero distribution), and three knots chosen for overnight admissions in the past year (with knots placed at the 10th, 70th, and 90th percentile of the non-zero distribution) (appendix **Figure 2.7**). In the latter two covariates, placement of knots at fixed intervals was not possible due to limited variation in the truncated distributions. Adjusted models retained the same number of knots, which remained significant ($p<0.001$) and were visualised to assess over-fitting.

Nine interaction terms included in previous analyses(86) were added individually to a single multivariate model including all hospital Trusts and found to significantly improve the model fit (lowered the BIC); these interactions were between age and Charlson score, admission method and calendar year, admission method and overnight admissions in the past year, admission specialty and overnight admissions in the past year, admission specialty and Charlson score, overnight admissions in the past year and Charlson score, ever had a complex overnight admission in the past year and Charlson score, overnight admissions in the past year and ever had a complex overnight admission in the last year, age and overnight admissions in the last year.

Table 2.2 Changes to HES provider codes due to splitting and mergers between NHS hospital Trusts, April 2010 to March 2017

Old provider code [i]	Updated provider code	Financial year when change occurred
Worthing and Southlands Hospitals NHS Trust (RPL), and Royal West Sussex NHS Trust (RPR)	Western Sussex Hospitals NHS Foundation Trust (RYR)	2010/11
Nuffield Orthopaedic Centre NHS Trust (RBF)	Oxford University Hospitals NHS Foundation Trust (RTH)	2012/13
Winchester and Eastleigh Healthcare NHS Trust (RN1)	Hampshire Hospitals NHS Foundation Trust (RN5)	2012/13
Trafford Healthcare NHS Trust (RM4)	Central Manchester University Hospitals NHS Foundation Trust (RW3)	2012/13
Scarborough and North East Yorkshire Health Care NHS Trust (RCC)	York Teaching Hospital NHS Foundation Trust (RCB)	2012/13
Newham University Hospital NHS Trust (RNH), and Whipps Cross University Hospital NHS Trust (RGC), and Barts and The London NHS Trust (RNJ)	Barts Health NHS Trust (R1H)	2013/14
South London Healthcare NHS Trust (RYQ)	King's College Hospital NHS Foundation Trust (RJZ) [ii]	2014/15
Mid Staffordshire NHS Foundation Trust (RJD)	University Hospitals of North Midlands NHS Trust (RJE) [ii]	2014/15
Ealing Hospital NHS Trust (RC3), and North West London Hospitals NHS Trust (RV8)	London North West Healthcare NHS Trust (R1K)	2015/16
Barnet and Chase Farm Hospitals NHS Trust (RVL)	Royal Free London NHS Foundation Trust (RAL)	2015/16
Royal National Hospital for Rheumatic Diseases NHS Foundation Trust (RBB)	Royal United Hospitals Bath NHS Foundation Trust (RD1)	2015/16
Heatherwood and Wexham Park Hospitals NHS Foundation Trust (RD7)	Frimley Health NHS Foundation Trust (RDU)	2015/16
West Middlesex University Hospital NHS Trust (RFW)	Chelsea and Westminster Hospital NHS Foundation Trust (RQM)	2015/16
Torbay and Southern Devon Health and Care NHS Trust (R1G)	Torbay and South Devon Health Care NHS Foundation Trust (RA9)	2015/16

[i] Note other Trust mergers and splitting occurred during the study timeframe, but are excluded from this table because the provider codes were either already updated in the raw data files or excluded.

[ii] RYQ split into RJZ, RJ2, and RPG. RJD split into RJE and RL4. Since RJZ and RJE had the most admissions per year(136), these provider codes were assumed.

Table 2.3 Classification broad-spectrum, Access, Watch, and Reserve antibiotics, adapted by UKHSA from the World Health Organization 2017 Essential Medicines List

Antibiotics	ATC index code 2020	Classified as broad-spectrum[1]	AWaRe England category
Amikacin	J01GB06	No	Watch
Amoxicillin	J01CA04	No	Access
Amoxicillin/Clavulanic Acid	J01CR02	Yes	Watch
Ampicillin	J01CA01	No	Access
Ampicillin/Flucloxacillin	J01CA51	No	Access
Ampicillin/Flucloxacillin	J01CR50	No	Access
Avibactam/Ceftazidime	J01DD52	Yes	Reserve
Azithromycin	J01FA10	No	Access or Watch[2]
Aztreonam	J01DF01	Yes	Reserve
Cefaclor[3]	J01DC04	No	Watch
Cefadroxil	J01DB05	No	Watch
Cefalexin	J01DB01	No	Watch
Cefixime	J01DD08	Yes	Access or Watch[2]
Cefotaxime	J01DD01	Yes	Access or Watch[2]
Cefpodoxime Proxetil	J01DD13	Yes	Watch
Cefradine	J01DB09	No	Watch
Ceftaroline Fosamil	J01DI02	Yes	Reserve
Ceftazidime	J01DD02	Yes	Watch
Ceftobiprole Medocaril	J01DI01	Yes	Reserve
Ceftolozane/Tazobactam	J01DI54	Yes	Reserve
Ceftriaxone	J01DD04	Yes	Access or Watch[2]
Cefuroxime	J01DC02	Yes	Watch
Cefuroxime Axetil	J01DC02	Yes	Watch
Chloramphenicol	J01BA01	No	Watch
Cilastatin/Imipenem	J01DH51	Yes	Reserve
Ciprofloxacin	J01MA02	Yes	Access or Watch[2]
Clarithromycin	J01FA09	No	Access or Watch[2]
Clindamycin	J01FF01	No	Watch
Colistin	J01XB01	Yes	Reserve
Dalfopristin/Quinupristin	J01FG02	No	Watch
Daptomycin	J01XX09	No	Reserve
Doripenem	J01DH04	Yes	Reserve
Doxycycline	J01AA02	No	Access
Ertapenem	J01DH03	Yes	Reserve
Erythromycin	J01FA01	No	Watch
Fidaxomicin[4]	A07AA12	No	Watch
Flucloxacillin	J01CF05	No	Access
Fosfomycin	J01XX01	No	Reserve

Antibiotics	ATC index code 2020	Classified as broad-spectrum[1]	AWaRe England category
Fusidic Acid	J01XC01	No	Access
Gentamicin	J01GB03	No	Access
Levofloxacin	J01MA12	Yes	Watch
Linezolid	J01XX08	No	Reserve
Lymecycline	J01AA04	No	Watch
Meropenem	J01DH02	Yes	Reserve
Metronidazole	J01XD01	No	Access
Minocycline	J01AA08	No	Watch
Moxifloxacin	J01MA14	Yes	Watch
Nitrofurantoin	J01XE01	No	Access
Norfloxacin	J01MA06	Yes	Watch
Ofloxacin	J01MA01	Yes	Watch
Oxytetracycline	J01AA06	No	Watch
Penicillin G	J01CE01	No	Access
Penicillin V	J01CE02	No	Access
Piperacillin	J01CA12	Yes	Watch
Piperacillin/Tazobactam	J01CR05	Yes	Access or Watch[2]
Pivmecillinam	J01CA08	No	Access
Sulfamethoxazole/Trimethoprim	J01EE01	No	Access
Tedizolid	J01XX11	No	Reserve
Teicoplanin	J01XA02	No	Watch
Telavancin	J01XA03	No	Watch
Telithromycin	J01FA15	No	Watch
Temocillin	J01CA17	No	Watch
Tetracycline	J01AA07	No	Access
Tigecycline	J01AA12	Yes	Reserve
Tinidazole	J01XD02	No	Access
Tobramycin	J01GB01	No	Watch
Trimethoprim	J01EA01	No	Access
Vancomycin	J01XA01	No	Access or Watch[2]
Vancomycin[4]	A07AA09	No	Access or Watch[2]

[1] Broad-spectrum antibiotics were defined as co-amoxiclav, piperacillin/tazobactam, second (e.g. cefuroxime), third (e.g. ceftriaxone, ceftazidime), fourth (e.g. cefepime) or fifth (e.g. ceftobiprole) generation cephalosporins, cephalosporin-beta-lactamase inhibitor combinations (e.g. ceftazidime-avibactam, ceftolozane-tazobactam), carbapenems, quinolones, azithromycin, tigecycline, aztreonam, telithromycin, and polymyxins. All other antibiotics were classified as narrow-spectrum.

[2] AWaRe category depended on indication (not known, therefore analysed as a separate category)

[3] Although second-generation cephalosporins were classified as broad-spectrum, Cefaclor was not as it is administered orally and is not well absorbed.

[4] Though not absorbed, oral fidaxomicin (A07AA12) and vancomycin (A07AA09) are important treatment options for *C. difficile* and were retained as they accounted for 0.8% of total DDDs.

Table 2.4 Number of acute and general medical admissions in 135 acute NHS hospital Trusts, England, April 2010 to March 2017

NHS acute hospital Trust	Admissions (N=36,124,372)
Barts Health NHS Trust	655,728 (1.8%)
Heart of England NHS Foundation Trust	600,781 (1.7%)
University Hospitals of Leicester NHS Trust	592,378 (1.6%)
King's College Hospital NHS Foundation Trust	577,624 (1.6%)
Pennine Acute Hospitals NHS Trust	563,816 (1.6%)
Sheffield Teaching Hospitals NHS Foundation Trust	563,268 (1.6%)
Leeds Teaching Hospitals NHS Trust	535,954 (1.5%)
University Hospital of North Midlands NHS Trust	515,228 (1.4%)
East Kent Hospitals University NHS Foundation Trust	500,870 (1.4%)
The Royal Wolverhampton NHS Trust	497,593 (1.4%)
Shrewsbury and Telford Hospital NHS Trust	478,043 (1.3%)
Nottingham University Hospitals NHS Trust	475,821 (1.3%)
Imperial College Healthcare NHS Trust	455,868 (1.3%)
Royal Devon and Exeter NHS Foundation Trust	442,324 (1.2%)
Norfolk and Norwich University Hospitals NHS Foundation Trust	439,213 (1.2%)
University College London NHS Foundation Trust	431,899 (1.2%)
Oxford University Hospitals NHS Foundation Trust	431,286 (1.2%)
Royal Free London NHS Foundation Trust	424,790 (1.2%)
Frimley Health NHS Foundation Trust	423,701 (1.2%)
London North West Healthcare NHS Trust	422,383 (1.2%)
The Newcastle Upon Tyne Hospitals NHS Foundation Trust	415,388 (1.2%)
South Tees Hospitals NHS Foundation Trust	397,418 (1.1%)
Mid Yorkshire Hospitals NHS Trust	387,113 (1.1%)
York Teaching Hospital NHS Foundation Trust	374,825 (1.0%)
The Dudley Group NHS Foundation Trust	372,312 (1.0%)
Northumbria Healthcare NHS Foundation Trust	371,996 (1.0%)
Barking, Havering and Redbridge University Hospitals NHS Trust	366,606 (1.0%)
East Lancashire Hospitals NHS Trust	364,950 (1.0%)
County Durham and Darlington NHS Foundation Trust	355,572 (1.0%)
Royal Cornwall Hospitals NHS Trust	355,337 (1.0%)
Hull and East Yorkshire Hospitals NHS Trust	352,449 (1.0%)
Central Manchester University Hospitals NHS Foundation Trust	346,302 (1.0%)
United Lincolnshire Hospitals NHS Trust	344,500 (1.0%)
The Royal Bournemouth and Christchurch Hospitals NHS Foundation Trust	341,723 (1.0%)
University Hospital Birmingham NHS Foundation Trust	337,519 (0.9%)
University Hospitals Coventry and Warwickshire NHS Trust	335,214 (0.9%)
Portsmouth Hospitals NHS Trust	330,267 (0.9%)
University Hospital Southampton NHS Foundation Trust	329,881 (0.9%)

NHS acute hospital Trust	Admissions (N=36,124,372)
Gloucestershire Hospitals NHS Foundation Trust	326,409 (0.9%)
Derby Teaching Hospitals NHS Foundation Trust	325,309 (0.9%)
Western Sussex Hospitals NHS Foundation Trust	323,149 (0.9%)
Cambridge University Hospitals NHS Foundation Trust	317,725 (0.9%)
St George's University Hospitals NHS Foundation Trust	310,309 (0.9%)
Aintree University Hospital NHS Foundation Trust	304,050 (0.8%)
Worcestershire Acute Hospitals NHS Trust	302,599 (0.8%)
Royal Liverpool and Broadgreen University Hospitals NHS Trust	295,000 (0.8%)
Doncaster and Bassetlaw Teaching Hospitals NHS Foundation Trust	293,277 (0.8%)
Sandwell and West Birmingham Hospitals NHS Trust	290,692 (0.8%)
Bradford Teaching Hospitals NHS Foundation Trust	290,347 (0.8%)
Guy's and St Thomas' NHS Foundation Trust	288,611 (0.8%)
Wirral University Teaching Hospital NHS Foundation Trust	279,952 (0.8%)
Lancashire Teaching Hospitals NHS Foundation Trust	276,100 (0.8%)
Blackpool Teaching Hospitals NHS Foundation Trust	272,910 (0.8%)
Brighton and Sussex University Hospitals NHS Trust	272,827 (0.8%)
East Sussex Healthcare NHS Trust	265,799 (0.7%)
Hampshire Hospitals NHS Foundation Trust	262,829 (0.7%)
North Tees and Hartlepool NHS Foundation Trust	262,443 (0.7%)
University Hospitals Bristol NHS Foundation Trust	255,811 (0.7%)
Maidstone and Tunbridge Wells NHS Trust	255,481 (0.7%)
Salford Royal NHS Foundation Trust	255,241 (0.7%)
West Hertfordshire Hospitals NHS Trust	252,202 (0.7%)
Chelsea and Westminster Hospital NHS Foundation Trust	250,648 (0.7%)
Royal United Hospitals Bath NHS Foundation Trust	244,224 (0.7%)
North Bristol NHS Trust	243,843 (0.7%)
Basildon and Thurrock University Hospitals NHS Foundation Trust	243,711 (0.7%)
Northern Lincolnshire and Goole NHS Foundation Trust	243,420 (0.7%)
Calderdale and Huddersfield NHS Foundation Trust	238,032 (0.7%)
Plymouth Hospitals NHS Trust	236,544 (0.7%)
St Helens and Knowsley Teaching Hospitals NHS Trust	235,402 (0.7%)
Southend University Hospital NHS Foundation Trust	233,632 (0.7%)
Ipswich Hospital NHS Trust	229,172 (0.6%)
Royal Berkshire NHS Foundation Trust	226,441 (0.6%)
University Hospitals of Morecambe Bay NHS Foundation Trust	225,491 (0.6%)
Colchester Hospital University NHS Foundation Trust	225,100 (0.6%)
City Hospitals Sunderland NHS Foundation Trust	224,731 (0.6%)
University Hospital of South Manchester NHS Foundation Trust	224,210 (0.6%)
Great Western Hospitals NHS Foundation Trust	216,419 (0.6%)
Wrightington, Wigan and Leigh NHS Foundation Trust	216,327 (0.6%)

NHS acute hospital Trust	Admissions (N=36,124,372)
Epsom and St Helier University Hospitals NHS Trust	215,461 (0.6%)
Sherwood Forest Hospitals NHS Foundation Trust	213,972 (0.6%)
Kettering General Hospital NHS Foundation Trust	213,856 (0.6%)
North Cumbria University Hospitals NHS Trust	207,188 (0.6%)
Stockport NHS Foundation Trust	204,489 (0.6%)
The Lewisham and Greenwich NHS Trust	204,330 (0.6%)
Torbay and South Devon Health Care NHS Foundation Trust	204,304 (0.6%)
Taunton and Somerset NHS Foundation Trust	203,272 (0.6%)
Buckinghamshire Healthcare NHS Trust	199,735 (0.6%)
Northampton General Hospital NHS Trust	198,621 (0.6%)
Mid Essex Hospital Services NHS Trust	196,750 (0.5%)
Surrey and Sussex Healthcare NHS Trust	195,349 (0.5%)
Luton and Dunstable University Hospital NHS Foundation Trust	195,087 (0.5%)
Gateshead Health NHS Foundation Trust	192,158 (0.5%)
East and North Hertfordshire NHS Trust	191,475 (0.5%)
Chesterfield Royal Hospital NHS Foundation Trust	190,801 (0.5%)
The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust	190,063 (0.5%)
Bolton NHS Foundation Trust	188,364 (0.5%)
Barnsley Hospital NHS Foundation Trust	181,844 (0.5%)
Poole Hospital NHS Foundation Trust	179,480 (0.5%)
Walsall Healthcare NHS Trust	178,890 (0.5%)
Ashford and St. Peter's Hospitals NHS Foundation Trust	176,576 (0.5%)
Peterborough and Stamford Hospitals NHS Foundation Trust	173,905 (0.5%)
Medway NHS Foundation Trust	169,650 (0.5%)
The Princess Alexandra Hospital NHS Trust	167,726 (0.5%)
The Rotherham NHS Foundation Trust	165,268 (0.5%)
North Middlesex University Hospital NHS Trust	160,818 (0.5%)
Warrington and Halton Hospitals NHS Foundation Trust	159,990 (0.4%)
James Paget University Hospitals NHS Foundation Trust	157,301 (0.4%)
Croydon Health Services NHS Trust	156,603 (0.4%)
Southport and Ormskirk Hospital NHS Trust	155,726 (0.4%)
Milton Keynes Hospital NHS Foundation Trust	152,350 (0.4%)
The Mid Cheshire Hospitals NHS Foundation Trust	151,949 (0.4%)
Tameside Hospital NHS Foundation Trust	150,714 (0.4%)
Dartford and Gravesham NHS Trust	150,517 (0.4%)
Bedford Hospital NHS Trust	148,377 (0.4%)
Kingston Hospital NHS Foundation Trust	147,291 (0.4%)
Burton Hospitals NHS Foundation Trust	144,039 (0.4%)
Airedale NHS Foundation Trust	142,587 (0.4%)
Salisbury NHS Foundation Trust	139,310 (0.4%)

NHS acute hospital Trust	Admissions (N=36,124,372)
Homerton University Hospital NHS Foundation Trust	138,945 (0.4%)
The Hillingdon Hospitals NHS Foundation Trust	138,454 (0.4%)
Dorset County Hospital NHS Foundation Trust	131,000 (0.4%)
South Warwickshire NHS Foundation Trust	130,656 (0.4%)
West Suffolk NHS Foundation Trust	129,822 (0.4%)
Northern Devon Healthcare NHS Trust	127,192 (0.4%)
Countess of Chester Hospital NHS Foundation Trust	124,431 (0.3%)
Royal Surrey County Hospital NHS Foundation Trust	120,376 (0.3%)
South Tyneside NHS Foundation Trust	120,055 (0.3%)
The Whittington Hospital NHS Trust	117,822 (0.3%)
Harrogate and District NHS Foundation Trust	113,270 (0.3%)
Hinchingbrooke Health Care NHS Trust	103,636 (0.3%)
Yeovil District Hospital NHS Foundation Trust	103,543 (0.3%)
George Eliot Hospital NHS Trust	94,417 (0.3%)
East Cheshire NHS Trust	93,900 (0.3%)
Weston Area Health NHS Trust	89,741 (0.3%)
Wye Valley NHS Trust	86,567 (0.2%)

Table 2.5 Clinical Classifications Software (CCS) diagnosis group identified among acute and general medical admissions in 135 acute NHS hospital Trusts, England, April 2010-March 2017

CCS diagnosis groups collapsed into 29 categories	Admissions (N=36,124,372)
Nonspecific chest pain; Cardiac dysrhythmias; Coronary atherosclerosis and other heart disease; Pulmonary heart disease; Essential hypertension, Hypertension with complications/secondary hypertension; Peri/endo/myo-carditis; cardiomyopathy (except if due to tuberculosis/sexually transmitted disease); Conduction disorders; Heart valve disorders; Cardiac arrest and ventricular fibrillation; Other/ill-defined heart disease or circulatory disease	4,649,241 (12.9%)
Cancer of bladder, bone, connective tissue, thyroid, malignant neoplasm without specification of site, cancer of brain, nervous system, breast, bronchus, lung, cervix, other female genital organs, colon, head and neck, kidney/renal pelvis, other urinary organs, liver/intrahepatic bile duct, oesophagus, other GI organs, peritoneum, ovary, pancreas, prostate, testis, or other male genital organs, rectum/anus, stomach, uterus; Maintenance chemotherapy; radiotherapy; Cancer, other respiratory/intrathoracic; Hodgkin's disease, Non-Hodgkin's lymphoma; Leukaemias; Melanomas of skin, Other non-epithelial cancer of skin; Multiple myeloma; Neoplasms of unspecified nature/uncertain behaviour, Non-malignant breast conditions; Pathological fracture; Secondary malignancies	3,848,695 (10.7%)
Headache; including migraine, Cataract, Retinal detachments; defects; vascular occlusion; and retinopathy, Glaucoma, Blindness and vision defects,	2,520,624 (7.0%)

CCS diagnosis groups collapsed into 29 categories	Admissions (N=36,124,372)
Inflammation; infection of eye (except that caused by tuberculosis or sexually transmitted disease), Other eye disorders, Otitis media and related conditions, Conditions associated with dizziness or vertigo, Other ear and sense organ disorders; Syncope; Epilepsy; convulsions; Other nervous system disorders; Malaise and fatigue; Multiple sclerosis, Other hereditary and degenerative nervous system conditions; Parkinson's disease; Paralysis, Late effects of cerebrovascular disease	
Regional enteritis and ulcerative colitis; Diverticulosis and diverticulitis, Anal and rectal conditions; Non-infectious gastroenteritis; Abdominal pain; Abdominal hernia; Intestinal obstruction without hernia; Appendicitis and other appendiceal conditions, Peritonitis and intestinal abscess	2,288,317 (6.3%)
Other gastrointestinal disorders; Gastritis and duodenitis, Other disorders of stomach and duodenum; Gastroduodenal ulcer (except haemorrhage)	2,062,927 (5.7%)
Deficiency and other anaemia, Acute post-haemorrhagic anaemia; Nutritional deficiencies, Disorders of lipid metabolism, Other nutritional; endocrine; and metabolic disorders; Diseases of white blood cells	1,779,260 (4.9%)
Allergic reactions, Rehabilitation care; fitting of prostheses; and adjustment of devices, Administrative/social admission, Medical examination/evaluation, Other aftercare, Other screening for suspected conditions (not mental disorders or infectious disease), Residual codes; unclassified, E Codes: All (external causes of injury and poisoning); Poisoning by psychotropic agents, Poisoning by other medications and drugs, Poisoning by nonmedicinal substances	1,755,860 (4.9%)
Oesophageal disorders; Gastrointestinal haemorrhage	1,546,058 (4.3%)
Other connective tissue disease; Gout and other crystal arthropathies, Rheumatoid arthritis and related disease, Osteoarthritis, Acquired foot deformities, Other acquired deformities, Systemic lupus erythematosus and connective tissue disorders, Other bone disease and musculoskeletal deformities; Other non-traumatic joint disorders	1,530,254 (4.2%)
Pneumonia (except that caused by tuberculosis or sexually transmitted disease); Aspiration pneumonitis; food/vomitus	1,392,396 (3.9%)
Chronic obstructive pulmonary disease and bronchiectasis; Cystic fibrosis, Other lower respiratory disease; Lung disease due to external agents	1,228,385 (3.4%)
Other, outcome risk above overall risk	1,148,971 (3.2%)
Other, outcome risk is below overall risk	1,058,630 (2.9%)
Urinary tract infections; Genitourinary symptoms and ill-defined conditions	838,968 (2.3%)
Acute cerebrovascular disease; Occlusion or stenosis of precerebral arteries, Other and ill-defined cerebrovascular disease, Transient cerebral ischemia; Coma; stupor; and brain damage	805,434 (2.2%)
Acute bronchitis; Asthma	756,552 (2.1%)
Nephritis; nephrosis; renal sclerosis, Chronic renal failure; Calculus of urinary tract, other diseases of kidneys/ureters, other diseases of bladder/urethra	754,331 (2.1%)

CCS diagnosis groups collapsed into 29 categories	Admissions (N=36,124,372)
Other liver diseases; Biliary tract disease; Liver disease; alcohol-related; Pancreatic disorders (not diabetes)	646,577 (1.8%)
Mental retardation, Senility and organic mental disorders; Alcohol-related mental disorders, Substance-related mental disorders, Affective disorders, Anxiety; somatoform; dissociative; and personality disorders; Other psychoses; Schizophrenia and related disorders, Preadult disorders, Other mental conditions, Personal history of mental disorder	641,475 (1.8%)
Superficial injury; contusion; Open wounds of head; neck; and trunk; Crushing injury or internal injury; Fracture of upper limb; Intracranial injury; Other injuries and conditions due to external causes; Open wounds of extremities; Fracture of lower limb; Fracture of neck of femur (hip); Burns	637,873 (1.8%)
Skin and subcutaneous tissue infections; Other inflammatory condition of skin, Chronic ulcer of skin, Other skin disorders	559,582 (1.6%)
Acute myocardial infarction	498,736 (1.4%)
Spondylosis; intervertebral disc disorders; other back problems, Osteoporosis; Joint disorders and dislocations; trauma-related, Spinal cord injury, Skull and face fractures, Other fractures, Sprains and strains	481,424 (1.3%)
Low risk (probability of outcome <0.5%)	478,387 (1.3%)
Immunity disorders, Sickle cell anaemia, Coagulation and haemorrhagic disorders, Other hematologic conditions	458,254 (1.3%)
Influenza, Acute/chronic tonsillitis, Other upper respiratory infections, Other upper respiratory disease, Disorders of teeth/jaw, Diseases of mouth	456,453 (1.3%)
Congestive heart failure; non-hypertensive	454,455 (1.3%)
Intestinal infection	442,769 (1.2%)
Phlebitis; thrombophlebitis and thromboembolism, Varicose veins of lower extremity, Haemorrhoids, Other disease of veins and lymphatics; Lymphadenitis, Gangrene; Peripheral and visceral atherosclerosis; Aortic and peripheral arterial embolism or thrombosis; Aortic; peripheral; and visceral artery aneurysms	403,484 (1.1%)

Table 2.6 Random-effects meta-regression of the association between the adjusted probability of death within 30 days of admission (in or out of hospital) and different metrics of hospital-level antibiotic use, among 36,124,372 general medicine admissions to 135 NHS acute care hospital Trusts, April 2010 – March 2017.

Metric of antibiotic use **For each hospital-level higher:**	**Change in the adjusted probability of 30-day mortality (95% CI)** [1]	**P-value**
500 Total DDDs per 1000 bed-days	-0.010% (-0.064, +0.044)	0.718
500 Total DDDs per 1000 admissions	-0.021% (-0.044, +0.001)	0.065
500 Inpatient DDDs per 1000 bed-days	+0.004% (-0.076, +0.085)	0.917
500 Inpatient DDDs per 1000 admissions	-0.014% (-0.045, +0.017)	0.362
500 Outpatient DDD per 1000 bed-days	-0.021% (-0.081, +0.038)	0.483
500 Outpatient DDD per 1000 admissions	-0.020% (-0.049, +0.008)	0.161
500 Oral DDDs per 1000 bed-days	-0.029% (-0.087, +0.028)	0.316
500 Oral DDDs per 1000 admissions	**-0.028% (-0.053, -0.003)**	**0.028**
500 Parenteral DDDs per 1000 bed-days	**+0.284% (+0.031, +0.538)**	**0.028**
500 Parenteral DDDs per 1000 admissions	+0.040% (-0.065, +0.145)	0.454
500 Narrow-spectrum DDDs per 1000 bed-days	-0.025% (-0.096, +0.046)	0.483
500 Narrow-spectrum DDDs per 1000 admissions	-0.026% (-0.055, +0.003)	0.074
500 Broad-spectrum DDDs per 1000 bed-days	+0.012% (-0.101, +0.125)	0.836
500 Broad-spectrum DDDs per 1000 admissions	-0.019% (-0.072, +0.034)	0.475
500 Access DDDs per 1000 bed-days	-0.033% (-0.126, +0.059)	0.476
500 Access DDDs per 1000 admissions	-0.032% (-0.072, +0.007)	0.102
500 Watch DDDs per 1000 bed-days	+0.035% (-0.083, +0.152)	0.561
500 Watch DDDs per 1000 admissions	-0.009% (-0.065, +0.047)	0.751
500 Access or Watch (depending on indication) DDDs per 1000 bed-days	-0.091% (-0.272, +0.090)	0.323
500 Access or Watch (depending on indication) DDDs per 1000 admissions	-0.078% (-0.157, +0.001)	0.050
100 Reserve DDDs per 1000 bed-days	-0.027% (-0.240, +0.186)	0.800
100 Reserve DDDs per 100 per 1000 admissions	-0.046% (-0.131, +0.040)	0.293
100 Piperacillin/tazobactam and meropenem DDDs per 1000 bed-days	+0.002% (-0.143, +0.147)	0.978
100 Piperacillin/tazobactam and meropenem DDDs per 1000 admissions	-0.023% (-0.085, +0.039)	0.464

[1] Probability of death derived from 135 separate (hospital-specific) Poisson models, with each model adjusting for all the factors in **Table 2.1**. Plots for each model are displayed in appendix **Figure 2.9**.

Table 2.7 Random-effects meta-regression of the association between the adjusted probability of death within 30 days of admission (in or out of hospital) and different metrics of hospital-level antibiotic use, among a more narrowly defined cohort of 19,023,144 general medicine admissions, April 2010 – March 2017

Metric of antibiotic use For each hospital-level higher:	Change in the adjusted probability of 30-day mortality (95% CI) [1]	P-value
500 Total DDDs per 1000 bed-days	+0.006% (-0.058, +0.071)	0.844
500 Total DDDs per 1000 admissions	-0.007% (-0.034, +0.020)	0.626
500 Inpatient DDDs per 1000 bed-days	+0.042% (-0.051, +0.136)	0.374
500 Inpatient DDDs per 1000 admissions	+0.005% (-0.031, +0.041)	0.779
500 Outpatient DDD per 1000 bed-days	-0.021% (-0.092, +0.050)	0.557
500 Outpatient DDD per 1000 admissions	-0.016% (-0.049, +0.017)	0.343
500 Oral DDDs per 1000 bed-days	-0.017% (-0.087, +0.053)	0.630
500 Oral DDDs per 1000 admissions	-0.016% (-0.046, +0.014)	0.299
500 Parenteral DDDs per 1000 bed-days	**+0.373% (+0.082, +0.663)**	**0.012**
500 Parenteral DDDs per 1000 admissions	+0.113% (-0.008, +0.235)	0.068
500 Narrow-spectrum DDDs per 1000 bed-days	-0.013% (-0.097, +0.071)	0.761
500 Narrow-spectrum DDDs per 1000 admissions	-0.014% (-0.048, +0.020)	0.431
500 Broad-spectrum DDDs per 1000 bed-days	+0.066% (-0.069, +0.201)	0.337
500 Broad-spectrum DDDs per 1000 admissions	+0.017% (-0.046, +0.080)	0.592
500 Access DDDs per 1000 bed-days	-0.027% (-0.137, +0.082)	0.624
500 Access DDDs per 1000 admissions	-0.019% (-0.065, +0.027)	0.407
500 Watch DDDs per 1000 bed-days	+0.078% (-0.063, +0.218)	0.276
500 Watch DDDs per 1000 admissions	+0.022% (-0.044, +0.089)	0.508
500 Access or Watch (depending on indication) DDDs per 1000 bed-days	-0.042% (-0.258, +0.173)	0.698
500 Access or Watch (depending on indication) DDDs per 1000 admissions	-0.038% (-0.133, +0.056)	0.426
100 Reserve DDDs per 1000 bed-days	+0.206% (-0.037, +0.450)	0.096
100 Reserve DDDs per 100 per 1000 admissions	+0.055% (-0.043, +0.154)	0.270
100 Piperacillin/tazobactam and meropenem DDDs per 1000 bed-days	+0.096% (-0.072, +0.265)	0.260
100 Piperacillin/tazobactam and meropenem DDDs per 1000 admissions	+0.026% (-0.047, +0.099)	0.486

[1] To improve model stability, five hospital Trusts with fewer than 50,000 admissions were excluded, including City Hospitals Sunderland NHS Foundation Trust, Harrogate and District NHS Foundation Trust, Royal Surrey County Hospital NHS Foundation Trust, The Hillingdon Hospitals NHS Foundation Trust, and The Whittington Hospital NHS Trust. Thus, the probability of death was derived from 130 separate (hospital-specific) Poisson models, with each model adjusting for all the factors in **Table 2.1** excluding admission specialty. Two out of 130 models required the removal of an interaction term between admission method and admission year in order for the model to converge.

Table 2.8 Random-effects meta-regression of the association between the adjusted probability of death within 30 days of admission (in/out of hospital) and different metrics of hospital-level antibiotic use, among 16,492,990 general medicine admissions, 01/April/2014 – 31/March/2017

Metric of antibiotic use For each hospital-level higher:	Change in the adjusted probability of re-admissions (95% CI) [1]	P-value
500 Total DDDs per 1000 bed-days	+0.013 (-0.053, +0.080)	0.691
500 Total DDDs per 1000 admissions	+0.005 (-0.023, +0.034)	0.713
500 Inpatient DDDs per 1000 bed-days	+0.061 (-0.039, +0.160)	0.231
500 Inpatient DDDs per 1000 admissions	+0.021 (-0.017, +0.060)	0.281
500 Outpatient DDD per 1000 bed-days	-0.020 (-0.093, +0.054)	0.592
500 Outpatient DDD per 1000 admissions	-0.009 (-0.044, +0.026)	0.612
500 Oral DDDs per 1000 bed-days	+0.029 (-0.090, +0.148)	0.634
500 Oral DDDs per 1000 admissions	+0.003 (-0.028, +0.035)	0.848
500 Parenteral DDDs per 1000 bed-days	+0.160 (-0.161, +0.481)	0.326
500 Parenteral DDDs per 1000 admissions	+0.075 (-0.057, +0.207)	0.261
500 Narrow-spectrum DDDs per 1000 bed-days	-0.001 (-0.090, +0.088)	0.979
500 Narrow-spectrum DDDs per 1000 admissions	+0.002 (-0.034, +0.039)	0.908
500 Broad-spectrum DDDs per 1000 bed-days	+0.063 (-0.077, +0.204)	0.373
500 Broad-spectrum DDDs per 1000 admissions	+0.031 (-0.034, +0.096)	0.353
500 Access DDDs per 1000 bed-days	-0.025 (-0.142, +0.093)	0.678
500 Access DDDs per 1000 admissions	-0.004 (-0.053, +0.046)	0.886
500 Watch DDDs per 1000 bed-days	+0.092 (-0.054, +0.238)	0.214
500 Watch DDDs per 1000 admissions	+0.043 (-.026, +0.112)	0.215
500 Access or Watch (depending on indication) DDDs per 1000 bed-days	+0.026 (-0.198, +0.250)	0.818
500 Access or Watch (depending on indication) DDDs per 1000 admissions	+0.012 (-0.087, +0.111)	0.811
100 Reserve DDDs per 1000 bed-days	-0.033 (-0.303, +0.238)	0.812
100 Reserve DDDs per 100 per 1000 admissions	-0.023 (-0.131, +0.084)	0.666
100 Piperacillin/tazobactam and meropenem DDDs per 1000 bed-days	+0.001 (-0.180, +0.181)	0.994
100 Piperacillin/tazobactam and meropenem DDDs per 1000 admissions	+0.002 (-0.075, +0.080)	0.952

[1] To improve model stability, four hospital Trusts with fewer than 50,000 admissions were excluded, including East Cheshire NHS Trust, George Eliot Hospital NHS Trust, Weston Area Health NHS Trust, and Wye Valley NHS Trust. Royal Surrey County Hospital was also excluded as it had just nine admissions in the reference group of admission specialty. City Hospitals Sunderland was excluded as it had no remaining admissions in the reference group of admission specialty. The probability of death was therefore derived from 129 separate (hospital-specific) Poisson models, with each model adjusting for all the factors in **Table 2.1**. Charlson comorbidity index and IMD score were modelled as linear predictors, rather than using natural cubic splines, as the use of splines led to poorly fitting models or the failure of models to converge.

Table 2.9 Random-effects meta-regression of the association between the adjusted probability of death within 14 days of admission (in/out of hospital) and different metrics of hospital-level antibiotic use, among 36,124,372 general medicine admissions, 01/ April/2010 – 31/March/2017

Metric of antibiotic use **For each hospital-level higher:**	**Change in the adjusted probability of re-admissions (95% CI)** [1]	**P-value**
500 Total DDDs per 1000 bed-days	-0.009% (-0.054, +0.037)	0.699
500 Total DDDs per 1000 admissions	**-0.019% (-0.038, -0.001)**	**0.047**
500 Inpatient DDDs per 1000 bed-days	+0.003% (-0.065, +0.071)	0.928
500 Inpatient DDDs per 1000 admissions	-0.014% (-0.040, +0.012)	0.290
500 Outpatient DDD per 1000 bed-days	-0.015% (-0.065, +0.036)	0.564
500 Outpatient DDD per 1000 admissions	-0.016% (-0.040, +0.008)	0.186
500 Oral DDDs per 1000 bed-days	-0.027% (-0.076, +0.022)	0.273
500 Oral DDDs per 1000 admissions	**-0.026% (-0.046, -0.005)**	**0.016**
500 Parenteral DDDs per 1000 bed-days	**+0.267% (+0.054, +0.479)**	**0.014**
500 Parenteral DDDs per 1000 admissions	+0.039% (-0.050, +0.128)	0.384
500 Narrow-spectrum DDDs per 1000 bed-days	-0.025% (-0.085, +0.035)	0.412
500 Narrow-spectrum DDDs per 1000 admissions	**-0.026% (-0.050, -0.002)**	**0.036**
500 Broad-spectrum DDDs per 1000 bed-days	+0.023% (-0.072, +0.119)	0.630
500 Broad-spectrum DDDs per 1000 admissions	-0.011% (-0.055, +0.034)	0.639
500 Access DDDs per 1000 bed-days	-0.034% (-0.112, +0.044)	0.392
500 Access DDDs per 1000 admissions	-0.032% (-0.065, +0.001)	0.052
500 Watch DDDs per 1000 bed-days	+0.036% (-0.063, +0.135)	0.475
500 Watch DDDs per 1000 admissions	-0.007% (-0.054, +0.041)	0.784
500 Access or Watch (depending on indication) DDDs per 1000 bed-days	-0.069% (-0.222, +0.083)	0.370
500 Access or Watch (depending on indication) DDDs per 1000 admissions	-0.064% (-0.130, +0.002)	0.056
100 Reserve DDDs per 1000 bed-days	+0.029% (-0.150, +0.208)	0.750
100 Reserve DDDs per 100 per 1000 admissions	-0.021% (-0.093, +0.051)	0.563
100 Piperacillin/tazobactam and meropenem DDDs per 1000 bed-days	-0.003% (-0.125, +0.120)	0.966
100 Piperacillin/tazobactam and meropenem DDDs per 1000 admissions	-0.022% (-0.075, +0.030)	0.399

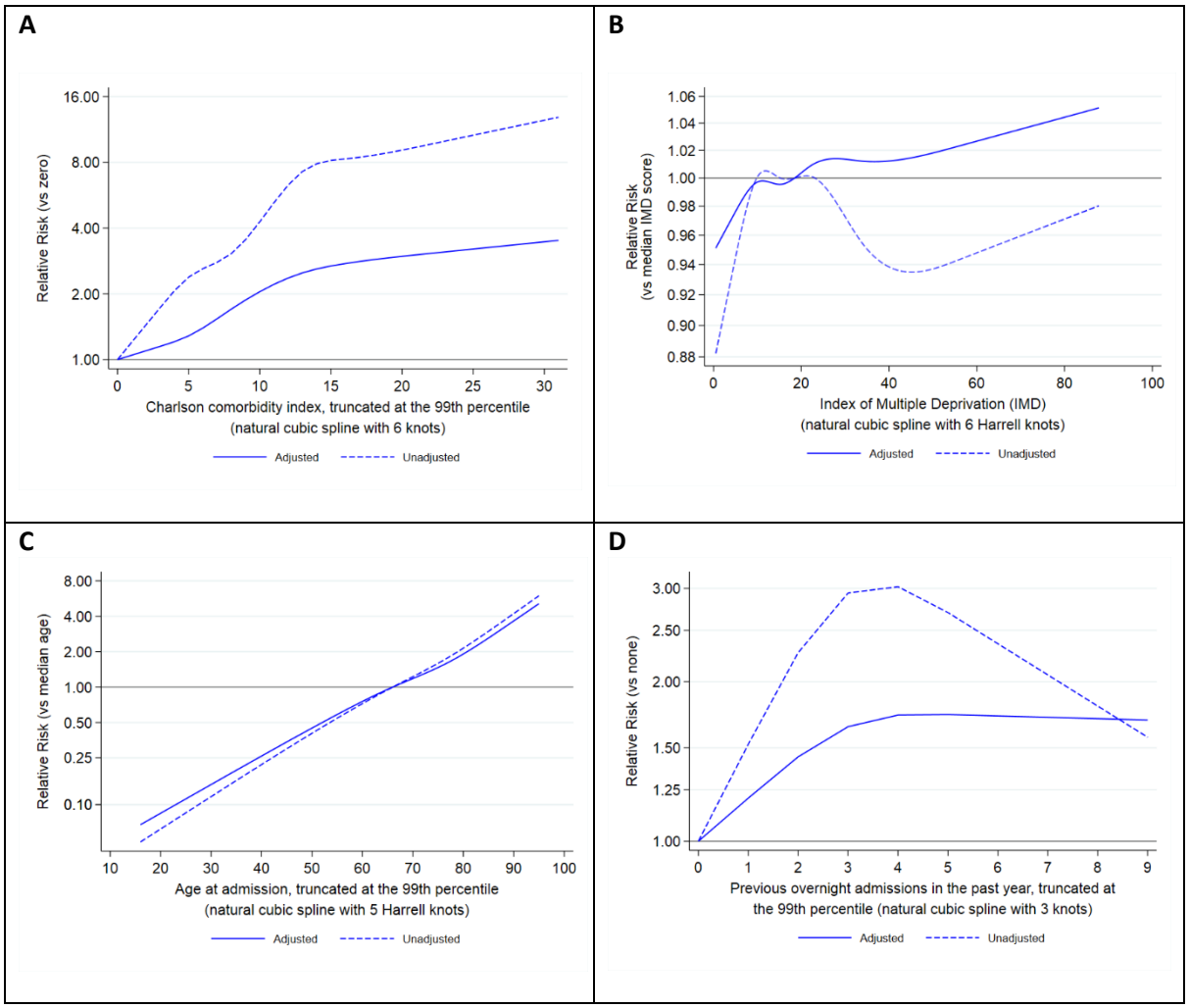
[1] Each model adjusting for all the factors in **Table 2.1**. However, 14 models would only converge with Charlson comorbidity index and IMD score modelled as linear predictors, rather than as natural cubic splines.

Table 2.10 Random-effects meta-regression of the association between the adjusted probability of non-elective re-admission within 30 days of discharge and different metrics of hospital-level antibiotic use, among 34,427,698 general medicine admissions discharged alive, 01/April/2010 – 28/February/2017

Metric of antibiotic use **For each hospital-level higher:**	**Change in the adjusted probability of re-admissions (95% CI)** [1]	**P-value**
500 Total DDDs per 1000 bed-days	**+0.404 (+0.157, +0.652)**	**0.002**
500 Total DDDs per 1000 admissions	**+0.208 (+0.106, +0.311)**	**<0.001**
500 Inpatient DDDs per 1000 bed-days	+0.149 (-0.234, +0.531)	0.443
500 Inpatient DDDs per 1000 admissions	+0.117 (-0.028, +0.261)	0.114
500 Outpatient DDD per 1000 bed-days	**+0.380 (+0.103, +0.657)**	**0.008**
500 Outpatient DDD per 1000 admissions	**+0.213 (+0.082, +0.344)**	**0.002**
500 Oral DDDs per 1000 bed-days	**+0.448 (+0.183, +0.713)**	**0.001**
500 Oral DDDs per 1000 admissions	**+0.233 (+0.120, +0.345)**	**<0.001**
500 Parenteral DDDs per 1000 bed-days	+0.431 (-0.793, +1.655)	0.488
500 Parenteral DDDs per 1000 admissions	+0.463 (-0.031, +0.956)	0.066
500 Narrow-spectrum DDDs per 1000 bed-days	**+0.521 (+0.194, +0.848)**	**0.002**
500 Narrow-spectrum DDDs per 1000 admissions	**+0.254 (+0.124, +0.384)**	**<0.001**
500 Broad-spectrum DDDs per 1000 bed-days	+0.514 (-0.018, +1.046)	0.058
500 Broad-spectrum DDDs per 1000 admissions	**+0.330 (+0.085, +0.575)**	**0.009**
500 Access DDDs per 1000 bed-days	**+0.592 (+0.164, +1.021)**	**0.007**
500 Access DDDs per 1000 admissions	**+0.315 (+0.137, +0.493)**	**0.001**
500 Watch DDDs per 1000 bed-days	**+0.719 (+0.173, +1.266)**	**0.010**
500 Watch DDDs per 1000 admissions	**+0.437 (+0.181, +0.692)**	**0.001**
500 Access or Watch (depending on indication) DDDs per 1000 bed-days	+0.697 (-0.161, +1.554)	0.110
500 Access or Watch (depending on indication) DDDs per 1000 admissions	**+0.436 (+0.067, +0.806)**	**0.021**
100 Reserve DDDs per 1000 bed-days	-0.162 (-1.175, +0.851)	0.753
100 Reserve DDDs per 100 per 1000 admissions	+0.028 (-0.385, +0.440)	0.894
100 Piperacillin/tazobactam and meropenem DDDs per 1000 bed-days	-0.109 (-0.800, +0.581)	0.755
100 Piperacillin/tazobactam and meropenem DDDs per 1000 admissions	+0.073 (-0.222, +0.368)	0.624

[1] Each model adjusted for all the factors in **Table 2.1**. Charlson comorbidity score was modelled as a linear predictors, rather than using natural cubic splines, as the use of splines led to poorly fitting models or the failure of models to converge.

Figure 2.7 Non-linear relative risks of death within 30 days of admission (in or out of hospital) among model covariates fit as natural cubic splines



Adjusted risk estimates derived from a single Poisson model containing NHS hospital Trust as a factor and all admission characteristics in **Table 2.1**.

Figure 2.8 Top 10 most commonly used antibiotics as a fraction of total DDDs within each of 135 NHS acute hospital Trusts (2016), ranked by amoxicillin/clavulanic acid

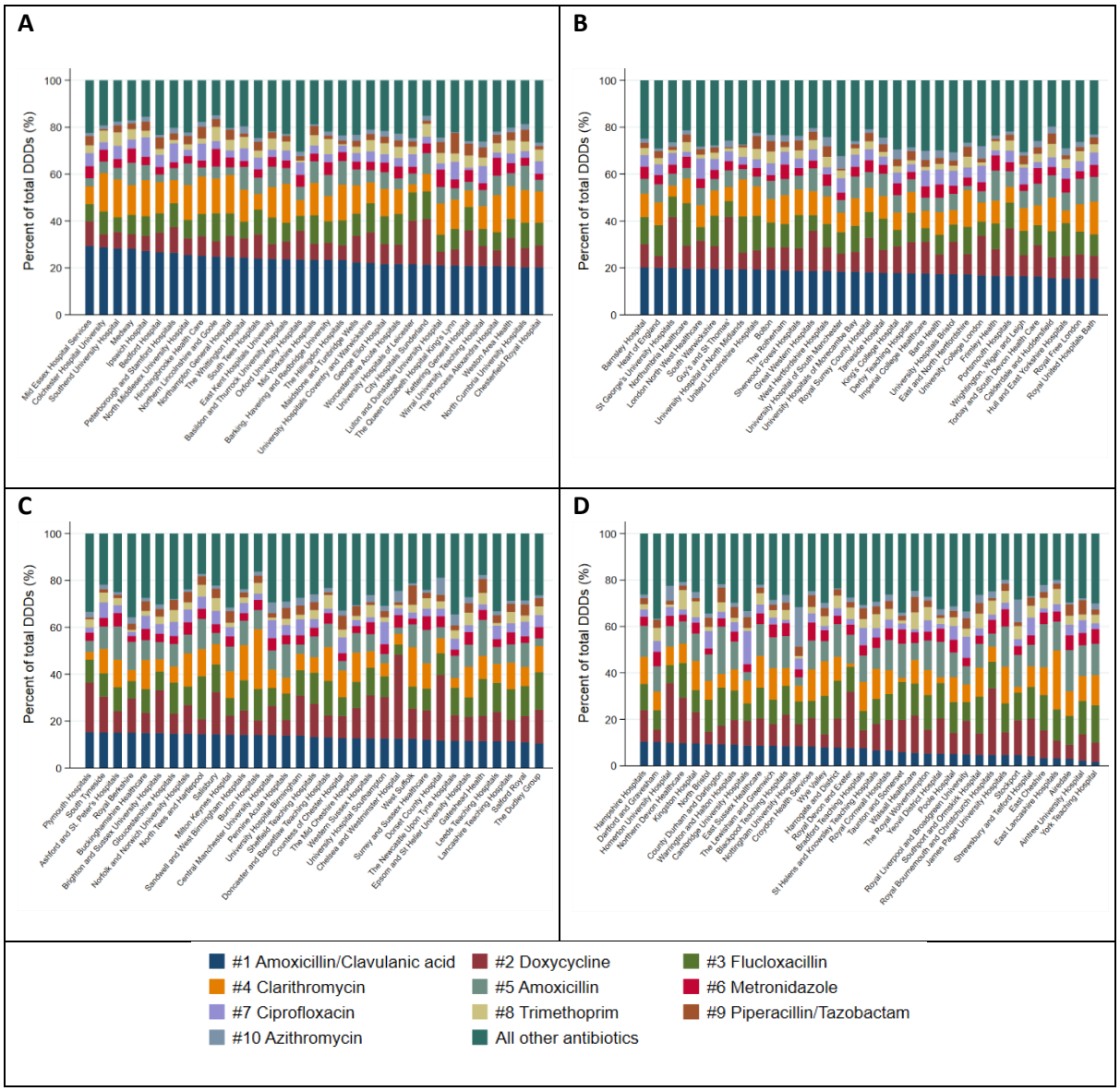
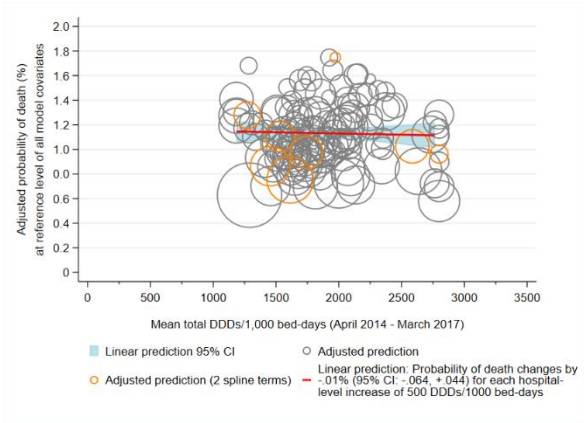
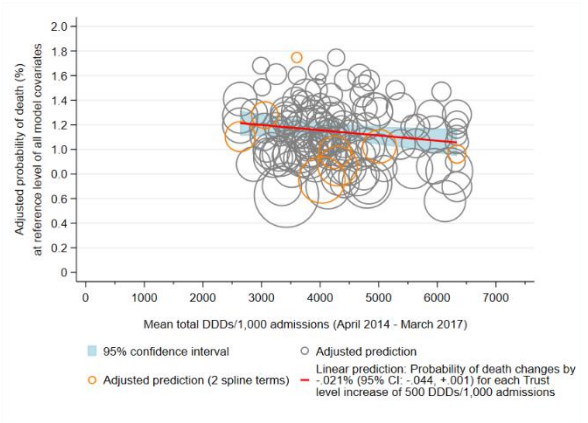


Figure 2.9 Random-effects meta-regression of the association between the adjusted probability of death within 30 days of admission (in or out of hospital) and hospital-level antibiotic use

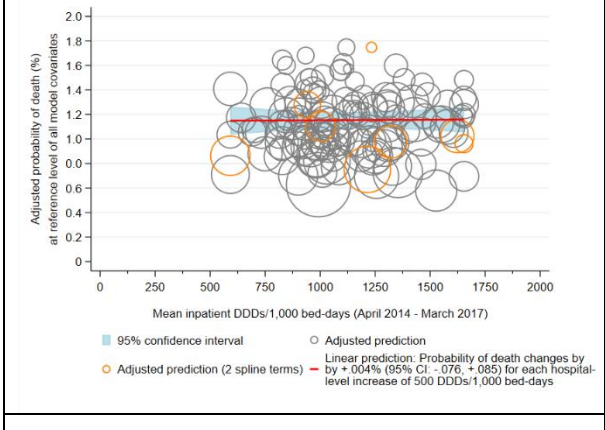
a1) Total DDDs/1,000 bed-days



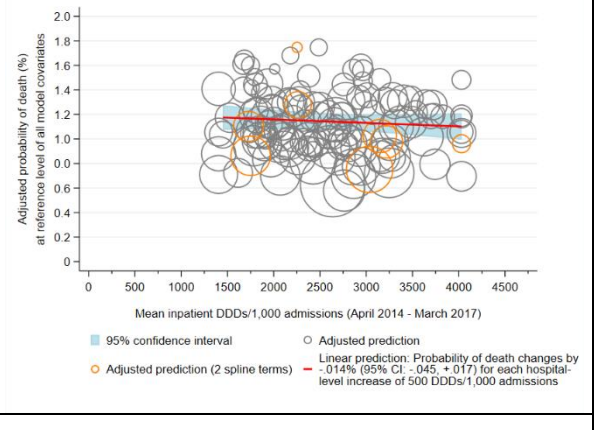
a2) Total DDDs/1,000 admissions



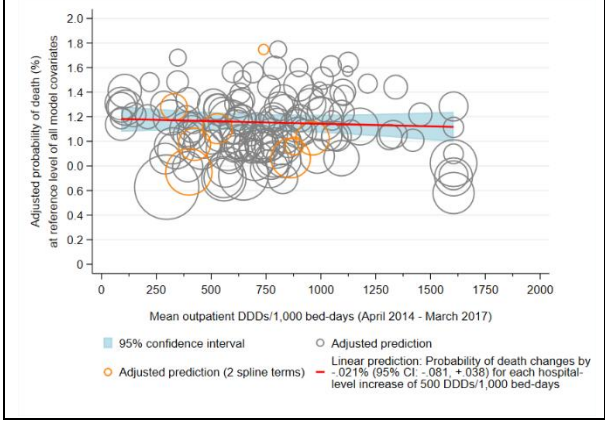
b1) Inpatient DDDs/1,000 bed-days



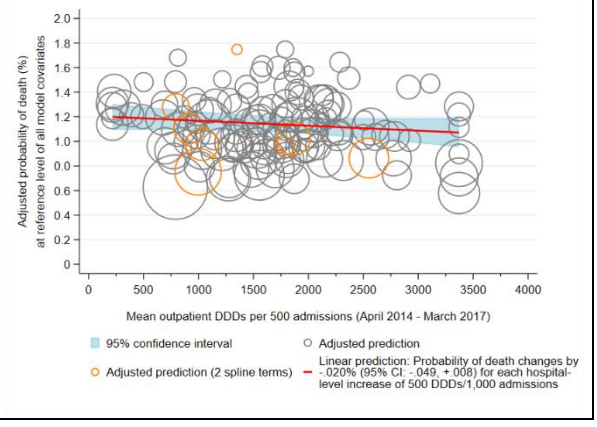
b2) Inpatient DDDs/1,000 admissions



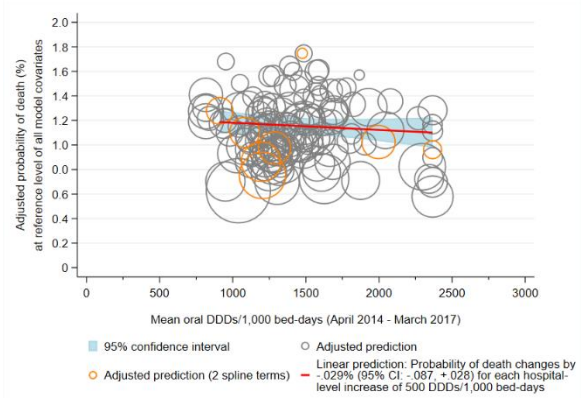
c1) Outpatient DDDs/1,000 bed-days



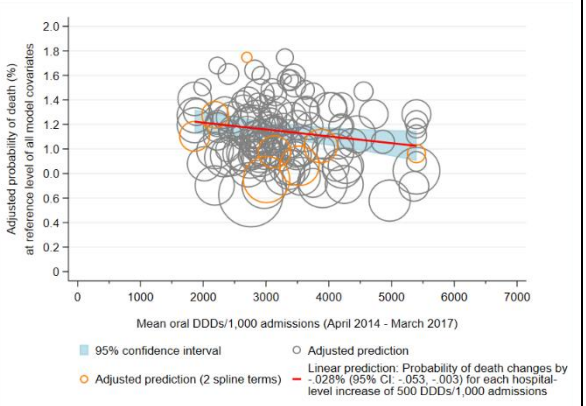
c2) Outpatient DDDs/1,000 admissions



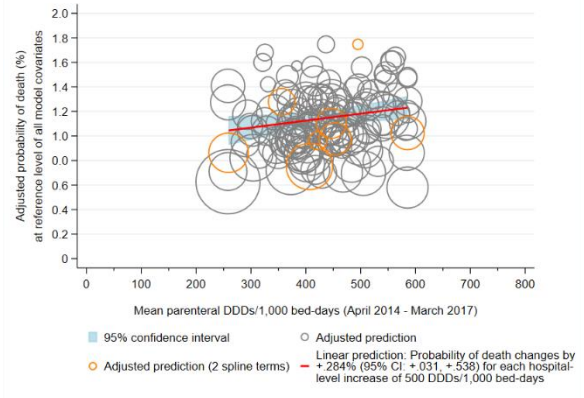
d1) Oral DDDs/1,000 bed-days



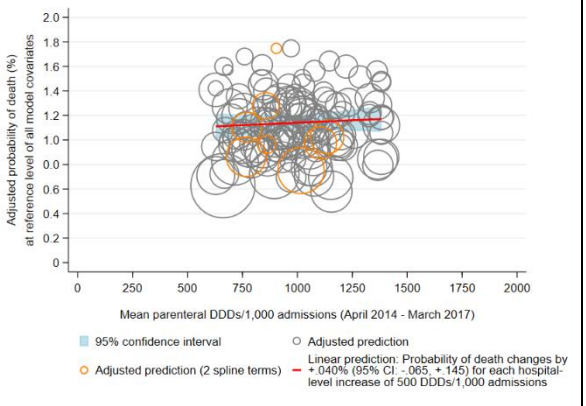
d2) Oral DDDs/1,000 admissions



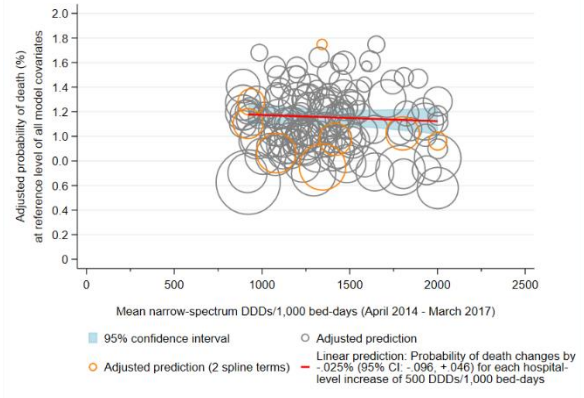
e1) Parenteral DDDs/1,000 bed-days



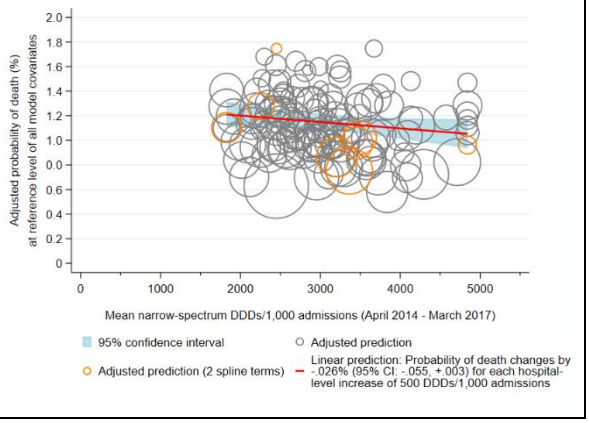
e2) Parenteral DDDs/1,000 admissions



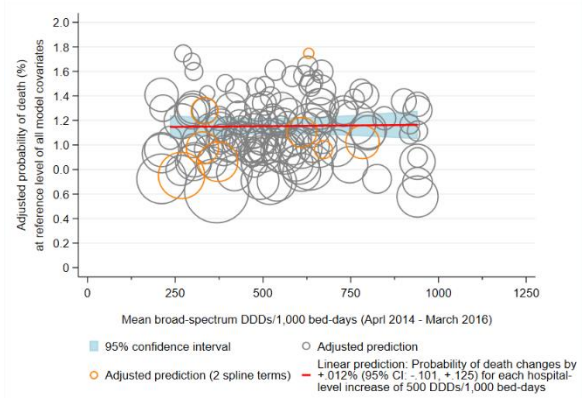
f1) Narrow-spectrum DDDs/1,000 bed-days



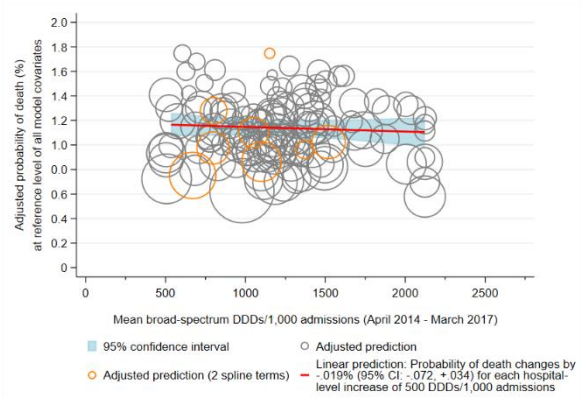
f2) Narrow-spectrum DDDs/1,000 admissions



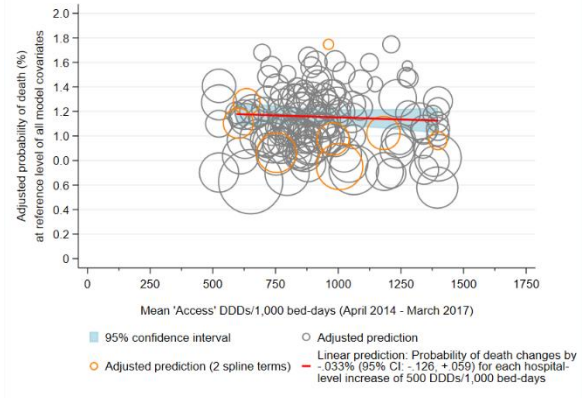
g1) Broad-spectrum DDDs/1,000 bed-days



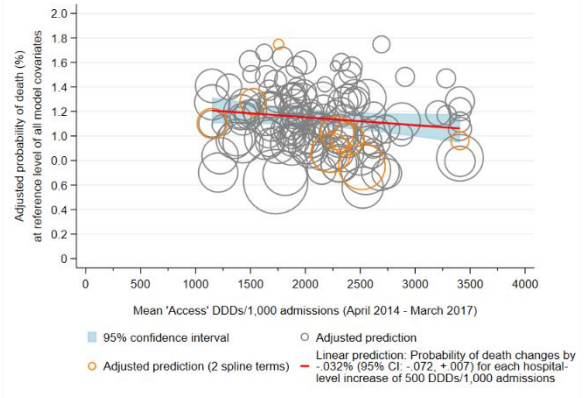
g2) Broad-spectrum DDDs/1,000 admissions



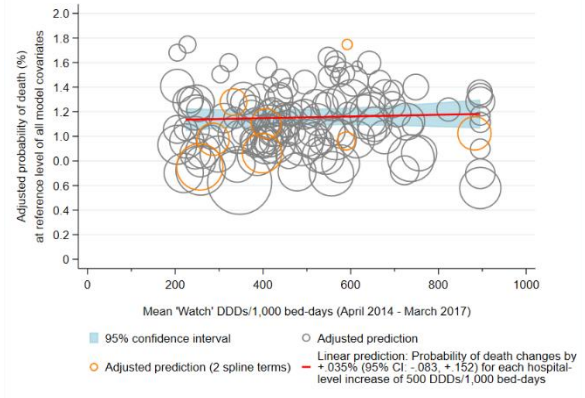
h1) Access DDDs/1,000 bed-days



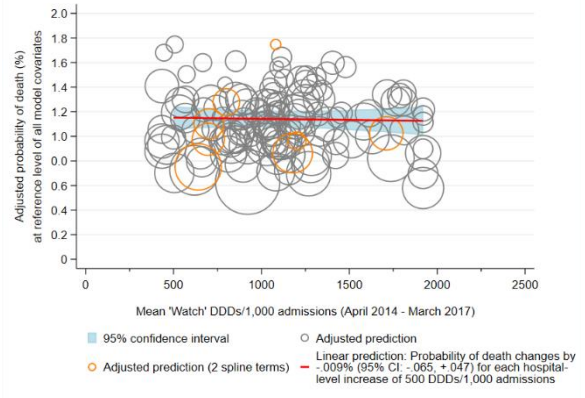
h2) Access DDDs/1,000 admissions



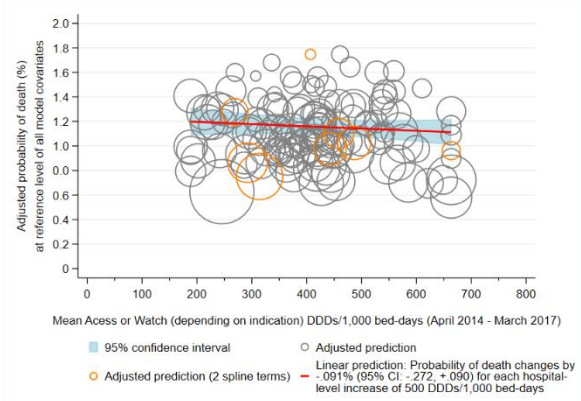
i1) Watch DDDs/1,000 bed-days



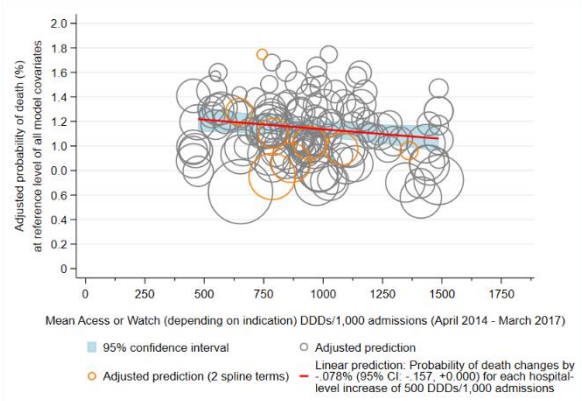
i2) Watch DDDs/1,000 admissions



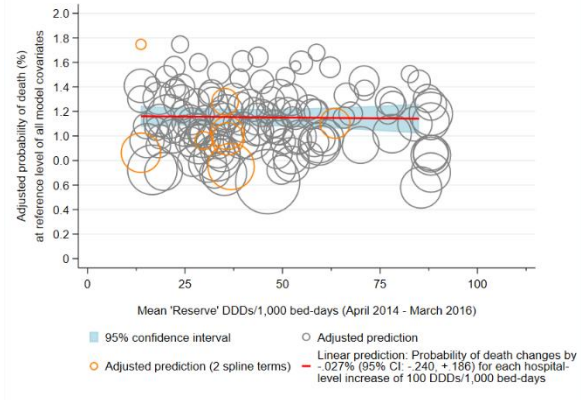
j1) Access or Watch (depending on indication) DDDs/1,000 bed-days



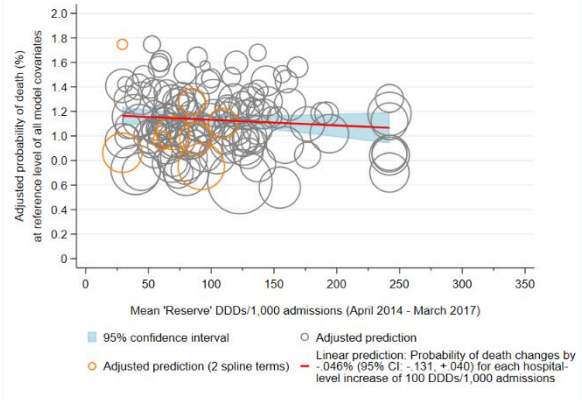
j2) Access or Watch (depending on indication) DDDs/1,000 admissions



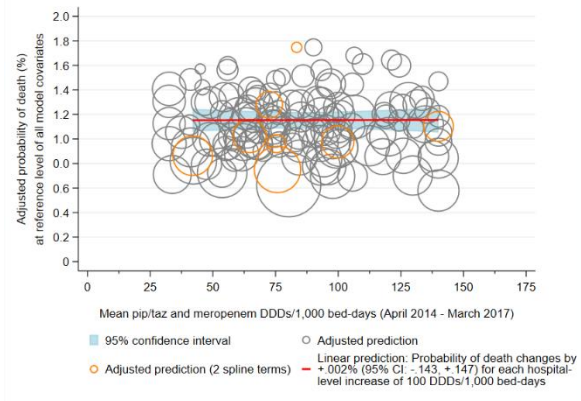
k1) Reserve DDDs/1,000 bed-days



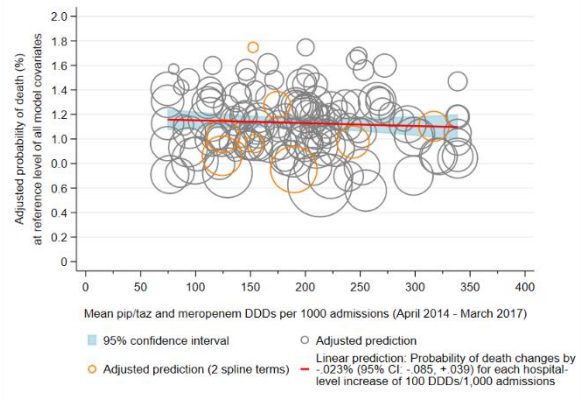
k2) Reserve DDDs/1,000 admissions



l1) Piperacillin/tazobactam and meropenem DDDs/1,000 bed-days



l2) Piperacillin/tazobactam and meropenem DDDs/1,000 admissions



[1] Marginal effects were derived from 135 separate (hospital-specific) models, represented here as circles sized according to the precision of the estimate (inverse of the within-hospital variance). Most (127/135) hospital-specific multivariate models (grey circles) included four spline terms (i.e. for age, Charlson score, IMD score, and overnight admissions in the past year), however models for eight hospital Trusts (orange circles) would only converge with two spline terms (i.e. age and overnight admissions in the past year). Antibiotic use was truncated below the 2.5th percentile and above the 95th percentile. Probability estimates were truncated above the 99th percentile.

Chapter 3: Emulating a sequentially randomised target trial to estimate the effect of stopping antibiotic treatment using electronic health records from University Hospitals Birmingham NHS Foundation Trust

3.1 Introduction

Antibiotic prescribers need to balance the risks of under-treating infections against the risks of future antibiotic resistance. However, doing so is complicated by widespread uncertainty regarding the optimal treatment duration for many bacterial infections(61), as illustrated by the wide variation in recommended antibiotic prescribing duration for key indications in acute NHS Trusts(60). Although the "optimal" course length will depend on many host- and pathogen-specific factors, including the priority attached to different short- and long-term considerations such as treatment outcomes, side effects (e.g. toxicity and other adverse events), and the risks of prolonged or excessively broad-spectrum use (e.g. resistance selection), a growing number of studies have demonstrated the safety and effectiveness of shorter antibiotic courses as one means of reducing antibiotic overuse(21,148,151,153,162–164). For example, over 120 RCTs have found shorter antibiotic courses to be non-inferior to longer antibiotic courses across a range of common bacterial infections with few exceptions(165), including for the treatment of:

- mild or moderate community-acquired pneumonia in children(166) and adults(150,167–169) (recommended 3 days treatment if clinically stable and have source control for infection), and ventilator-associated pneumonia (minimum 7-8 days treatment)(170)
- complicated and postoperative intra-abdominal infections (recommended 4-7 days treatment after source control is achieved)(149,171,172)
- skin and soft tissue infections such as uncomplicated cellulitis (recommended 5 days treatment if clinically stable and excluding deep-seated infections)(173)
- osteomyelitis (recommended 3-6 weeks for diabetic foot osteomyelitis, or 6 weeks for

vertebral osteomyelitis)(174–177)

- complicated urinary tract infections (recommended 7 days treatment for male cystitis or male/female pyelonephritis)(178–180)
- uncomplicated gram-negative bacteraemia (recommended 7 days treatment in patients with source control for infection)(181–183)
- uncomplicated acute bacterial sinusitis (recommended 3-7 days treatment)(184)

Despite this growing evidence, prescription length often exceeds guideline recommendations in both primary(185,186) and secondary care(57,187,188). In Chapter 2, I assessed whether there was any evidence to substantiate the key prescriber concern that substantial reductions in hospital-level antibiotic use may lead to clinical harm. In this chapter, I build on this work by instead using patient-level antibiotic and clinical outcome data to investigate the causal effect of using shorter vs longer antibiotic courses on the grounds that such an analysis could provide early evidence of harm, were this to be associated with the rollout of an antimicrobial stewardship initiative designed to reduce treatment duration across antibiotic classes, such as the ARK-Hospital trial(105). The range of possible inferences that might be drawn from this observational analysis are illustrated in **Table 3.1**. As shown, finding no evidence of harm (i.e. not being in the red cell) is necessary, though not sufficient on its own (gray cells), to conclude no harm from shorter antibiotic courses.

Although observational analyses cannot demonstrate non-inferiority of clinical outcomes due to the potential for residual confounding by indication, finding no evidence of harm may nevertheless provide reassuring evidence to practitioners and the public that shorter antibiotic treatment could be safe, and by extension, that strategies to reduce treatment duration are worth adopting and evaluating in RCTs.

Table 3.1 Possible inferences from an observational analysis of antibiotic treatment duration

	Unbiased inference regarding the impact of reduced antibiotic treatment duration, as might be derived from an ideal RCT	
Observational analysis of antibiotic treatment duration	Reduced treatment duration is non-inferior with respect to mortality	Reduced treatment duration is inferior with respect to mortality (causes harm)
If residual confounding by indication is considered unlikely (i.e. target trial protocol is adequately emulated, providing approximately unbiased inference)	Would expect to find reduced treatment duration is not associated with clinical outcome	Would expect to find reduced treatment duration is associated with clinical **harm**
If residual confounding by indication is considered likely (i.e. bias from more acutely unwell patients receiving longer antibiotic courses and having worse outcomes)	Would expect to find reduced treatment duration is associated with clinical **benefit**	Would expect to find reduced treatment duration is not associated with clinical outcome

3.2 Methods

3.2.1 Study population

This observational analysis used routinely collected electronic health records from adult (≥16 years) general medical inpatients admitted to the University Hospitals Birmingham (UHB) NHS Foundation Trust in England, between 01/January/2010-31/December/2015 inclusive (timeframe determined by the funding period of the ARK trial grant, analyses originally conducted in parallel with intervention development). All patients were cared for by consultants contracted to general medicine (HES field mainspef=300) during at least one episode in their admission spell. Since patients in general medicine may be treated under different adult specialties, and coding practices vary by hospital(130), this definition is narrower than that used in Chapter 2 (**Figure 2.1**) and therefore is likely to have very high specificity but relatively lower sensitivity (e.g. it may exclude some patients cared for by consultants who worked in general medicine (HES field tretspef=300) but were not contracted to general medicine). Whilst in theory, I could have assessed this by comparing the two

populations (i.e. admissions to UHB in HES data from Chapter 2 vs the cohort from this chapter), I was required by NHS Digital to delete the HES dataset used in Chapter 2 as part of the conditions of its release before realising this comparison could be useful.

3.2.2 Data source

The extracted dataset from UHB contained patient information that is routinely sent to the national warehouse held by NHS Digital, including:

- Date and time of hospital admission, discharge, and each episode start and end
- Demographics, including sex, age, ethnicity, multiple deprivation percentile (based on the postal code where the patient resided at admission, but postal codes were not collected)
- Admission method and source
- Discharge destination and method
- Diagnoses (primary and secondary ICD-10 codes) in each episode
- Specialty under which attending consultants worked or were contracted in each episode
- Date and time associated with intensive care unit (ICU) admissions (if applicable)

Patient and spell identifiers were pseudonymised by the local informatics department at UHB and linked with date of death (in or out of hospital) from the national death registry locally before pseudonymisation. The informatics team also calculated:

- Measures of past hospital exposure, reflecting overnight admissions to the same Trust
- Days to next non-elective readmission after discharge (to the same Trust, in any specialty)
- Dates of any positive MRSA specimens or screening swabs (before and/or after admission)

The following patient-level time-updated datasets were also provided by the informatics department:

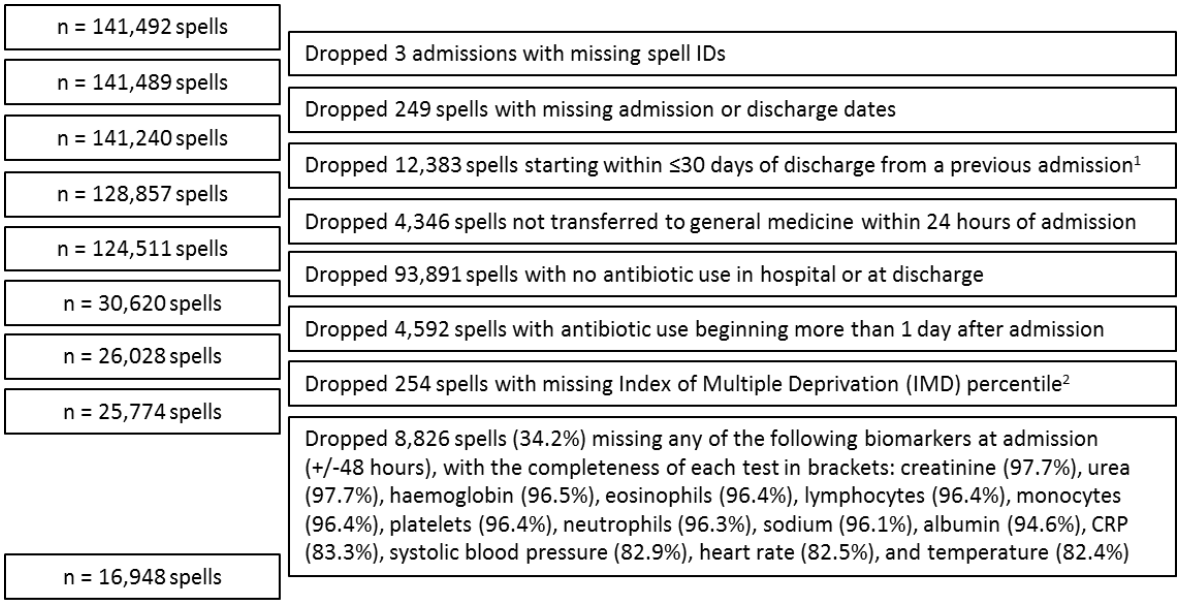
- Antibiotic dispensing in hospital, including drug, dose, route of administration, prescriber type, and the date and time of each administered dose
- Antibiotic prescriptions given at discharge, including drug, dose, route of administration, prescriber type, prescription start and end dates, and doses per day
- Haematology tests, including haemoglobin (Hb), neutrophils, eosinophils, lymphocytes, monocytes, and platelets
- Blood biochemistry tests, including CRP, albumin, alanine transaminase (ALT) (not routinely reported before October 2011), urea, sodium, and creatinine (since body weight was unavailable, measures of creatinine clearance or glomerular filtration could not be estimated)
- Physiological investigations (not routinely reported before March 2011), including heart rate, temperature, systolic and diastolic blood pressure
- Positive blood and urine cultures, including specimen collection date/time, organism(s) isolated, and minimum inhibitory concentration (MIC) values from Vitek 2. However, clinical breakpoints and drug susceptibility interpretations were not provided, nor were negative cultures.
- *Clostridioides difficile* test results, including specimen collection date and time, test type (glutamase dehydrogenate (GDH) screen, toxin polymerase chain reaction (PCR), or toxin enzyme immunoassay), and test result (positive, negative, equivocal). Prior to 01/April/2012, specimens were screened by GDH and positive screens were tested by toxin enzyme immunoassay. From 01/April/2012, the testing algorithm changed and positive GDH screens were followed by PCR, and positive PCR isolates were followed by a toxin enzyme immunoassay. Negative GDH screens were not reported before 01/April/2012.

With the exception of haemoglobin, units for blood biochemistry and haematology test results were not included in the dataset and were inferred based on reference ranges supplied by the hospital.

3.2.3 Cleaning spells data

Before the data was received, the hospital informatics department removed 513 admissions from patients who opted out of having their health records shared for research purposes, or who were on a restricted list due to a potentially stigmatising diagnosis. After their removal, the transferred data contained 141,492 adult (≥16 years) admissions from 80,264 patients. Cleaning steps used to derive the analytic cohort are described in **Figure 3.1**. These steps ensured all included patients were admitted or transferred to general medicine within 24 hours of admission, initiated antibiotics within 1 calendar day of admission, and had select haematology, biochemistry, physiological investigations recorded at baseline (admission +/-48h), many of which may reflect underlying infection.

Figure 3.1 Cleaning steps applied to admissions data



[1] Admissions beginning within 30 of discharge from a previous spell were excluded because I initially planned to examine non-elective readmission within 30 days of discharge as a secondary endpoint, and reasoned that spells beginning within this 30-day window were likely to share the same underlying indication for admission and treatment as the previous (included) spell.

[2] Missing deprivation index was first imputed from the nearest admission if available (n=15 spells).

Vital status was unknown in 0.2% of spells (n=38) (i.e. the NHS number was not found during death registry linkage); 7/38 were recorded as discharged dead and were thus treated as having died on

the day of discharge, while the remaining 31/38 were censored as alive at the time of discharge from their final spell. One patient was verified as alive through linkage to the death registry but was recorded as having been discharged dead; since there was no evidence of subsequent admissions or laboratory results, the hospital data was treated as correct and the patient was considered dead at discharge. Where date of death in the death registry and the hospital data differed, the hospital data was treated as correct (n=122 spells); in all but 2 spells these dates differed by ≤4 days. Since the last admission began on 31/December/2015 and death status was known through 09/June/2016, death was considered completely ascertained in analyses of 30-day mortality.

3.2.4 Cleaning antibiotic data

I restricted analyses to systemic antibiotics (i.e. where ATC code begins with "J"), dropping topical agents (e.g. eye and ear drops), antifungals (e.g. amphotericin B, fluconazole, fungizone), antivirals (e.g. acyclovir, ganciclovir), and antiprotozoal medications (e.g. pyrimethamine, mepacrine). Antibiotics used in the suspected treatment of tuberculosis were also dropped (e.g. pyrazinamide, isoniazid, streptomycin, and rifampicin if combined with other antituberculosis agents or if tuberculosis was flagged among the patient's primary or secondary diagnosis codes). Additional cleaning steps are described in the appendix.

3.2.5 Statistical methods

I assessed the causal effect of stopping vs continuing antibiotics on 30-day (all-cause) mortality by treating time-varying antibiotic use as sequence of non-randomised nested "trials", using an approach proposed by Miguel Hernán *et al*(104). Starting on the day of antibiotic initiation (which was within 1 calendar day of admission), I estimated the effect of stopping antibiotics (versus continuing) on that day on death within 30 days of the "trial" starting, adjusting for baseline admission characteristics and biomarkers (further details below). Those who continued antibiotics were cloned, and included in a new pseudo-trial starting the following day, and their exposure status

(stopped vs continued on day 2) and outcome (death within 30 days of the “trial” starting, i.e. from day 2 to day 31) were reassigned. In contrast, those who stopped antibiotics on day 1 of treatment were excluded from the subsequent pseudo-trials (i.e. excluded from the “trial” starting on day 2 of treatment, and all subsequent nested “trials”). The eligibility criteria for each “trial” was therefore:

- Age ≥16y when admitted to acute/general medicine within 24h of hospital admission
- Started systemic antibiotics within 1 calendar day of admission
- Still on systemic antibiotics today

In an ideal sequentially randomised trial, I would have randomised each patient to stop vs continue antibiotics each day after treatment initiation, where “stopping antibiotics” means that the last dose of antibiotics is taken today (the day the trial starts) and no antibiotics are taken tomorrow (see definitions below). In the pseudo-trials, I used the observational data, but adjusted for characteristics up to and including each "trial” baseline, in order to try to adjust for time-dependent confounding in the decision to continue or stop antibiotics. By comparing mortality risk in patients who stopped vs continued, regardless of whether they subsequently stopped or restarted, the estimated causal contrast was analogous to an intention-to-treat (ITT) effect in a randomised controlled trial.

This procedure was continued up to 10 days after starting antibiotics, thus yielding 10 separate “trials”. A separate record was created for each day of antibiotic treatment in hospital or after discharge (as post-discharge antibiotics were included in the dataset) up to this time point, with each record reflecting a separate pseudo-trial. There were two reasons I restricted to only the first 10 “trials”. First, among those who stopped antibiotic treatment (as opposed to dying or being censored), 80% (n=13,118/16,404) did so within ≤10 days of treatment initiation and those continuing treatment beyond 10 days may therefore be less representative of those who started. Second, the “trials” became progressively smaller over time as patients stopped antibiotics and were excluded from subsequent trials and this greatly reduced power and resulted in covariates with

levels with no deaths in multivariate models after the 10th "trial".

In the primary analysis, I first estimated the causal effect of stopping inpatient antibiotics specifically. Patients were classified as having stopped inpatient treatment on day *n* if there were no administered doses on day *n+1*, or if the patient was discharged on day *n*, as it cannot be known whether a prescription given at discharge was taken as prescribed. That is, there was an additional trial eligibility criteria, namely still in hospital today (start of the "trial"), and those previously discharged were not eligible or included in trials starting post-discharge. If antibiotics were administered on the day of death, the patient was classified as having continued treatment, as it was unclear if continued antibiotic treatment was planned. This definition of stopping antibiotics is henceforth described as definition #1, and is illustrated in **Table 3.2**.

Next, I estimated the causal effect of stopping antibiotics irrespective of whether the antibiotics were administered in hospital or after discharge, assuming antibiotics given at discharge were taken as prescribed up to the earliest of prescription end or death on treatment. Thus, if a patient received 5 continuous days of antibiotics in hospital and was discharged with an antibiotic prescription for 5 more days without a ≥1-day break between inpatient and outpatient treatment, then I created 10 records in the dataset (one per day of treatment) and the causal effect of stopping vs continuing was estimated for each of these 10 days, with the first 9 "trials" having an exposure status indicating that the patient continued treatment and the 10th trial indicating that antibiotics were stopped. This definition of stopping is described as definition #2 (also illustrated in **Table 3.2**).

In each analysis, the crude and adjusted relative hazard of stopping antibiotics versus not stopping was estimated using Cox proportional hazards models. Person-time began at the start of each "trial" and continued until the earliest of death or censoring after 30 days of follow-up (assumed 0.5 days of follow-up if the subject died or was censored on the day of a "trial"). A target trial protocol and summary of how it was emulated in the primary analysis is summarised in **Table 3.3**.

Table 3.2 First 10 "trials" in a patient who was treated with antibiotics for 5 consecutive days in hospital, then discharged on antibiotics for another 5 days (assume taken as prescribed), then died 33 days after initiating antibiotic treatment

Trial	Discharged	Administered antibiotics at trial baseline	Stopped antibiotics (no lag), definition 1	Stopped antibiotics (1 day lag), definition 1	Stopped antibiotics (2 day lag), definition 1	Stopped antibiotics (no lag), definition 2	Stopped antibiotics (1 day lag), definition 2	Stopped antibiotics (2 day lag), definition 2	Follow-up ends
1	No	Yes, in hospital	No			No			Censored as alive 30 days after trial start, on day 31
2	No	Yes, in hospital	No	No		No	No		Censored as alive 30 days after trial start, on day 32
3	No	Yes, in hospital	No	No	No	No	No	No	Died 30 days after trial start, on day 33
4	No	Yes, in hospital	No	No	No	No	No	No	Died 29 days after trial start, on day 33
5	Yes	Yes, in hospital	**Yes**	No	No	No	No	No	Died 28 days after trial start, on day 33
6		Yes, after discharge		**Yes**	No	No	No	No	Died 27 days after trial start, on day 33
7		Yes, after discharge			**Yes**	No	No	No	Died 26 days after trial start, on day 33
8		Yes, after discharge				No	No	No	Died 25 days after trial start, on day 33
9		Yes, after discharge				No	No	No	Died 24 days after trial start, on day 33
10		Yes, after discharge				**Yes**	No	No	Died 23 days after trial start, on day 33

Table 3.3 Emulation of a sequentially randomised target trial for stopping definition 1

Target trial protocol	Target trial	Emulation (primary analysis)
Eligibility criteria	• Age ≥16y when admitted to general medicine within <24h of hospital admission • Started systemic antibiotics within ≤1 calendar day of admission • Still on systemic antibiotics today • Not yet discharged[1]	Same, identified from routinely collected electronic health records and patient-level prescribing data
Treatment strategies	Randomly assign subjects to stop inpatient antibiotics today, or continue treatment today[1]	Subjects were classified as having "stopped" inpatient treatment if routine data recorded treatment being administered today but not tomorrow, or if discharged today
Causal contrast	ITT effect	Observational analogue of an ITT effect
Assignment procedure	Those who are assigned to continue treatment today are re-randomised the following day, and this is repeated from the 1st to the 10th consecutive day of treatment. Strategy assignment is unblinded.	Effect of stopping vs continuing treatment was modelled as a sequence of up to 10 non-randomised nested pseudo-trials. The sequential comparability (exchangeability) of patients was assumed conditional on "trial" baseline covariates.
Endpoint	Death within ≤30 days after strategy assignment	Same, with death date determined through death registry linkage (performed by UHB informatics team)
Follow-up Period	From strategy assignment to the earliest of death or 30 days after assignment	Aligned eligibility with treatment assignment and start of follow-up, which spanned from assignment to the earliest of death, censoring at discharge (if date of death was unknown), or 30 days after assignment
Analyses	Compute relative hazard of stopping vs continuing antibiotic treatment for each of first 10 days of treatment	Same – modelling approach summarised in **Table 3.4**

[1] Definition 2 differs in that there is no inclusion criteria relating to being in hospital today, and stopping/continuing antibiotics includes both antibiotics administered in hospital and those prescribed at discharge, assuming that those prescribed at discharge were taken as prescribed.

To derive causal effect estimates, all 10 pseudo-trials were first pooled a single Cox model, with a robust variance estimator to account for patients contributing time to multiple "trials" (affecting

standard errors, but not model coefficients). To allow for the effect of stopping antibiotics to vary with time since starting antibiotics, this pooled model included an interaction term between the primary exposure (stopped versus not stopped antibiotics) and trial day (fit as a 4 knot natural cubic spline). Since patients may stop antibiotics when death is considered to be imminent (leading to reverse causation between stopping antibiotics and death), each pooled model was modified by lagging the primary exposure (stopped vs continued) by 1 day, such that the effect of stopping antibiotics on day *n* of treatment was estimated using each subject's exposure status from day *n-1* (thus, no effect estimate was derived on the first day of treatment, and an extra "trial" was included to account for the lagged stop date, provided *n*<10 and the patient did not die on day *n*). A 2-day lag was also considered (illustrated in **Table 3.2**). Next, the pooled model was fit with trial day as a categorical predictor, with one category per day of treatment up to the earliest of death, stopping, censoring at discharge, or the 10th "trial". Finally, I estimated the causal effect of stopping antibiotics using 10 separate (unpooled) Cox models (one for each "trial" day, with no robust variance adjustment by patient), allowing each potential confounder to have a different impact on the outcome in each "trial" (**Table 3.4**).

Table 3.4 Modelling approach used to estimate the causal effect of stopping antibiotics on 30-day mortality, among general medical inpatients who initiated treatment within 1 day of admission

Model	Cox model	How was trial day adjusted for?	Was the primary exposure (stopped vs continued treatment) lagged?
1	Pooled	Natural cubic spline (4 Harrell knots[1])	No
2			Yes, lagged by 1 day
3			Yes, lagged by 2 days
4		Categorical	No
5			Yes, lagged by 1 day
6			Yes, lagged by 2 days
7	Unpooled	N/A	No
8			Yes, lagged by 1 day
9			Yes, lagged by 2 days

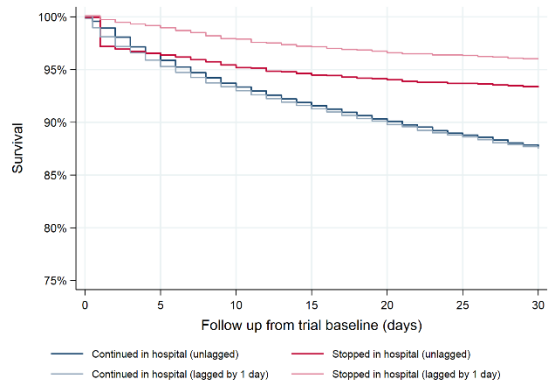
[1] Knot values based on Harrell's recommended percentiles (i.e. 5th, 35th, 65th, 95th)(161), generated using the Stata command "mkspline, cubic".

In addition to reverse causation from patients who stop active therapy due to imminent death, my emulation of the target trial may misclassify a treatment "stop" if antibiotics are administered today, but not tomorrow, due to a death occurring tomorrow before planned treatment could be administered. In this scenario, treatment was not withdrawn and the patient should have been classified as "continuing" treatment. Kaplan-Meier survival curves show this misclassification often results in patients dying exactly 1 day after "stopping", and this problem affects all trials (examples in **Figure 3.2**). As shown, the unlagged survival curves are substantially biased by this misclassification. However, since this problem is resolved by lagging the exposure variable (stopped vs continued), I have only presented adjusted effect estimates from the lagged models in the main results.

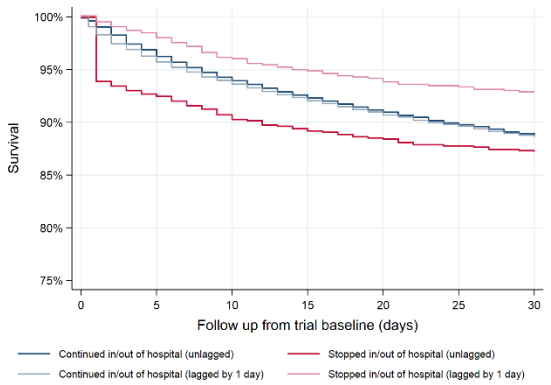
Of note, the misclassification of "stopped" vs "continued" treatment is more pronounced in analyses of stopping definition 2. This is because a relatively greater proportion of patients who "stop" antibiotics early on treatment using definition 2 (vs definition 1) remain admitted and are at high risk of death, while a relatively greater proportion of patients who "stop" antibiotics early on treatment using definition 1 (vs definition 2) are discharged alive and are at low risk of death; this is an artefact of how stopping has been defined, since any patient who is discharged on antibiotics without a 1-day gap in treatment will be classified as having stopped under definition 1, but not definition 2, as illustrated in **Table 3.2**.

Figure 3.2 Kaplan-Meier survival function showing the lagged and unlagged unadjusted effect of stopping antibiotics (vs continuing), with "stopped" defined as per definition 1 (A1-A3; inpatient antibiotics only) or definition 2 (B1-B3; inpatient and discharge antibiotics)

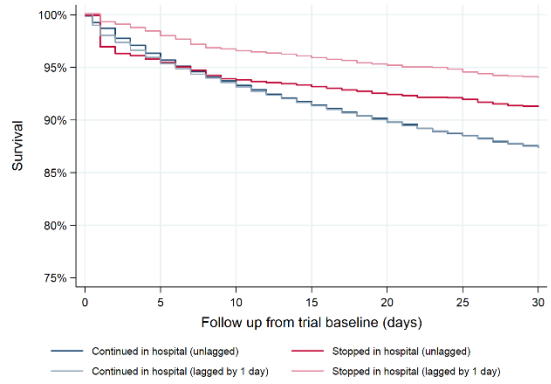
A1) Trial #1, stopping definition 1



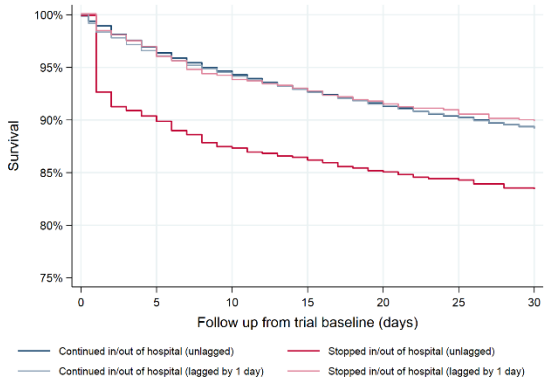
B1) Trial #1, stopping definition 2



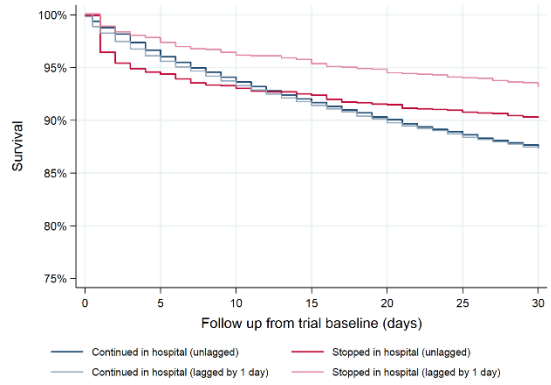
A2) Trial #2, stopping definition 1



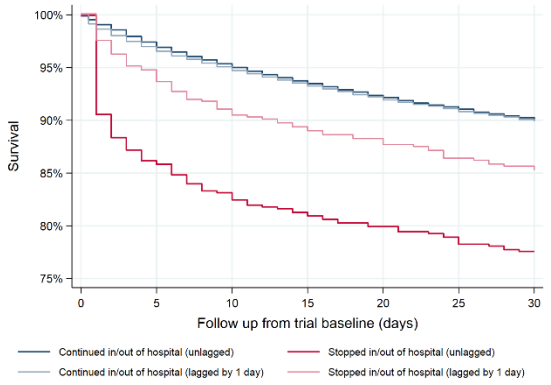
B2) Trial #2, stopping definition 2



A3) Trial #3, stopping definition 1



B3) Trial #3, stopping definition 2



3.2.6 Confounding adjustment – admission factors

Since I was primarily concerned with confounding adjustment, rather than the effect of individual covariates on mortality per se, I first adjusted each "trial" for the main exposure (stopped versus not stopped antibiotics) and 13 admission (i.e. constant across "trials") factors included in a previous analysis of mortality risk in NHS hospital admissions(86), irrespective of statistical significance, including:

- Sex (male, female)
- Ethnicity (white, Asian, black, other, unknown)
- Age at admission (continuous in integer years, centred at median)
- Multiple deprivation percentile (continuous on a 100 point scale, centred at median)
- Charlson co-morbidity index (Charlson score), estimated from the first episode in general medicine using secondary diagnosis codes and weights from NHS Digital(159) (continuous)
- Admission day of the week (weekend, weekday)
- Admission time of day (hour and minute), modelled using a sin() + cos() function (2df) to ensure a smooth transition in risk from hour to hour
- Admission day of year, modelled using a sin() + cos() function (2df) to ensure a smooth transition in risk from day to day
- Admission year (one category per year, from 2010-2015)
- Admission method (A&E, elective and other non-emergency, emergency via general practitioner (GP) or other source)
- Number of overnight admissions in the past year within the same Trust (continuous)
- Ever had a complex admission (≥2 consultant episodes) in the past year within the same Trust (yes, no)
- CCS group, identified from the first episode in general medicine and linked with 142 CCS groups identified by NHS Digital(189). To improve model stability, these categories were

collapsed into 43 subgroups following an approach described by Walker *et al*(86). CCS categories with low mortality (<1.5%) within 30 days of admission were then combined into a "low-risk" group, while the smallest categories (<130 spells each, cumulatively accounting for 5% of spells) were combined into two "other" groups based on whether 30-day mortality was above or below the overall 30-day mortality risk at admission. This process yielded 28 CCS groups, the largest of which (pneumonia) was used as the reference group in regression models.

Age, multiple deprivation percentile, Charlson score, and the number of overnight admissions in the past year were all fit using natural cubic splines to account for non-linearity as the use of splines (rather than fitting these factors as linear continuous predictors) improved the model fit (lowered AIC by >55) in a pooled Cox model adjusted only for the factor in question, the trial day, the primary exposure (stopped vs not stopped, using 1st definition above), and its interaction with trial day (model 1 in **Table 3.4**). Each factor was fit using 4 Harrell knots, except overnight admissions in the past year, where I instead placed 3 knots manually due to limited variation in the truncated distribution, at the 50th, 85th, and 95th percentiles of the non-zero distribution.

I also included other admission factors that were not considered in this previous analysis(86). Specifically, I adjusted for the following 4 admission factors irrespective of statistical significance (pre-specified in the Statistical Analysis Plan), although they also each individually lowered the AIC of a multivariate pooled Cox model (model 1 in **Table 3.4**, adjusted for the 13 factors above):

- Immunosuppression (yes, no), estimated from the first episode in general medicine using secondary diagnosis codes indicating metastatic cancer, severe liver disease, or HIV.
- Blood microbiological isolations nearest to admission (+/-48h), categorised as *Escherichia coli* (*E. coli*) or other lactose-fermenting colony (LFC) (e.g. *Klebsiella pneumoniae*), gram positive, other (includes mixed species), no result available, and "probable contaminants"—defined as bacteria typically belonging to the skin flora, as per criteria used elsewhere(190).

- Urine culture nearest admission (+/-48h), categorised as *E. coli* or other LFC, other, or no result available/"probable contaminant"—defined as isolations with >1 bacterial species
- Any positive MRSA specimen or swab >48 hours before admission (yes, no).

I considered adjusting for other potential confounders – namely admission source (usual place of residence, NHS general ward, other), *C. difficile* infection at admission (+/-48h), and *C. difficile* infection before admission. However, there was insufficient variability in these factors (i.e. the reference groups of these factors accounted for 99% of spells), so they were not included in adjusted models.

In model 1 (**Table 3.4**), fitting trial day using natural cubic splines with 4 Harrell knots produced a model with the lowest AIC value, compared with 3, or 5 knots, when adjusting for all 18 factors above (appendix **Table 3.7**). Thus, trial day was modelled using 4 Harrell knots in all models.

3.2.7 Confounding adjustment – interactions between admission factors

Eight interaction terms included in a previous analysis(86) were added individually to a multivariate pooled Cox model (model 1 in **Table 3.4**) adjusting for the primary exposure, trial day, and all 18 factors listed above, namely between:

1. Admission method and the number of overnight admissions in the last year
2. Admission method and calendar year
3. Charlson score and age
4. Charlson score and the number of prior admissions in the last year
5. Charlson score and any complex admission in the last year
6. Number of admissions in the last year and age
7. Number of admissions in the last year and any complex admission in the last year
8. Admission day of the week (weekend vs weekday) and admission hour of day

I planned to include interactions if significant ($p<0.05$) or if they changed the primary effect estimate by >5%; however, none of the 8 interactions met either criterion and therefore none were retained in the final adjusted model.

3.2.8 Confounding adjustment – time-updated factors reflecting illness severity

To adjust for potential residual confounding by illness severity, I also adjusted for 15 biomarkers, including 3 physiological tests (systolic blood pressure, temperature, and heart rate) and 12 commonly performed haematology and biochemistry investigations, including acute inflammatory and infection biomarkers (CRP, neutrophils, eosinophils, lymphocytes, monocytes) and tests for renal (sodium, urea, creatinine), hepatic (albumin, ALT), and haematologic dysfunction (haemoglobin, platelets). Since biomarker tests are often repeated after admission, up to three separate test results were available for adjustment in each "trial", including:

- Biomarker test result at admission
- Current ("trial" baseline) biomarker test result, with the value being carried forward from the most recent previous test result before each "trial"
- Change between the current and admission test result

To determine which biomarker test result(s) to adjust for, I fitted up to 6 separate pooled Cox models (model 1 in **Table 3.4**) per biomarker, with each model adjusting for the primary exposure, trial day, the 18 admission factors listed above, and one of the test result(s) below:

a) Biomarker test result at admission, or
b) Most recent biomarker test result (up to the "trial" day), or
c) Change between the 'current' and admission biomarker test result, or
d) Both result at admission and the current measure, or
e) Both result at admission and the change in value, or
f) Both the current test result and the change in value

For each biomarker, the model (a-f) with the lowest AIC was selected as the best fitting model and the corresponding test result(s) were included in adjusted models irrespective of statistical significance (appendix **Table 3.8**). Since admission and current measures of ALT, haemoglobin, platelets and creatinine were all highly correlated (Spearman's rho ≥0.88, p<0.0001), test results at admission were not considered for these biomarkers (i.e. only models b, c, and f were considered). Most (13/15) biomarker test results were modelled using 4 knot natural cubic splines to account for non-linearity and visualised to assess over-fitting, with two exceptions:

- ALT was modelled as binary (ever vs never tested) as it was not routinely reported before October 2011 and therefore only recorded at admission in 81.3% (13,782/16,948) of spells. Models adjusting for changes in biomarker results (c, f) were therefore not considered for ALT.
- Eosinophil results were fit as categorical (0, 0.1, 0.2-0.3, 0.4-0.5x10^9/L) as there was limited variation in the truncated distribution. Therefore only models a, b, and d were considered.

To improve model stability, all continuous covariates were truncated at the 99th percentile (age, multiple deprivation percentile, number of overnight admissions in the past year, creatinine, neutrophils, CRP), the 1st and 99th percentile (systolic blood pressure, heart rate, temperature), the 95th percentile (Charlson score, urea, Hb, lymphocytes, monocytes, platelets, eosinophils), or the 5th and 95th percentile (sodium, albumin).

Lastly, since restricting the analysis to patients with non-missing infection biomarkers and physiology tests excluded one-third of spells (34.2%, n=8,826/25,774) (**Figure 3.1**), I compared the distribution of admission factors, treatment characteristics, and clinical outcomes among complete cases (n=16,948) vs those in the full dataset (25,774), to gauge if complete cases were broadly representative of all patients starting antibiotics at admission.

3.2.9 Confounding adjustment – time-updated factor reflecting discharge

In analyses using stopping definition 2, each "trial" also adjusted for a "trial baseline" covariate reflecting whether the patient had already been discharged on antibiotics (no, yes to another hospital, yes but not to another hospital). This factor was included in models as 30-day mortality risk might be expected to be lower among patients already discharged on antibiotics than among inpatients on antibiotics. Thus, if a patient was discharged after 5 days of inpatient antibiotic treatment and remained on antibiotics after discharge, this factor equalled "no" in trials 1-5 and "yes" from trial 6 to the earliest of trial 10 or end of the discharge prescription. This factor was not relevant when modelling the effect of stopping inpatient antibiotics specifically (definition 1), since these patients were only eligible for each trial if they had not yet been discharged (remained on antibiotics in hospital).

Analyses were conducted in Stata/MP 17, with all data stored on NHS servers.

3.3 Results

Among 124,511 admissions to general medicine, antibiotics were prescribed in 24.6% (n=30,620) of spells, of which most (85.0%, n=26,028/30,620) initiated treatment within 1 day of admission (i.e. 16,680 started on the day of admission, and 9,348 the following day). Those who did not start antibiotics within 1 day were dropped (n=4,592/30,620), as were patients missing IMD scores (n=254/26,028). Of the remaining patients, 34.2% (n=8,826/25,774) were missing at least one of the available biomarkers/physiology tests reflecting illness severity, leaving 16,948 complete cases, all of whom were included in analyses using stopping definition 2 (**Figure 3.1**). Among the complete cases, 2.8% (482/16,948) were not administered inpatient antibiotics within 1 day of admission. Instead, they were discharged on antibiotics within 1 day of admission, and therefore were excluded from analyses using stopping definition 1, which included 16,466 spells.

The distribution of baseline admission factors was broadly similar in complete cases (n=16,948) and those in the full dataset (n=25,774), with the exception of admission year, reflecting the fact that temperature, systolic blood pressure, and heart rate were not routinely reported before March 2011 so more early admissions were excluded (**Table 3.5**). Among complete cases, median age at admission was 73 years (IQR: 56-83) and pneumonia was the most common CCS primary diagnosis group (19.7%, n=3,340). A positive blood culture was recorded at admission (+/-48h) in 4.0% (n=683) of spells (excluding probable contaminants), while 6.0% (n=1,021) had a positive urine culture at admission.

As shown in **Table 3.6**, antibiotic treatment characteristics and clinical outcomes were also broadly similar between complete cases and the full dataset. For example, days to first antibiotic stop was similar (median 5 days using the 1st definition of "stopping", and median 7 days using the 2nd definition of "stopping" in both), as was death within 30 days of starting antibiotics (11.3% vs 10.5%). Rates of re-starting inpatient antibiotic treatment within 30 days of stopping were also similar, whether I considered antibiotics re-started in hospital or at discharge (60.3% vs 59.3%, 1st definition of "stopping") or in hospital only (15.9% vs 15.4%, 1st definition). However, median length of stay was slightly longer in complete cases (5.9 days vs 4.9 days overall) and a slightly greater proportion of complete cases:

- Initiated broad-spectrum (56.9% vs 54.1%) and IV antibiotics (58.5% vs 53.0%)
- Had drug modifications to their initial prescription (45.1% vs 39.9%)
- Escalated from narrow- to broad-spectrum (17.1% vs 14.0%) or oral to IV (12.7% vs 10.3%)

Table 3.5 Baseline characteristics of general medical inpatients who initiated antibiotics within 1 day of admission, University Hospitals Birmingham NHS Foundation Trust, 2010-2015

Characteristic		Admissions (full dataset)	Admissions (complete cases)	Among complete cases: died ≤30 days of starting antibiotics (first "trial")	
		N=25,774 (col %)	N=16,948 (col %)	No N=15,029 (col %)	Yes N=1,919 (col %)
Age (years)	Median [IQR]	71 [54-82]	73 [56-83]	71 [54-82]	81 [71-88]
Sex	Female[1]	13,887 (53.9%)	8,887 (52.4%)	7,960 (53.0%)	927 (48.3%)
	Male	11,887 (46.1%)	8,061 (47.6%)	7,069 (47.0%)	992 (51.7%)
Ethnic category	White[1]	21,252 (82.5%)	14,041 (82.9%)	12,356 (82.2%)	1,685 (87.8%)
	Asian	2,598 (10.1%)	1,781 (10.5%)	1,623 (10.8%)	158 (8.2%)
	Black	728 (2.8%)	464 (2.7%)	433 (2.9%)	31 (1.6%)
	Other	621 (2.4%)	380 (2.2%)	356 (2.4%)	24 (1.3%)
	Unknown	575 (2.2%)	282 (1.7%)	261 (1.7%)	21 (1.1%)
Deprivation percentile (0 is most deprived)	Median [IQR]	31.7 [19.9-49.7]	32.1 [20.2-49.7]	32.3 [20.3-50.0]	28.9 [19.7-47.2]
Immunosuppression	No[1]	22,563 (87.5%)	14,509 (85.6%)	13,152 (87.5%)	1,357 (70.7%)
	Yes	3,211 (12.5%)	2,439 (14.4%)	1,877 (12.5%)	562 (29.3%)
Charlson score	Median [IQR]	4 [0-13]	4 [0-14]	4 [0-13]	14 [4-18]
Admission day of week	Weekday[1]	19,563 (75.9%)	12,800 (75.5%)	11,355 (75.6%)	1,445 (75.3%)
	Weekend (Saturday or Sunday)	6,211 (24.1%)	4,148 (24.5%)	3,674 (24.4%)	474 (24.7%)
Admission year	2010[2]	3,962 (15.4%)	741 (4.4%)	629 (4.2%)	112 (5.8%)
	2011	3,893 (15.1%)	2,253 (13.3%)	1,997 (13.3%)	256 (13.3%)
	2012	4,974 (19.3%)	3,694 (21.8%)	3,265 (21.7%)	429 (22.4%)
	2013	5,030 (19.5%)	3,954 (23.3%)	3,522 (23.4%)	432 (22.5%)
	2014	4,595 (17.8%)	3,642(21.5%)	3,267 (21.7%)	375 (19.5%)
	2015[1]	3,320 (12.9%)	2,664 (15.7%)	2,349 (15.6%)	315 (16.4%)

Characteristic		Admissions (full dataset)	Admissions (complete cases)	Among complete cases: died ≤30 days of starting antibiotics (first "trial")	
				No	Yes
		N=25,774 (col %)	N=16,948 (col %)	N=15,029 (col %)	N=1,919 (col %)
Admission day of year	January to March	7,175 (27.8%)	4,397 (25.9%)	3,848 (25.6%)	549 (28.6%)
	April to June	6,138 (23.8%)	4,100 (24.2%)	3,676 (24.5%)	424 (22.1%)
	July to September	5,763 (22.4%)	3,814 (22.5%)	3,401 (22.6%)	413 (21.5%)
	October to December	6,698 (26.0%)	4,637 (27.4%)	4,104 (27.3%)	533 (27.8%)
Admission hour of day	Median [IQR]	3PM [9AM-7PM]	3PM [9AM-7PM]	3PM [9AM-7PM]	4PM [10AM-7PM]
Admission method	A&E[1]	18,817 (73.0%)	12,829 (75.7%)	11,291 (75.1%)	1,538 (80.1%)
	Elective & Other non-emergency	1,061 (4.1%)	265 (1.6%)	250 (1.7%)	15 (0.78%)
	Emergency via GP or other source	5,896 (22.9%)	3,854 (22.7%)	3,488 (23.2%)	366 (19.1%)
Admission source (not in adjusted models)	Usual place of residence	25,449 (98.7%)	16,808 (99.2%)	14,906 (99.2%)	1,902 (99.1%)
	NHS general ward	219 (0.9%)	70 (0.4%)	62 (0.4%)	8 (0.4%)
	Other	106 (0.4%)	70 (0.4%)	61 (0.4%)	9 (0.5%)
Number of overnight admissions in past year	Median [IQR]	0 [0-2]	0 [0-2]	0 [0-2]	1 [0-2]
Any complex overnight admission in past year (>1 consultant episode)	No[1]	18,976 (73.6%)	12,147 (71.7%)	10,885 (72.4%)	1,262 (65.8%)
	Yes	6,798 (26.4%)	4,801 (28.3%)	4,144 (27.6%)	657 (34.2%)
Blood culture at admission (nearest +/-48 hours)	Not done or negative[1,3]	24,565 (95.3%)	16,034 (94.6%)	14,249 (94.8%)	1,785 (93.0%)
	E. coli or Other LFC	426 (1.7%)	335 (2.0%)	289 (1.9%)	46 (2.4%)
	Gram positive	313 (1.2%)	239 (1.4%)	201 (1.3%)	38 (2.0%)
	Probable contaminant[4]	326 (1.3%)	231 (1.4%)	198 (1.3%)	33 (1.7%)
	Other	144 (0.6%)	109 (0.6%)	92 (0.6%)	17 (0.9%)
Urine culture at admission (nearest +/-48 hours)	Not done or negative or probable contaminant[3]	24,310 (94.3%)	15,927 (94.0%)	14,088 (93.7%)	1,839 (95.8%)
	E. coli or other LFC	1,215 (4.7%)	834 (4.9%)	774 (5.2%)	60 (3.1%)
	Other	249 (1.0%)	187 (1.1%)	167 (1.1%)	20 (1.0%)

Characteristic		Admissions (full dataset)	Admissions (complete cases)	Among complete cases: died ≤30 days of starting antibiotics (first "trial")	
		N=25,774 (col %)	N=16,948 (col %)	No N=15,029 (col %)	Yes N=1,919 (col %)
***C difficile* infection at admission (+/-48 hours) (not in adjusted models)**	No	25,729 (99.8%)	16,919 (99.8%)	15,009 (99.9%)	1,910 (99.5%)
	Yes	45 (0.2%)	29 (0.2%)	20 (0.1%)	9 (0.5%)
***C difficile* infection >48h before admission (not in adjusted models)**	No	201,984 (98.9%)	16,735 (98.7%)	14,846 (98.8%)	1,889 (98.4%)
	Yes, >1 year ago	1,679 (0.8%)	159 (0.9%)	141 (0.9%)	18 (0.9%)
	Yes, in past year	609 (0.3%)	54 (0.3%)	42 (0.3%)	12 (0.6%)
Any positive MRSA specimen or swab >48h before admission	No[1]	23,637 (91.7%)	15,449 (91.2%)	13,746 (91.5%)	1,703 (88.7%)
	Yes	2,137 (8.3%)	1,499 (8.8%)	1,283 (8.5%)	216 (11.3%)
Clinical Classifications Software (CCS) diagnosis group (abbreviated)	Pneumonia (excluding tuberculosis)[1]	4,355 (16.9%)	3,340 (19.7%)	2,672 (17.8%)	668 (34.8%)
	Chronic obstructive pulmonary disease	2,523 (9.8%)	1,702 (10.0%)	1,588 (10.6%)	114 (5.9%)
	Urinary tract infections	2,462 (9.6%)	1,692 (10.0%)	1,599 (10.6%)	93 (4.8%)
	Acute bronchitis, asthma	2,262 (8.8%)	1,485 (8.8%)	1,407 (9.4%)	78 (4.1%)
	Skin and subcutaneous tissue infections	1,692 (6.6%)	1,113 (6.6%)	1,077 (7.2%)	36 (1.9%)
	Liver diseases, biliary tract disease	1,491 (5.8%)	700 (4.1%)	622 (4.1%)	78 (4.1%)
	Nonspecific chest pain, cardiac dysrhythmias	1,146 (4.5%)	616 (3.6%)	556 (3.7%)	60 (3.1%)
	Headache / other nervous system disorders	1,000 (3.9%)	594 (3.5%)	556 (3.7%)	38 (2.0%)
	Low risk (<1.5% 30-day mortality at admission, in complete cases)	853 (3.3%)	547 (3.2%)	541 (3.6%)	6 (0.3%)
	Other, 30-day mortality risk is < overall risk at admission (in complete cases)	856 (3.3%)	502 (3.0%)	472 (3.1%)	30 (1.6%)
	Other connective tissue disease, osteoarthritis	622 (2.4%)	430 (2.5%)	397 (2.6%)	33 (1.7%)
	Mental disorders/retardation, senility	676 (2.6%)	412 (2.4%)	368 (2.4%)	44 (2.3%)

Characteristic		Admissions (full dataset)	Admissions (complete cases)	Among complete cases: died ≤30 days of starting antibiotics (first "trial")	
		N=25,774 (col %)	N=16,948 (col %)	No N=15,029 (col %)	Yes N=1,919 (col %)
	Other, 30-day mortality risk is ≥ overall risk at admission (in complete cases)	631 (2.5%)	377 (2.2%)	297 (2.0%)	80 (4.2%)
	Regional enteritis and ulcerative colitis	566 (2.2%)	328 (1.9%)	305 (2.0%)	23 (1.2%)
	Influenza, other upper respiratory infections	498 (1.9%)	323 (1.9%)	296 (2.0%)	27 (1.4%)
	Cancer related	527 (2.0%)	322 (1.9%)	188 (1.3%)	134 (7.0%)
	Congestive heart failure, non-hypertensive	417 (1.6%)	315 (1.9%)	248 (1.7%)	67 (3.5%)
	Superficial or internal injuries, fractures, burns	507 (2.0%)	305 (1.8%)	272 (1.8%)	33 (1.7%)
	Septicaemia (except in labour), shock	356 (1.4%)	274 (1.6%)	185 (1.2%)	89 (4.6%)
	Diabetes mellitus	291 (1.1%)	203 (1.2%)	197 (1.3%)	6 (0.3%)
	Intestinal infection	246 (1.0%)	202 (1.2%)	186 (1.2%)	16 (0.8%)
	Acute and unspecified renal failure	221 (0.9%)	190 (1.1%)	139 (0.9%)	51 (2.7%)
	Nausea/vomiting, fluid/electrolyte disorders	286 (1.1%)	185 (1.1%)	162 (1.1%)	23 (1.2%)
	Pleurisy, pneumothorax, pulmonary collapse	236 (0.9%)	174 (1.0%)	143 (1.0%)	31 (1.6%)
	Esophageal disorders	268 (1.0%)	165 (1.0%)	136 (0.9%)	29 (1.5%)
	Complication of surgical procedures/implant	316 (1.2%)	159 (0.9%)	152 (1.0%)	7 (0.4%)
	Phlebitis, thrombophlebitis/thromboembolism	252 (1.0%)	154 (0.9%)	141 (0.9%)	13 (0.7%)
	Other gastrointestinal disorders	218 (0.9%)	139 (0.8%)	127 (0.8%)	12 (0.6%)
Haematology results at admission (+/-48 hours)[5]	Haemaglobin (g/dL), median [IQR]	12.5 [11.1-13.8]	12.5 [11.0-13.7]	12.5 [11.1-13.8]	11.8 [10.3-13.2]
	Neutrophils (x10^9/L), median [IQR]	8.5 [5.7-12.3]	9.0 [6.2-12.8]	8.9 [6.1-12.6]	10.1 [7.0-14.4]
	Lymphocytes (x10^9/L), median [IQR]	1.2 [0.8-1.8]	1.1 [0.7-1.7]	1.2 [0.7-1.8]	0.9 [0.6-1.4]
	Monocytes (x10^9/L), median [IQR]	0.8 [0.5-1.1]	0.8 [0.6-1.1]	0.8 [0.6-1.1]	0.5 [0.8-1.1]
	Platelets (x10^9/L), median [IQR]	234 [178-301]	233 [176-301]	233 [178-299]	233 [160-319]
	Eosinophils (x10^9/L), median [IQR]	0.1 [0-0.1]	0 [0-0.1]	0 [0-0.1]	0 [0-0.1]

Characteristic		Admissions (full dataset)	Admissions (complete cases)	Among complete cases: died ≤30 days of starting antibiotics (first "trial")	
				No	Yes
		N=25,774 (col %)	N=16,948 (col %)	N=15,029 (col %)	N=1,919 (col %)
	Eosinophils (x10^9/L) = 0[1]	12,260 (49.4%)	8,845 (52.2%)	7,561 (50.3%)	1,284 (66.9%)
	Eosinophils (x10^9/L) = 0.1	6,495 (26.1%)	4,243 (25.0%)	3,882 (25.8%)	361 (18.8%)
	Eosinophils (x10^9/L) = 0.2-0.3	4,109 (16.5%)	2,582 (15.2%)	2,412 (16.0%)	170 (8.9%)
	Eosinophils (x10^9/L) = 0.4-0.5	1,980 (8.0%)	1,278 (7.5%)	1,174 (7.8%)	104 (5.4%)
Blood biochemistry at admission (+/-48 hours)[5]	CRP (mg/L), median [IQR]	49 [16-122]	50 [16-124]	47 [15-117]	80 [34-168]
	Albumin (g/L), median [IQR]	40 [35-43]	39 [35-43]	40 [36-43]	34 [30-38]
	ALT (IU/L), median [IQR] among non-missing	19 [13-30]	18 [13-29]	18 [13-29]	19 [13-36]
	ALT, tested at admission[1]	16,907 (65.6%)	13,782 (81.3%)	12,253 (81.5%)	1,529 (79.7%)
	ALT, not tested at admission	8,867 (34.4%)	3,166 (18.7%)	2,776 (18.5%)	390 (20.3%)
	Urea (mmol/L), median [IQR]	6.3 [4.4-9.7]	6.5 [4.5-10.1]	6.2 [4.4-9.3]	10.5 [6.7-16.9]
	Creatinine (µmol/L), median [IQR]	83 [65-113]	85 [66-117]	83 [66-113]	106 [74-166]
	Sodium (mmol/L), median [IQR]	138 [135-141]	138 [134-140]	138 [135-140]	137 [133-141]
Physiological tests at admission (+/-48 hours)[5]	Heart rate (bpm), median [IQR]	87 [75-100]	88 [76-100]	88 [76-100]	91 [79-105]
	Temperature (°C), median [IQR]	36.8 [36.4-37.3]	36.9 [36.4-37.4]	36.9 [36.4-37.4]	36.6 [36.1-37.1]
	Systolic blood pressure (mmHg), median [IQR]	124 [110-142]	124 [109-141]	124 [110-142]	117 [101-136]

[1] Used as reference group in Cox models.

[2] The lower percentage of complete cases in 2010 is because temperature, systolic blood pressure, and heart rate were not routinely reported before March 2011 and admissions missing these test results were not complete cases by definition.

[3] Since negative culture results were not available, it was not possible to differentiate between 'not done' and 'negative' results. With urine cultures, the very small number of probable contaminants (defined as isolations with >1 bacterial species) were grouped with 'not done or negative' results (n=26/16,948 spells). If the blood or urine culture that was nearest to admission was a contaminant, but there was another isolate in the +/-48h window, then the next closest isolate was used as the baseline.

[4] Defined as bacteria typically belonging to the skin flora, as per criteria described by Ammerlaan *et al*(190).

[5] The blood biomarker and physiological test results reported for the full dataset (n=25,774) reflect the available non-missing results (missingness varies by test, as reported in **Figure 3.1**).

Note: The complete cases in this table reflect the number of patients included on trial day 1 in unlagged models using stopping definition 2.

Table 3.6 Antibiotic treatment characteristics and clinical outcomes of general medical inpatients who initiated antibiotics within 1 day of admission

Antibiotic treatment characteristics and clinical outcomes		Full dataset N=25,774 (col %)	Complete cases N=16,948 (col %)
When were antibiotics administered?	In hospital (within 1 day of admission)	24,475 (95.0%)	16,466 (97.2%)
	In hospital (within 1 day of admission) and at discharge	13,420/24,475 (54.8%)	8,702/16,466 (52.8%)
	Only at discharge (within 1 day of admission)	1,299 (5.0%)	482 (2.8%)
Antibiotic regimen on first day of treatment[1]	Broad-spectrum agent(s)	13,932 (54.1%)	9,642 (56.9%)
	IV antibiotic(s)	13,653 (53.0%)	9,909 (58.5%)
	Piperacillin-tazobactam	4,583 (17.8%)	3,371 (19.9%)
	Clarithromycin + co-amoxiclav	1,922 (7.5%)	1,438 (8.5%)
	Co-amoxiclav	2,141 (8.3%)	1,394 (8.2%)
	Amoxicillin	2,111 (8.2%)	1,307 (7.7%)
	Doxycycline	1,798 (7.0%)	1,161 (6.9%)
	Flucloxacillin	1,735 (6.7%)	1,087 (6.4%)
	Ciprofloxacin	1,423 (5.5%)	631 (3.7%)
	Other(s)	10,061 (39.0%)	6,559 (38.7%)
Days to first antibiotic stop (1st definition)	Median (IQR) [10th, 90th percentile]	5 (2-7) [1-10]	5 (3-8) [1-11]
Days to first antibiotic stop (2nd definition)	Median (IQR) [10th, 90th percentile]	7 (5-9) [2-13]	7 (6-10) [2-14]
Cumulative antibiotic use at time of first stop (1st definition)	Median DDDs (IQR)	4.9 (2.0-9.8)	5.5 (2.8-10.9)
	Median days of therapy (IQR)[2]	5 (2-10)	6 (3-10)
Cumulative antibiotic use at time of first stop (2nd definition)	Median DDDs (IQR)	7.3 (4.5-13.8)	8 (6-14)
	Median days of therapy (IQR)[2]	8 (5-13)	8 (6-14)
Initial regimen was modified before stopping (1st definition) or death/censoring	Yes (%)	9,766/24,475 (39.9%)	7,429/16,466 (45.1%)
	Median days to first change (IQR)	1 (1-3)	1 (1-3)
Initial regimen was modified before stopping (2nd definition) or death/censoring	Yes (%)	10,639 (41.3%)	7,987 (47.1%)
	Median days to first change (IQR)	1 (1-3)	1 (1-3)
Escalated treatment before stopping (1st definition) or death/censoring	Yes (%), from narrow- to broad-spectrum agent(s)	1,520/10,863 (14.0%)	1,186/6,927 (17.1%)
	Median days to escalation to broad-spectrum (IQR)	2 (1-3)	2 (1-3)
	Yes (%), from oral to IV	1,118/10,834 (10.3%)	834/6,564 (12.7%)

Antibiotic treatment characteristics and clinical outcomes		Full dataset N=25,774 (col %)	Complete cases N=16,948 (col %)
	Median days to escalate to IV (IQR)	1 (1-3)	1 (1-3)
Escalated treatment before stopping (2nd definition) or death/censoring	Yes (%), from narrow- to broad-spectrum agent(s)	1,533/11,842 (12.9%)	1,196/7,306 (16.4%)
	Median days to escalation to broad-spectrum (IQR)	2 (1-3)	2 (1-3)
	Yes (%), from oral to IV	1,120/12,121 (9.2%)	836/7,039 (11.9%)
	Median days to escalate to IV (IQR)	1 (1-3)	1 (1-3)
Clinical outcomes			
Died within 30 days of starting antibiotics	N (%)	2,700 (10.5%)	1,919 (11.3%)
Re-started antibiotic treatment within 30 days of stopping (1st definition), excluding those who died on treatment or were censored at discharge)	Yes (%), including antibiotics re-started in hospital or at discharge	14,087/23,737 (59.3%)	9,620/15,965 (60.3%)
	Median days to re-start (IQR)	1 (1-1)	1 (1-1)
Re-started antibiotic treatment within 30 days of stopping (2nd definition), excluding those who died on treatment or were censored at discharge)	Yes (%), including antibiotics re-started in hospital or at discharge	3,427/24,966 (13.7%)	2,403/16,404 (14.6%)
	Median days to re-start (IQR)	3 (2-8)	4 (2-9)
Length of hospital stay	Median days (IQR)	4.9 (1.7-11.8)	5.9 (2.4-13.1)
Admitted to a critical care ward during spell	N (%)	1,330 (5.2%)	724 (4.3%)
Emergency readmission within 30 days of discharge alive (to any specialty)	N (%)	3,973/23,396 (17.0%)	2,752/15,238 (18.1%)
	Median days to readmission (IQR)	10 (4-19)	10 (4-19)

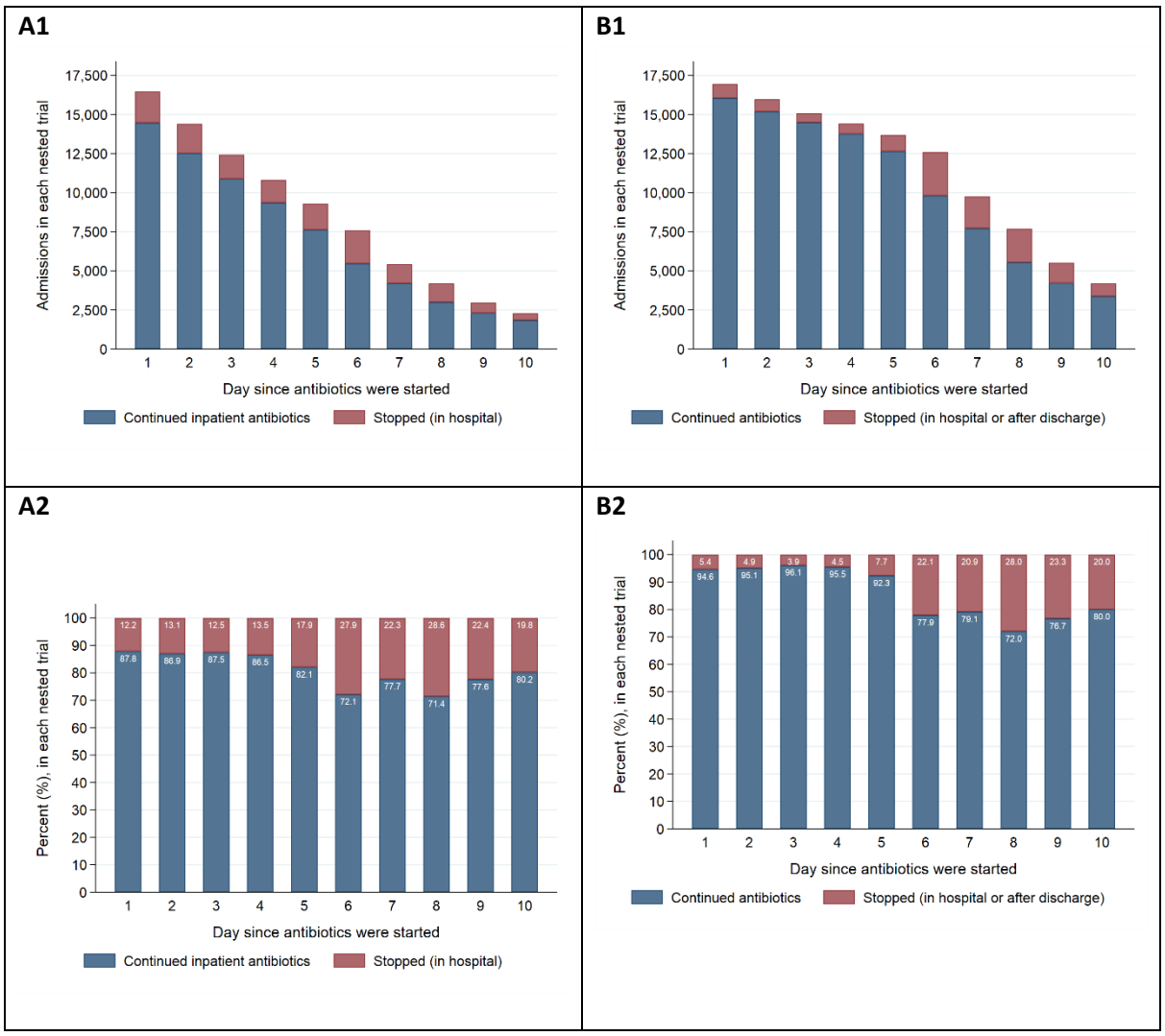
[1] Only listed regimens prescribed in >5% of spells in the full dataset. Broad-spectrum defined as: carbapenems, quinolones, polymyxins, co-amoxiclav, piperacillin/tazobactam, azithromycin, tigecycline, aztreonam, telithromycin, cephalosporin-beta-lactamase inhibitor combinations, and second-, third-, fourth-, or fifth-generation cephalosporins.

[2] Days of therapy (DOT) = sum of days between starting antibiotic treatment and stopping, with each antibiotic counted separately (i.e. 2 different antibiotics on a single day would have DOT=2)(157).

Note: 1st definition includes inpatient antibiotics only. 2nd definition of stopping includes inpatient antibiotics and antibiotics prescribed at discharge.

The proportion of complete cases stopping antibiotics remained relatively low in each of the first 5 "trials" and increased thereafter until the 10th day after treatment initiation (**Figure 3.3**). Using the 1st definition of stopping antibiotics, considering only inpatient antibiotics, 89.0% (n=14,202/15,965) of those who stopped did so within 10 days of treatment initiation (panel A1-A2). Using the 2nd definition of stopping antibiotics, including both inpatient and discharge antibiotics, 80.0% (n=13,118/16,404) of those who stopped did so within 10 days of treatment initiation (panel B1-B2).

Figure 3.3 Absolute number and proportion of complete cases stopping antibiotics on each "trial" day after treatment initiation, using either the 1st definition of stopping (A1-A2; inpatient antibiotics only) (N=16,466) or the 2nd definition of stopping (B1-B2; inpatient and discharge antibiotics) (N=16,948)



3.3.1 Causal effect of stopping vs continuing antibiotics (Models 2-3, 5-6, 8-9)

The causal effect of stopping vs continuing antibiotics on each of the first 10 days after treatment initiation is shown in **Figure 3.4**, with adjusted estimates from models lagged by 1-2 days (i.e. models 2-3, 5-6, and 8-9 from **Table 3.4**). Both pooled and unpooled models yielded effect estimates that were very similar in magnitude, as did models including trial day as categorical vs spline. I therefore focus on pooled models with a categorical effect of time (models 5-6).

Using the 1st definition of "stopping" (including inpatient antibiotics only), there was no evidence of harm in the lagged models (**Figure 3.4** panel A). Estimates from models 5-6 are shown with 95% confidence bounds in **Figure 3.5** panels A1-A2. Further there was at least some evidence for lower risk of 30-day mortality in those stopping vs continuing from day 8 onwards in these two lagged models. There was, however, also evidence of reduced risk in those stopping vs continuing on day 2, suggesting that some residual confounding may be remaining. Of note, here I am comparing trial days, thus for the 1-day lag, the day antibiotics were stopped is the previous day, i.e. trial day 2 refers to the effect of stopping antibiotics on the day of initiation (day 1). Unadjusted estimates for models 5-6 are presented in **Figure 3.6** panel A and show that, without any adjustment, stopping vs continuing antibiotics is associated with a greatly reduced 30-day mortality risk, which is consistent with confounding by indication (i.e. whereby patients who are less sick receive shorter antibiotic courses and have better clinical outcomes) and explains why adjusting for baseline confounding shifts point estimates towards the null.

Using the 2nd definition of "stopping" (**Figure 3.4** panel B; including inpatient antibiotics and those prescribed at discharge), the relative hazard of stopping antibiotics (vs continuing) peaked on trial day 5, using either a 1-day lag (model 5): HR 1.50, 95% CI: 1.19, 1.88, or a 2-day lag (model 6): HR 1.38, 95% CI: 1.08, 1.76). These estimates are also shown with 95% confidence bounds in **Figure 3.5** panels B1-B2. The fact that the 2-day lag was associated with a lower risk from stopping antibiotics compared with a 1-day lag suggests reverse causation (where active treatment, including antibiotics,

is stopped in patients who are at imminent risk of death) is still explaining some of the early harm associated with stopping antibiotics. In both of these lagged models, there was little to no evidence of harm from stopping vs continuing from trial day 6 onwards (95% CI crossed 1 or were very close to 1), with a trend towards decreasing risks from day 5 onwards. Unadjusted estimates for models 5-6 are presented in **Figure 3.6** panel B and show stopping vs continuing (using definition 2) reduced 30-day mortality risk in the first trial, but by the 3rd, 4th and 5th trials was associated with increased mortality risk. This appears counterintuitive, as patients stopping antibiotics early on treatment may be expected to be less ill and at lower risk of unadjusted mortality. However, the majority of patients who "stop" antibiotics (using definition 2) in these trials have not been discharged and remain admitted to hospital, while those who "continue" treatment may be discharged on antibiotics at a comparatively lower risk of death; as a result, those who "stop" have relatively higher unadjusted mortality than patients who "continue" antibiotics using definition 2, as can be seen in the Kaplan-Meier survival curves for the 1-day lagged exposure in **Figure 3.2** panel B3. Using stopping definition 1 the reverse is true, as the majority of patients classified as "stopping" antibiotics early on treatment are discharged and remain at lower unadjusted risk of death than patients who "continue" inpatient treatment (**Figure 3.2** panel A3). In adjusted models using stopping definition 2, having been previously discharged on antibiotics to a non-hospital destination greatly reduced 30-day mortality risk vs not having been discharged (model 5 HR 0.42, 95% CI 0.35, 0.50; model 6 HR 0.43, 95% CI: 0.36, 0.51), while having been discharged to another hospital (vs not discharged) had no impact on mortality risk (model 5 HR: 1.02, 95% CI 0.66, 1.59, model 6 HR 1.07, 95% CI: 0.69, 1.65). Although adjusting for trial baseline discharge status shifted the adjusted effect of stopping antibiotics towards the null, it did not completely eliminate the increased mortality risk on trial day 5 and some residual confounding from imperfect adjustment for this factor is possible.

Figure 3.4 Adjusted effects of stopping antibiotics (vs continuing) on 30-day mortality each day after treatment initiation, estimated in both pooled and unpooled lagged Cox models (models 2-3, 5-6, 8-9), with "stopped" defined as per definition 1 (A; inpatient antibiotics only) or definition 2 (B; inpatient antibiotics and antibiotics prescribed at discharge)

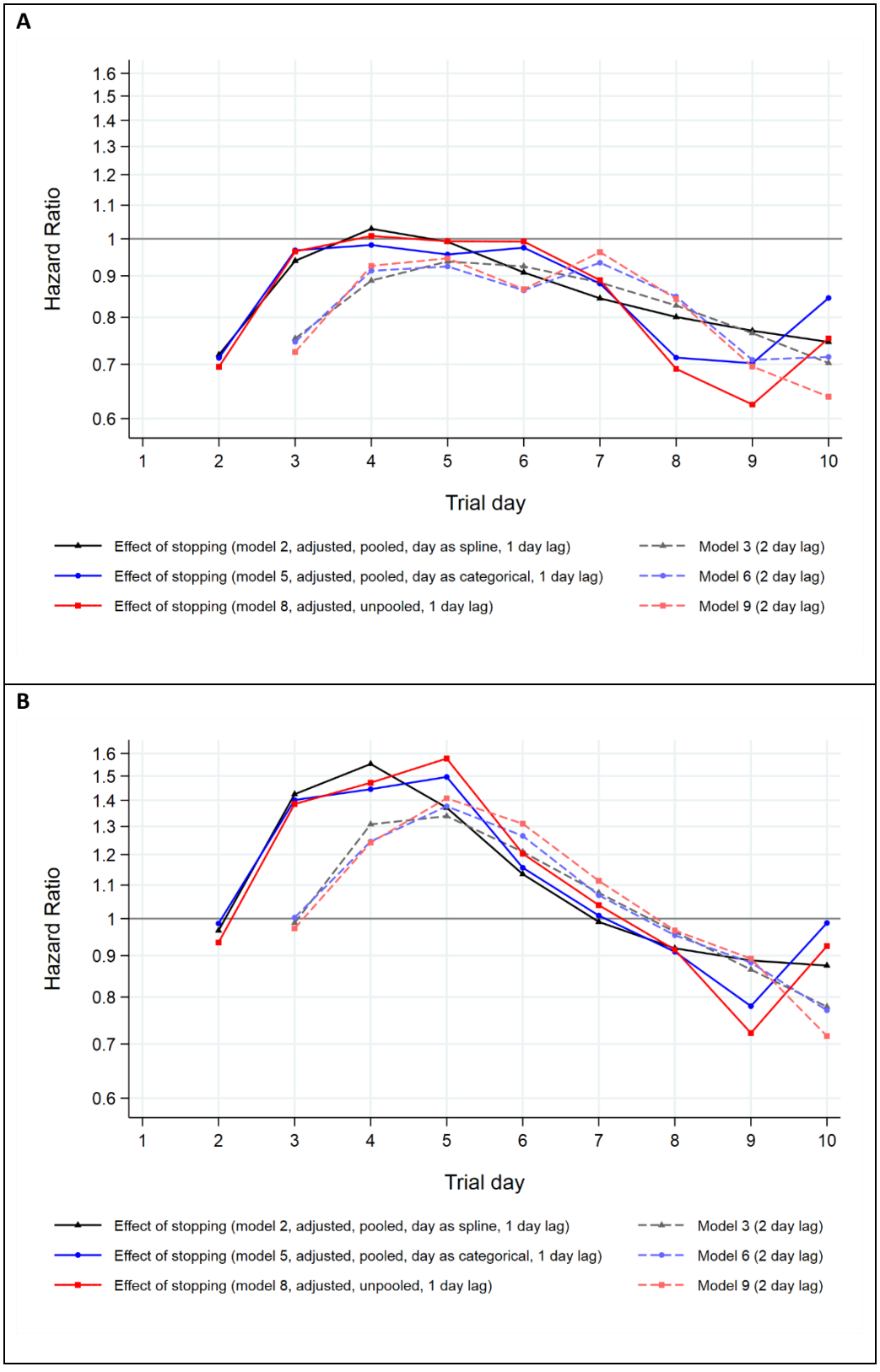
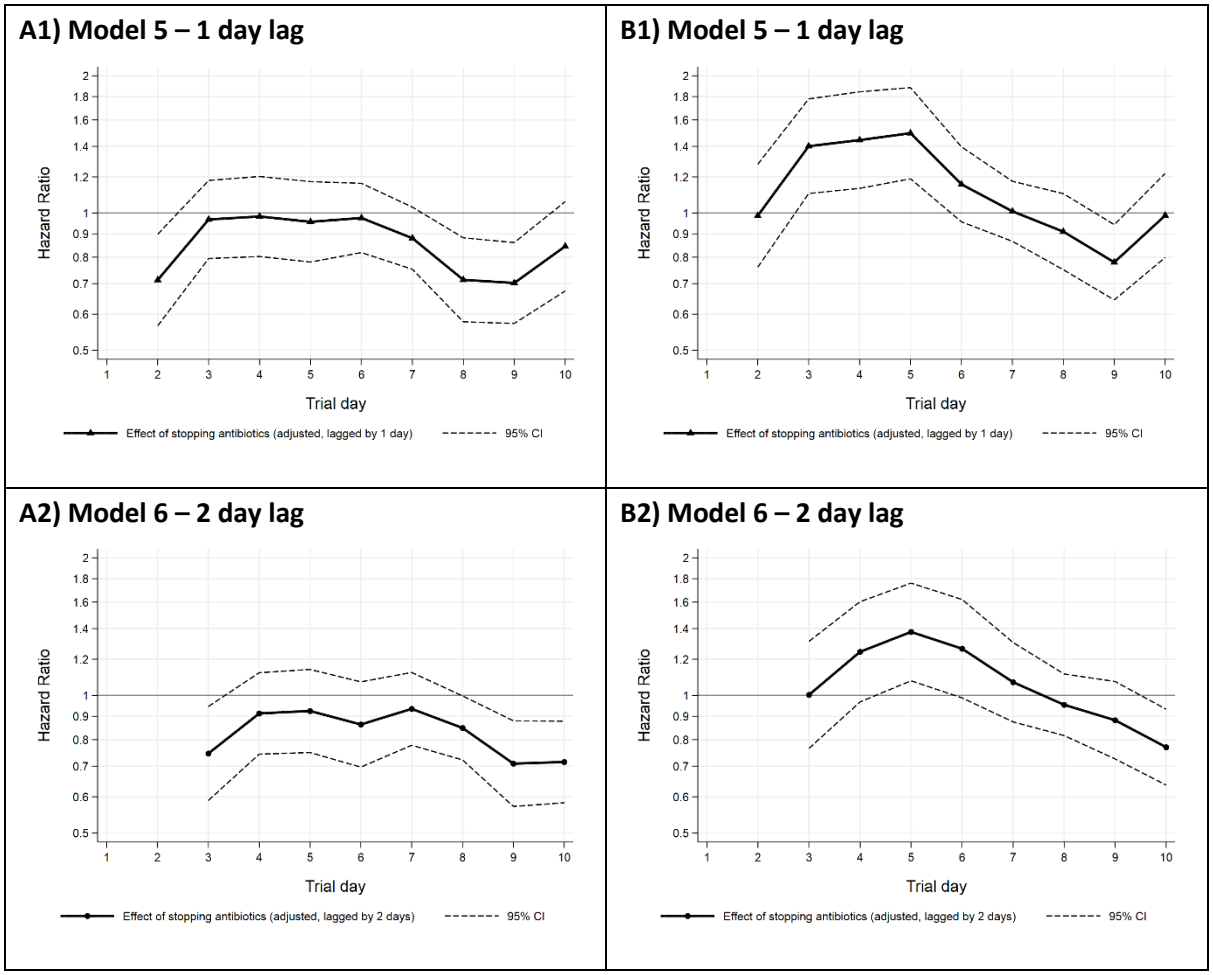
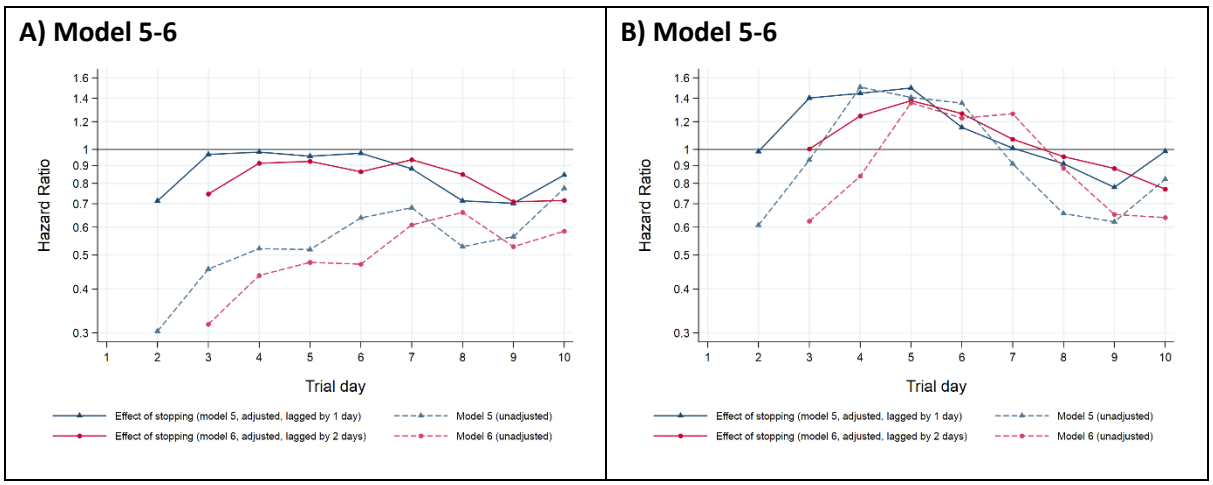


Figure 3.5 Adjusted effects of stopping antibiotics (vs continuing) on 30-day mortality each day after treatment initiation, estimated in multivariate pooled Cox models (Models 5-6), with "stopped" defined as per definition 1 (A1-A2) or definition 2 (B1-B2)



Note: These point estimates are shown in appendix **Table 3.9**.

Figure 3.6 Adjusted and unadjusted effects of stopping antibiotics (vs continuing) on 30-day mortality each day after treatment initiation (models 5-6), with "stopped" defined as per definition 1 (A; inpatient antibiotics only) or definition 2 (B; inpatient and discharge antibiotics)



Unadjusted estimates for each of models 2-3 and 8-9 are presented in appendix **Figure 3.7**. Adjusted and unadjusted associations between each biomarker and 30-day mortality are also presented in appendix **Figure 3.8**.

3.4 Discussion

In this chapter I have assessed the causal effect of antibiotic discontinuation on 30-day mortality by emulating the design of an ITT analysis in a sequentially randomised trial. The approach I used conceptualises the observational cohort as though it were a sequence of non-randomised trials, in which patients either stop or continue antibiotic therapy on each day after treatment initiation, providing insight into whether stopping antibiotics early on treatment (vs continuing) affects clinical outcomes, regardless of subsequent management, after adjusting for baseline characteristics in each trial-specific Cox model. Using this approach, I found no evidence that stopping inpatient antibiotics (definition 1) was associated with 30-day mortality, though there was some evidence of mortality risk being lower in those who stopped antibiotics vs continued antibiotics in the first trial (i.e. stopped antibiotics on the same day they started them), which likely reflects residual confounding, and also from the seventh day of treatment (trial day 8 in the models lagged by 1 day, or trial day 9 when lagged by 2 days). Stopping inpatient or discharge antibiotics (definition 2) was associated with an increased risk of death on the third day of treatment (trial day 5 in models lagged by 2 days, **Figure 3.5** panels B2); this result differs from the primary analysis due to structural differences in how stopping was defined, i.e. a larger share of patients who "stop" antibiotics early on treatment using definition 1 (vs definition 2) are discharged alive and are at low risk of death, largely because any patient who is discharged on antibiotics without a 1-day gap in treatment will be classified as having stopped under definition 1, but not definition 2.

In terms of my study population, I chose to restrict to a sample in whom I could use repeated blood test results and physiological measurements to adjust for evolving illness severity. Since failed assays are uncommon, missing biomarker tests likely reflect that the test was not ordered by the attending

physician. Thus, when compared to patients with missing test results, those tested at admission for all available biomarkers might be more acutely unwell and more likely to have serious bacterial infections that put them at higher risk of death, leading to a possible oversampling of severe cases among the complete cases and more frequent abnormal test results that are prognostic for mortality. As a result, aside from a loss of precision and power, the complete case analysis may provide biased effect estimates if test results are not "missing completely at random" (i.e. if systematic differences exist among subjects with missing and non-missing tests, as defined by Rubin(191), in a way that is not reflected in the values of the test results themselves, or in other covariates). If we assume these systematic differences are due to differences in measured covariates (tests are "missing at random"), this should be evident by comparing the distribution of these factors. However, as shown in **Table 3.5**, the distribution of baseline factors and available biomarker test results at admission in the full dataset vs the complete cases are broadly similar, as were clinical outcomes (**Table 3.6**), and as I adjusted for multiple covariates that reflect patient characteristics and comorbidities, any bias from oversampling of severe cases is likely to be limited. Also, over sampling of severe cases would, if anything, increase the observed risk of harm from early stopping if this harm truly existed.

I did not use multiple imputation to address missing test results because although it would increase power and could in theory enable unbiased estimation of causal effects (under a missing at random assumption), I considered the risk of mis-specifying complex imputation models to be relatively high given the large number of possible interactions, non-normally distributed test results, and non-linear associations that would need to be appropriately accounted for in each imputation model (one for each of the 14 missing biomarkers); making incorrect assumptions about any of these could have easily increased bias(192). Also, if the missing at random assumption is not plausible – e.g. due to unknown or unmeasured factors that are not among the covariates included in multivariable models – the bias resulting from multiple imputation may be substantial. Similarly, I did not assess interactions among the 15 biomarkers or between these test results and other baseline

characteristics due to concerns about correct model specification given the very large number of possible interactions and the multiple testing problem this posed, which was exacerbated by the fact that the Cox models adjusted for more than one test result for 13/15 biomarkers (e.g. biomarker test result at admission and at the "trial" baseline).

Although treating the cohort as a sequence of non-randomised nested "trials" avoids bias from misclassified or left-censored follow-up, this analysis has a number of limitations that complicate interpretation of these results. First, I cannot rule out unmeasured confounding, as potentially important confounders were not recorded in the electronic health records, including information regarding antibiotic use in the community before admission. Also, whilst I adjusted for positive blood and urine cultures at admission, MICs were not available before September 2010 and clinical breakpoints and drug susceptibility interpretations were not recorded, preventing adjustment for history of multi-resistant infections. Residual bias may also arise from covariates that have been measured with error, such as previous hospital exposure, which was estimated by the local informatics department at UHB and based only on previous admissions within the same Trust, and is therefore likely to be an underestimate if patients obtained previous care in another hospital. The reduced mortality risk associated with stopping inpatient antibiotics (definition 1) on the first day of treatment is also indicative of at least some residual confounding by indication from less sick patients stopping antibiotics quickly and being at lower risk of death.

A second limitation stems from the routinely collected diagnosis codes. Clinical coding departments use discharge summaries to compile the ICD-10 codes that reflect the main condition managed in an episode, and this information is then relied upon to derive CCS primary diagnosis group and comorbidity scores. Although this process is standardised and audited, the codes are primarily collected for administrative and reimbursement purposes and the clinical indication for prescribing antibiotics may be unrelated to the conditions that have been captured by the coding. As a result, adjusting for CCS group or comorbidities may effectively use the future to predict the past in a

proportion of patients, yet not adjusting for it will lead to residual confounding bias since diagnoses are not accounted for despite having greatly varied mortality risks. Unfortunately, it is not clear how this problem can be resolved when using routine electronic health records.

Third, although outcome misclassification errors are unlikely given that my outcome (mortality) was ascertained through linkage to the national death registry, some exposure misclassification may remain. For example, patients who miss inpatient doses on a given calendar day (or whose records contain date errors) may be erroneously classified as stopping. Similarly, patients with poor renal function (at higher risk of mortality) may receive intermittent doses of antibiotics such as vancomycin and gentamicin, with >24h between doses leading to a possible misclassification of the patient as having "stopped" treatment. However, this is unlikely to lead to substantial misclassification, as only 0.7% (n=122/16,948) of spells were on vancomycin or gentamicin (alone or in combination with other antibiotics) at the time of stopping within 10 days of treatment initiation. Also, some patients who appear to stop antibiotics on the day of discharge will be acutely unwell and transferred to another hospital where they continue inpatient antibiotics and remain at high risk of death. This bias would be expected to increase the hazard associated with "stopping" inpatient antibiotics (definition 1), but the lagged models revealed no such increase in the first 10 days of treatment. Finally, some patients who "stop" antibiotics will have been withdrawn from active treatment due to imminent death, and since these patients cannot be reliably identified from diagnosis codes, this may result in some reverse causation (e.g. if death occurs >2 days after treatment is withdrawn) despite the 1-2 day lag in the primary exposure that was designed to minimise this source of bias.

Fourth, I checked for interactions between select admission factors and justified some model fitting decisions (e.g. whether to fit certain confounders as spline terms, and which biomarker test results to include in adjusted models) based on the AIC and BIC of model 1. Since the primary exposure (stopped vs continued) was unlagged in this model, the estimates derived from it will be biased by

exposure misclassification (shown in **Figure 3.2**), though I had not yet identified the extent of this misclassification problem when building the adjusted model.

Where health records only capture prescription start and end dates, investigators will only be able to contrast the effect of treatment assignment (i.e. the observational analogue of an ITT effect), since adherence to the prescribed regimen will be unknown. However, since prescription assignment may provide a biased reflection of treatment efficacy, investigators with access to data showing whether or not individual doses were administered might reasonably be interested in estimating the effect of adherence to an assigned regimen, i.e. analogous to a per-protocol effect. To do so, I could have compared alternate treatment strategies (e.g. stopping antibiotics after 1 day, or 2 days, or 3 days, and so forth) using a 3-step cloning, censoring, and inverse probability of censoring weighting (IPCW) procedure, which emulates the per-protocol effect of a target RCT with full adherence(76,193) and has been recently applied to address questions of antibiotic treatment duration(98,99), de-escalation(194) and the impact of treatment delays(69). This approach considers each subject exposed to every treatment strategy at the start of follow-up, with one observed and the others counterfactual. This is achieved by creating perfect duplicates (clones) in an expanded dataset, with each clone assigned a different treatment strategy. By adopting this analytic approach, it is no longer possible for outcome events to be systematically assigned to one treatment strategy through misclassified follow-up, and the outcome risk in each treatment arm can be estimated as though exposure assignment had been random in the study population, achieving the desired exchangeability between observed and counterfactual risks. Clones are censored when they deviate from the assigned strategy, and this procedure is repeated for all subjects, with artificial censoring used to emulate compliance with the assigned treatment arm. Finally, to eliminate time-varying selection bias induced by censoring, IPCW would be used(195). This last step ensures each uncensored subject's weight at time *t* is equal to the inverse of the probability of remaining uncensored at time *t*, conditional on the subject surviving through time *t* and their treatment and covariate history. Probabilities and stabilised weights could be determined by combining fitted

values from a series of propensity score (PS) models at each time *t* (i.e. 1/PS for censored participants, and 1/(1-PS) for uncensored subjects), which may be estimated from pooled logit models, with stabilised weights being truncated to protect against model misspecification(196). By including baseline and time-varying factors associated with adherence to the assigned treatment strategy in the propensity scores models, censoring would be independent of the measured covariates in the weighted pseudo-population. However, this assumes confounders are accurately measured and models are properly specified with no unmeasured confounding, which is unlikely given the potential sources of unmeasured confounding and challenges with diagnosis codes being recorded retrospectively, as noted above. Also, this approach does not prevent reverse causation from imminent death leading to the withdrawal of active treatment. Given these challenges and the fact that the underpinning data is relatively old (potentially limiting generalisability, as it is unclear if these results would hold with antibiotics administered today), I have chosen not to pursue this alternate analytic strategy, or assess other clinical outcomes such as emergency readmission risk, length of stay, or critical care admission.

Given that the primary analysis emulated the target trial protocol relatively well and adjusted for a wide range of potential baseline confounding factors, it is unlikely that harm from early stopping of inpatient antibiotics (definition 1) is being masked by residual confounding. If time were available, I could in theory assess the robustness of these results to unmeasured confounding through the use of quantitative bias analysis(197) and by computing "E-values"(198). This would help to assess the magnitude and prevalence of unmeasured confounding required to change inferences, and whether such a confounder is plausible in light of known predictors of antibiotic duration and clinical outcomes. Nevertheless questions regarding generalisability would remain, given the age of the underlying data. Although the analysis of stopping inpatient or discharge antibiotics (definition 2) found some evidence of increased mortality risk from stopping on the third day of treatment, this analysis makes the large untestable assumption that antibiotics administered out of hospital are taken and stopped as prescribed, and it is possible that reverse causation from patients stopping

treatment due to imminent death continues to bias these estimates, given the shift in the point estimates towards the null between models with a 2-day vs 1-day lag. Looking ahead, were more recent patient-level antibiotic data available, the analytic approach described in this chapter could be replicated to emulate the ITT effect of other target trial protocols that compare a wide range of alternate antibiotic treatment durations that have not already been compared using RCTs.

3.5 Appendix

Antibiotics administered in hospital had a date and time associated with each dose. As a result of possible date errors, in the complete cases doses were recorded before admission in 1.6% of spells (n=268) by median 2.5 hours (IQR: 0.7-10.8) and after discharge in 4.5% or spells (n=765) by median 12.0 hours (IQR: 1.1-48.9). To address this problem, doses administered ≤1 day before admission (n=246/268) or after discharge (n=594/765) were shifted to equal the admission and discharge time respectively, while doses administered >1 day before or after the spell were dropped (n=351).

Antibiotics prescribed at discharge should have a prescription start and end date recorded. In 3 spells, discharge prescription start date was missing and set equal to the time prescribed, while in 81 spells the discharge prescription end date was missing and set equal to the lowest mode duration for other prescriptions of the same drug and route of administration (based on prescriptions from the same patient if available (n=14/81), otherwise based on prescriptions across all patients (n=67/81)). When a discharge prescription had start and end dates that overlapped with the patient's time in hospital, the prescription was either dropped if its start and end date were before discharge (i.e. complete overlap, n=964 spells), or the start date was brought forward to the discharge date (by a median of 6.0 hours, IQR: 2.4-21.8 hours, n=4,836 spells). If the first dose of a discharge prescription overlapped with the last inpatient dose of the same antibiotic (i.e. both administered on the day of discharge), the overlapping discharge dose was dropped (n=188).

A look-up table containing DDDs was updated to reflect the 2022 WHO ATC/DDD Index(199) and used to calculate DDDs associated with each administered dose. Where an administered dose was not clear (e.g. recorded as "500-750mg"), the midpoint dose was assumed (n=68 spells). Outlier doses were flagged (e.g. those that were an order of magnitude higher/lower than all others) and likely dose errors were corrected with input from a Consultant in Infectious Diseases (n=37 spells).

Table 3.7 Akaike information criterion from a pooled Cox model adjusted for the main effect, trial day, and 18 baseline factors (but no biomarkers/physiology tests)

How was trial day fit?	AIC[1]
Categorical	251,054.5
Continuous (linear)	251,106.7
Natural cubic spline with 3 Harrell knots	251,087.3
Natural cubic spline with 4 Harrell knots	251,036.3
Natural cubic spline with 5 Harrell knots	251,039.1

[1] "Best" model shown in shaded cell, based on minimising the AIC.

Table 3.8 Akaike information criterion from a pooled Cox model with trial day fit as a 4 knot natural cubic spline, adjusted for the main effect and 18 baseline factors

	Akaike information criterion (AIC) from model 1						AIC in best model – 2nd best model (best)[3]
Biomarker	Model a	Model b	Model c	Model d	Model e	Model f	
Platelets[1]		250,347.3	250,385.3			249,702.4	644.9 (f)
Creatinine[1]		250,228.1	250,851.6			250,150.5	77.5 (f)
Hb[1]		250,495.3	250,889.3			250,421.0	74.3 (f)
Eosinophils[2]	279,957.2	250,355.8		250,286.1			69.8 (d)
Systolic BP	250,082.3	249,649.1	251,013.7	249,494.6	249,720.4	249,545.5	50.9 (d)
Temperature	249,871.8	249,512.5	250,922.1	249,103.0	249,589.7	249,085.1	17.9 (f)
Urea	249,822.1	249,285.7	250,877.5	249,284.3	249,280.7	249,266.5	14.2 (f)
Neutrophils	250,940.5	250,217.3	250,287.8	249,830.3	249,815.4	249,823.2	7.8 (e)
CRP	250,358.8	250,062.3	250,951.4	250,009.6	250,028.2	250,015.7	6.0 (d)
Heart rate	250,459.9	249,340.6	250,745.0	249,305.3	249,300.6	249,317.6	4.7 (e)
Monocytes	250,980.3	250,878.7	250,973.1	250,798.4	250,936.0	250,794.6	3.9 (f)
Albumin	247,193.9	247,406.4	250,944.4	246,844.3	246,889.8	246,847.7	3.4 (d)
Na	250,585.0	250,144.8	250,893.4	250,022.1	250,485.3	250,019.1	3.0 (f)
Lymphocytes	250,908.2	250,555.4	250,875.8	250,499.9	250,617.4	250,501.1	1.3 (d)
Ever ALT[1]		251,028.8					N/A

[1] Test results at admission were highly correlated with 'current' results obtained after admission (Spearman's rho ≥0.88), therefore only 'current' measures were considered and models including baseline results at admission were not fitted (i.e. models a, d, and e). ALT was missing at admission in 18.7% (n=3,166) of spells and was therefore modelled as binary (i.e. ever vs never tested) and models adjusting for changes in ALT (models c, e, and f) were not fitted.

[2] Fit as categorical as there was limited variation in the truncated distribution. As a result, models adjusting for changes in test results (models c, e, and f) were not fitted.

[3] "Best" model shown in shaded cells, based on minimising the AIC.

Table 3.9 Causal effect of stopping antibiotics on 30-day mortality in of the first 10 days after treatment initiation (Models 5-6)

Stopping defined using definition 1				
Trial day	**Model 5**		**Model 6**	
	N[1]	**aHR (95% CI)**	**N[1]**	**aHR (95% CI)**
1				
2	16,400	0.71 (0.57, 0.90)		
3	14,305	0.97 (0.79, 1.18)	16,246	0.75 (0.59, 0.95)
4	12,359	0.98 (0.80, 1.20)	14,179	0.91 (0.74, 1.12)
5	10,740	0.96 (0.78, 1.17)	12,234	0.92 (0.75, 1.14)
6	9,245	0.98 (0.82, 1.16)	10,672	0.86 (0.70, 1.07)
7	7,536	0.88 (0.75, 1.03)	9,164	0.93 (0.78, 1.12)
8	5,398	0.71 (0.58, 0.88)	7,478	0.85 (0.72, 1.00)
9	4,167	0.70 (0.57, 0.86)	5,361	0.71 (0.57, 0.88)
10	2,956	0.85 (0.67, 1.06)	4,131	0.72 (0.58, 0.88)
Stopping defined using definition 2				
Trial day	**Model 5**		**Model 6**	
	N	**aHR (95% CI)**	**N**	**aHR (95% CI)**
1				
2	16,882	0.99 (0.76, 1.28)		
3	15,869	1.40 (1.10, 1.78)	16,728	1.00 (0.77, 1.31)
4	15,010	1.45 (1.13, 1.84)	15,742	1.25 (0.97, 1.60)
5	14,346	1.50 (1.19, 1.88)	14,883	1.38 (1.08, 1.76)
6	13,650	1.16 (0.96, 1.40)	14,273	1.27 (0.99, 1.62)
7	12,542	1.01 (0.87, 1.17)	13,566	1.07 (0.88, 1.31)
8	9,727	0.91 (0.75, 1.10)	12,483	0.95 (0.82, 1.11)
9	7,664	0.78 (0.64, 0.94)	9,688	0.88 (0.73, 1.07)
10	5,500	0.99 (0.80, 1.22)	7,627	0.77 (0.64, 0.93)

[1] The number of complete cases included in each "trial" is lower when stopped is defined using definition 1 because there are a small number of complete cases that are never administered inpatient antibiotics, but instead are only prescribed antibiotics at discharge (which takes place within 1 day of admission). Although these patients were included in analyses using definition 2, they were excluded from analyses of definition 1.
aHR: adjusted Hazard Ratio.

Figure 3.7 Adjusted and unadjusted effects of stopping antibiotics (vs continuing) on 30-day mortality each day after treatment initiation (models 2-3, 8-9), with "stopped" defined as per definition 1 (A1-A2; inpatient antibiotics only) or definition 2 (B1-B2; inpatient and discharge antibiotics)

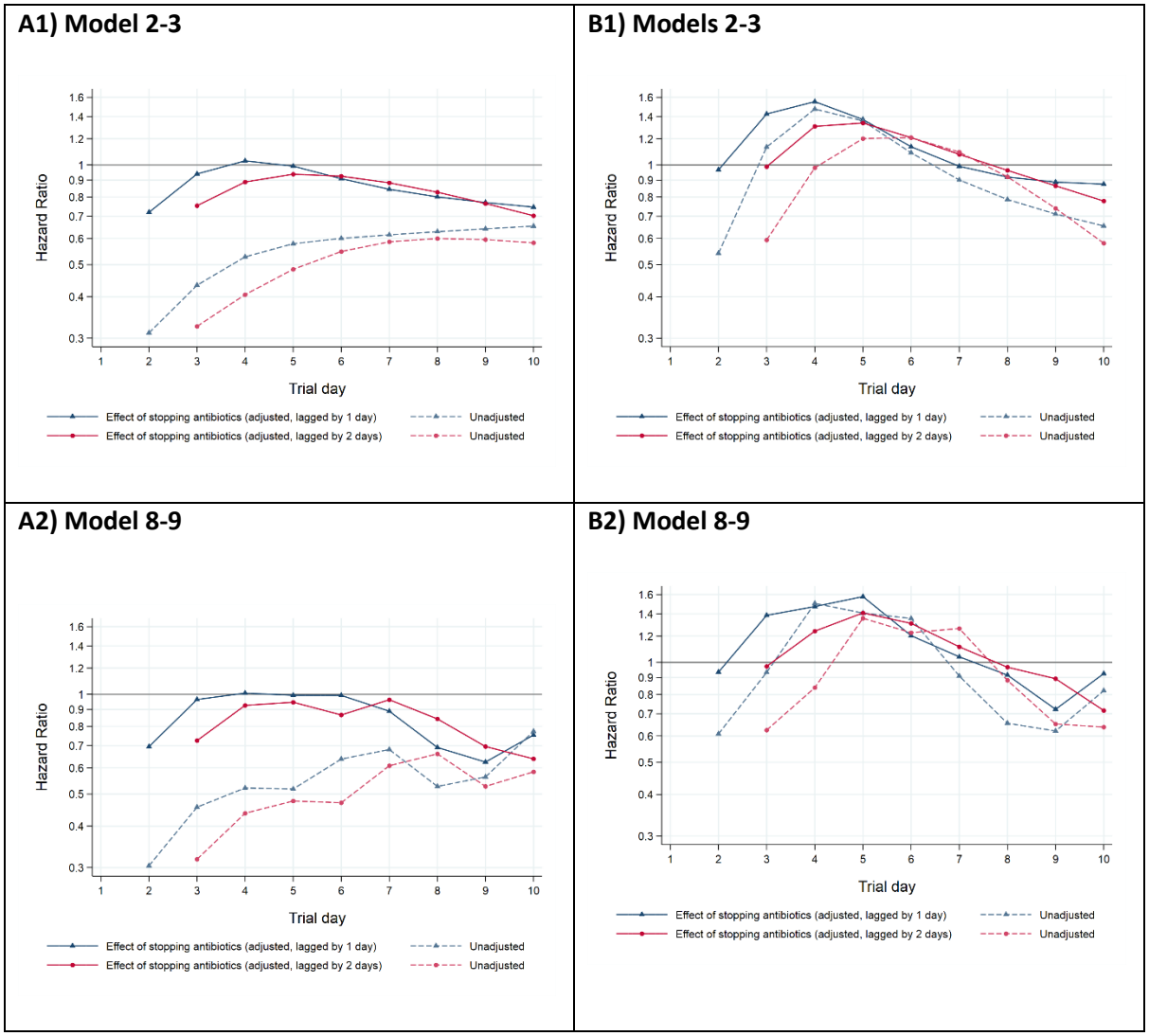
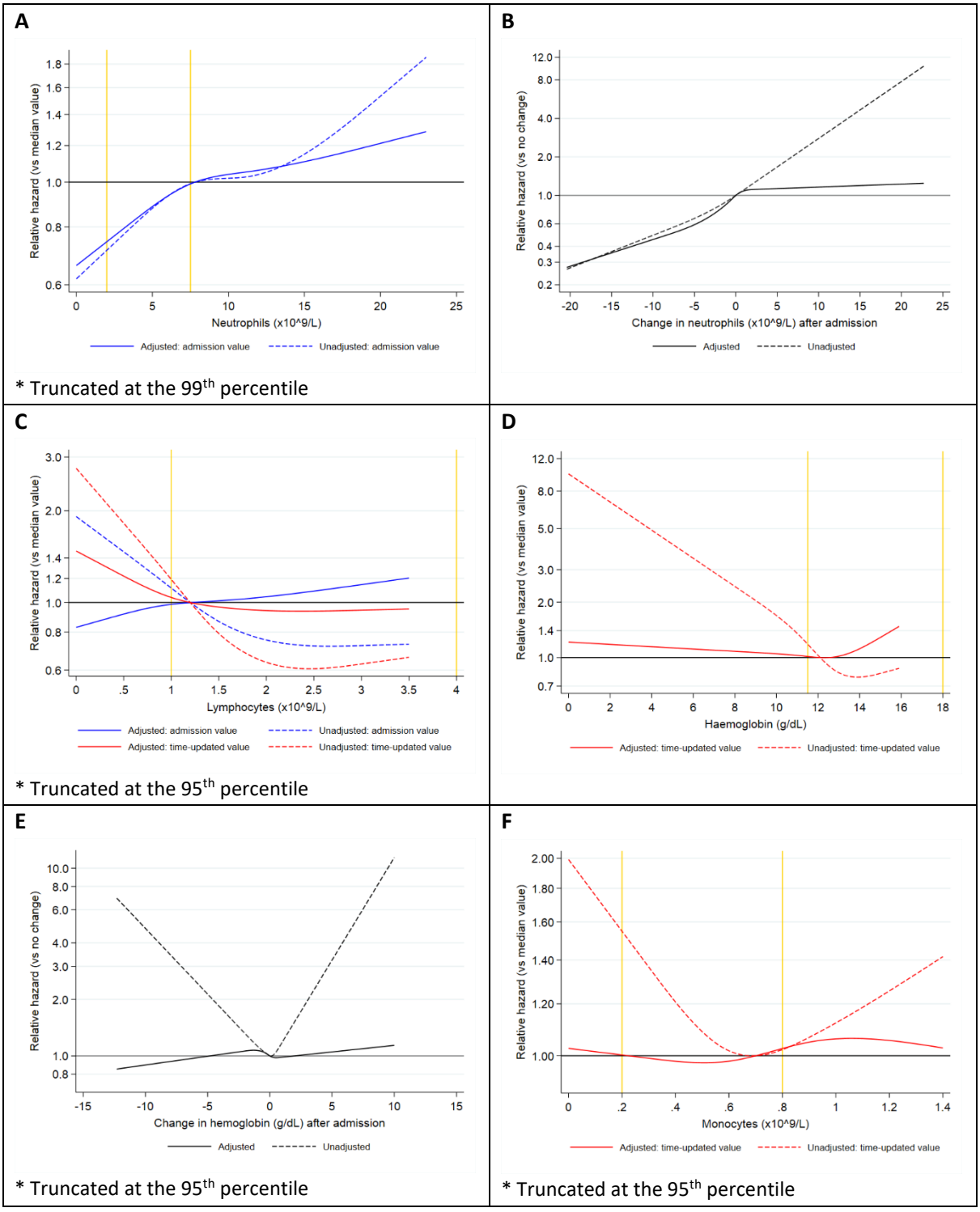


Figure 3.8 Unadjusted and adjusted associations between 13 biomarkers and 30-day mortality, as estimated in a pooled Cox model (adjusted estimates from model 2 as an exemplar) including neutrophils (A,B), lymphocytes (C), haemoglobin (D,E), monocytes (F,G), platelets (H,I), CRP (J), albumin (K), creatinine (L,M), urea (N,O), sodium (P,Q), heart rate (R,S), temperature (T,U), systolic BP (V)

Note: Reference ranges were supplied by the hospital and are shown in yellow. Eosinophils (fit as categorical) and ALT (fit as binary) are not included below.



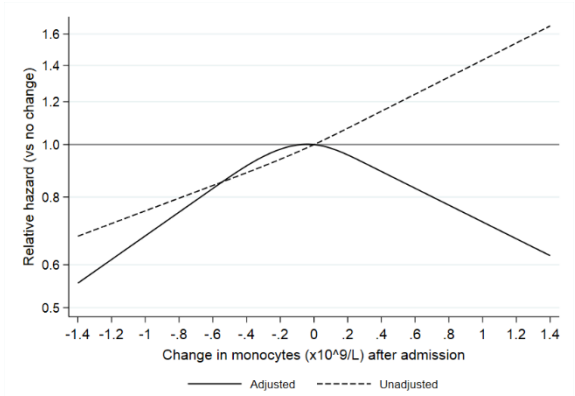
* Truncated at the 99th percentile

* Truncated at the 95th percentile

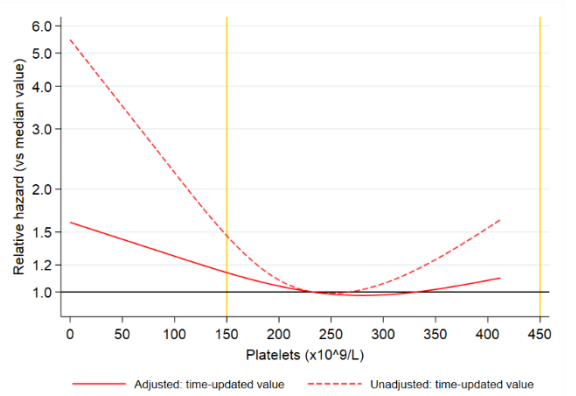
* Truncated at the 95th percentile

* Truncated at the 95th percentile

G

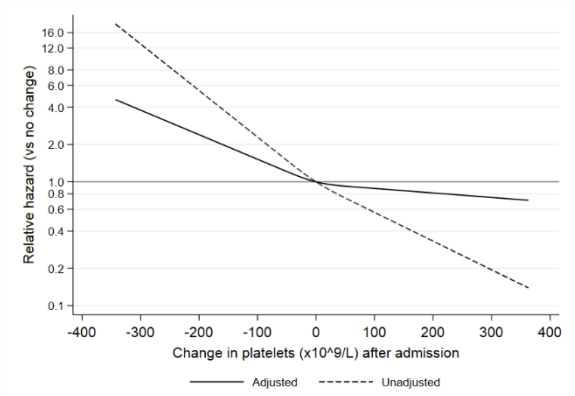


H

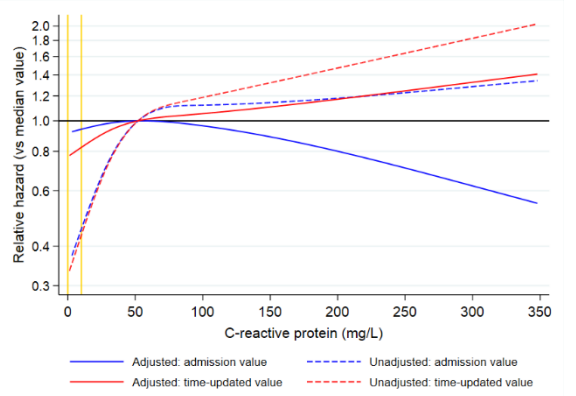


* Truncated at the 95th percentile

I

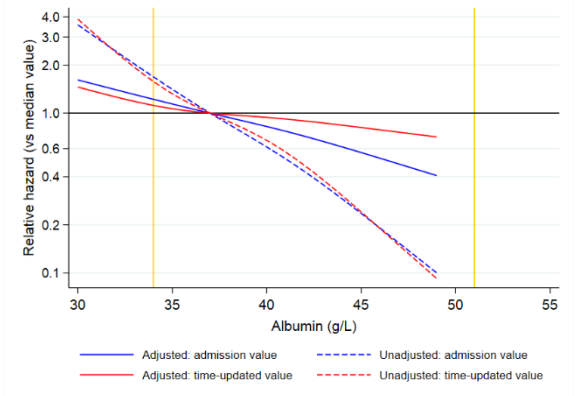


J



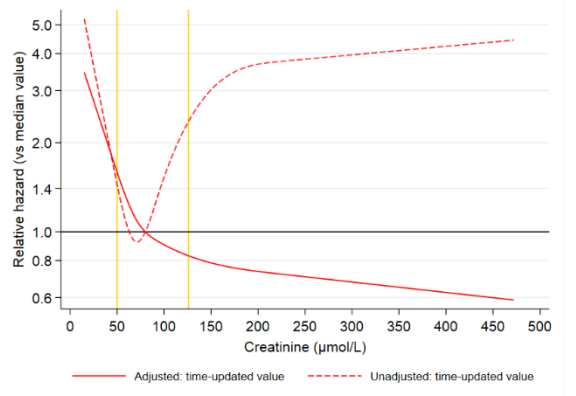
* Truncated at the 99th percentile

K



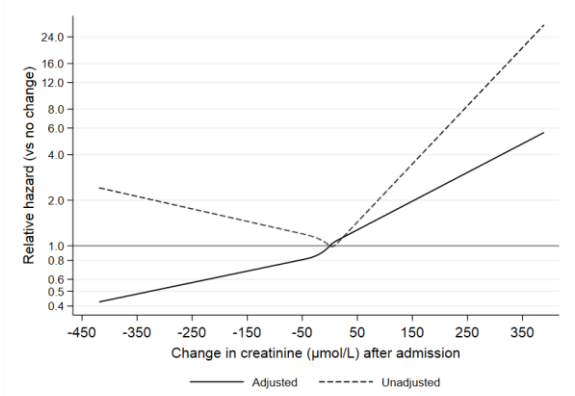
* Truncated at the 5th and 95th percentile

L

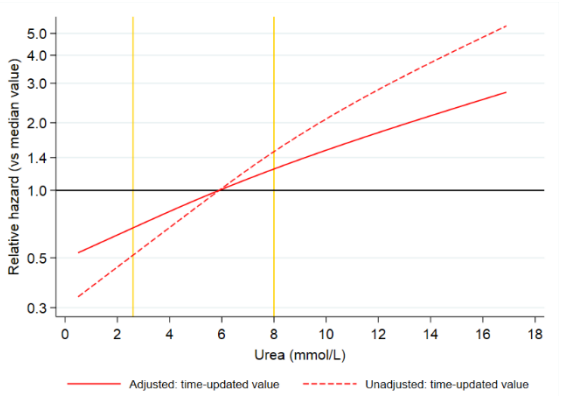


* Truncated at the 99th percentile

M

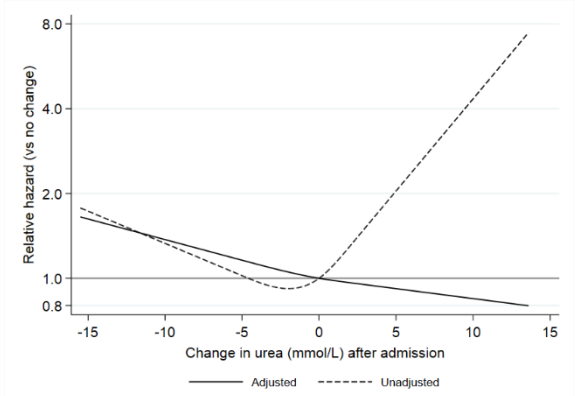


N

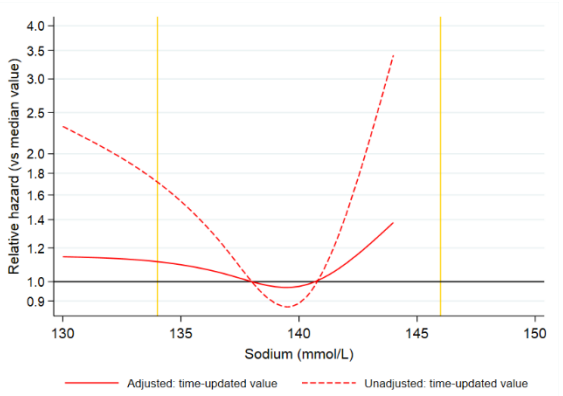


* Truncated at the 95th percentile

O

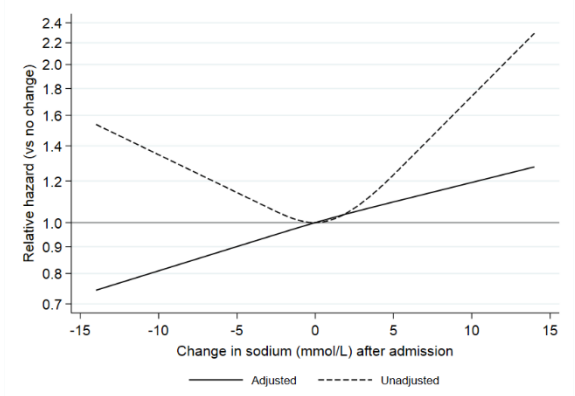


P

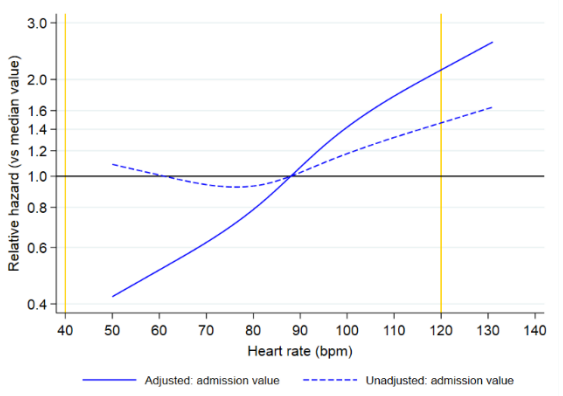


* Truncated at the 5th and 95th percentile

Q

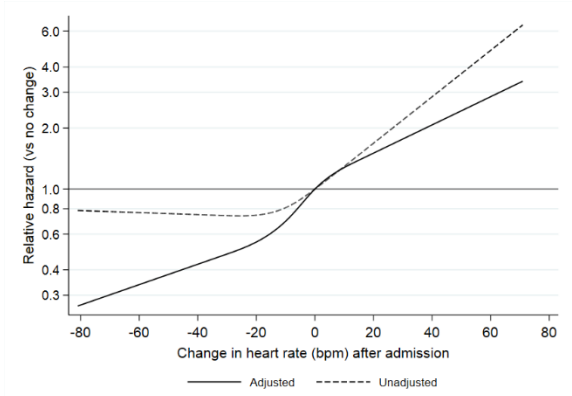


R

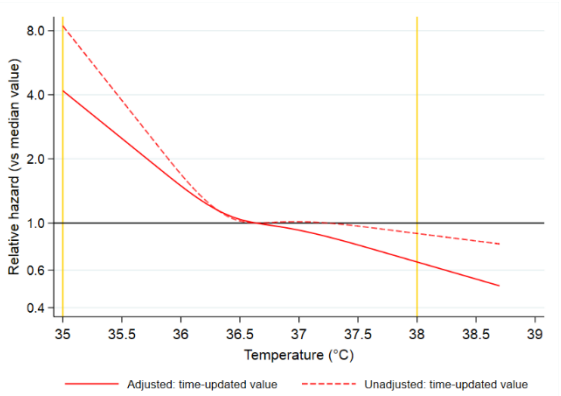


* Truncated at the 1st and 99th percentile

S

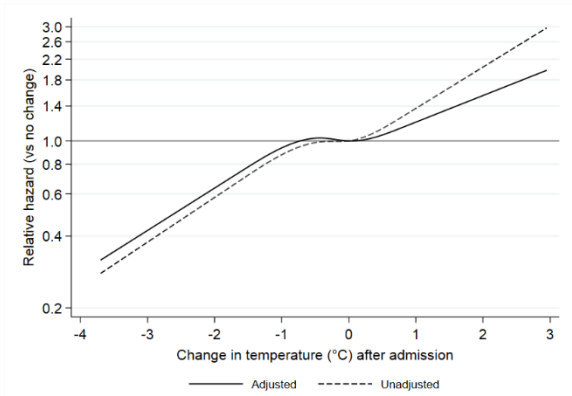


T

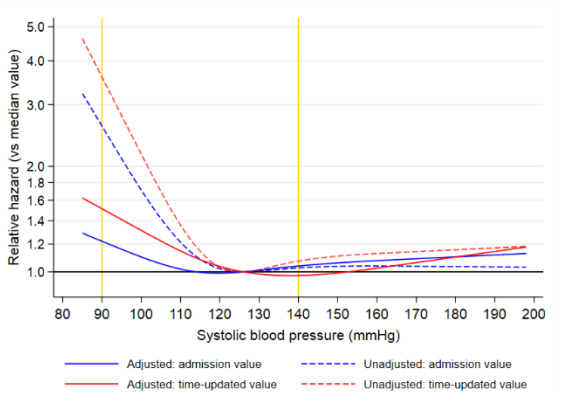


* Truncated at the 1st and 99th percentile

U



V



* Truncated at the 1st and 99th percentile

Chapter 4: Antibiotic Review Kit for Hospitals (ARK-Hospital): an expanded process evaluation for a stepped-wedge cluster-randomised controlled trial

Note: Through my employment in Nuffield Department of Medicine, I served as the statistician on the ARK-Hospital trial(105). The Statistical Analysis Plan for the trial pre-specified 8 factors to be explored as potential mediators of any intervention effect; during my analysis of these mediators, I developed a plan to assess multiple other mediators and also effects on other outcomes, as an additional secondary use of the dataset to try to explain the main trial findings. This chapter contains these analyses.

4.1 Introduction

Antimicrobial stewardship (AMS) programmes aim to optimise antibiotic use, treating bacterial infections effectively while avoiding excessively prolonged or broad-spectrum treatment(200–202). These programmes have been shown to reduce treatment costs(203–205) and the incidence of infections and colonisation with antibiotic-resistant bacteria(205,206), albeit inconsistently(207), and AMS forms an essential part of the UK's five-year plan to tackle AMR(40).

Although there is no 'one size fits all' approach to stewardship, safe and effective strategies to reduce overuse in primary care include 'restrictive' interventions, such as avoiding or delaying treatment of respiratory tract infections(208–211), though this may require careful targeting of low-risk patients(212) with support from resources such as the UK's TARGET (Treat Antibiotics Responsibly, Guidance, Education Tools) toolkit(213). By contrast, in hospital settings clinical urgency and diagnostic uncertainty mean antibiotics are often prescribed rapidly and empirically, as short delays can increase mortality risk in patients with severe bacterial infections(117). As a result, efforts to reduce hospital prescribing depend on prescribers reviewing and stopping antibiotics when appropriate. There is good evidence 'enablement' interventions, such as providing decision support or feedback to prescribers, can safely reduce hospital prescribing and improve guideline compliance,

though the risk of bias in these studies is often high and treatment duration is rarely included as an endpoint(207).

To promote hospital prescription reviews in England, the Department of Health adopted a policy called “Start Smart, then Focus” in 2012(49). Designed to reduce mortality risk associated with delays in receiving effective antibiotics for those presenting with sepsis, this policy provides guidance for rapid empiric prescribing followed by prescription reviews every 48-72 hours. However, antimicrobial pharmacists reported there was limited monitoring of compliance with this policy in 2014(60) and nearly 3 in 10 (105/148) acute Trusts still lacked formal procedures requiring 48-72h reviews 3 years later(214). Prescription stop rates also remained low in 2017 (<10%)(55) despite evidence that higher stop rates (20-30%) can be achieved safely(56) and the introduction financial incentives (CQUIN)(41) and targets to reduce overuse in NHS hospital contracts(45).

To support prescribers with stopping antibiotics at review, evidence-based tools are needed that can be readily adopted into routine care. Recommendations are available to optimise the design of AMS research(215–219), but reviews of AMS interventions show study design weaknesses are widespread, undermining external validity and the translation of evidence into clinical practice(47,220). Studies are often non-experimental, conducted in single centres, underpowered, and fail to report safety or microbiological endpoints or mediators of impact. The quality of intervention reporting is also generally poor(207), with key details often missing such as *who* should act and *when*(221), making replication and evidence synthesis difficult. Since the quality of evidence to support specific stewardship objectives is low and heterogeneous(222), it is often unclear what stewardship interventions are effective in terms of clinical and microbiological outcomes, or how those intervention can be adapted into clinical practice(223).

To address this limited evidence base, the ARK-Hospital programme was launched in 2016 to develop(121,224,225) and test(226) a multifaceted behaviour change intervention to safely reduce hospital antibiotic overuse by supporting prescriber decisions to stop prescriptions at review.

4.1.1 ARK study design and outcomes

The ARK intervention was tested in a stepped-wedge, cluster (hospital)-randomised controlled trial in 39 acute NHS Trusts providing unplanned (emergency) adult medical services(105). The intervention comprised 4 components, including:

- a prescribing decision aid, encouraging prescribers to clarify the level of diagnostic uncertainty at antibiotic initiation by recording infection risk as either "possible" or "probable", as this information helps prescription reviewers to understand whether antibiotics were prescribed merely as a precaution (hence intrinsically providing "permission" for a reviewer to stop antibiotics), or if a bacterial infection was considered the most likely diagnosis
- an online training tool to support its use(227)
- audit and feedback tools(225), and
- a patient leaflet(224)

By aiming to increase prescription stop rates, the intervention targeted treatment duration rather than the initial prescription choice.

Since the trial aimed to reduce antibiotic use safely, two co-primary endpoints were assessed: (i) antibiotic DDDs per adult acute/general medical admission (hospital-level, superiority), and (ii) all-cause mortality within 30 days of admission in or out of hospital (patient-level, non-inferiority, relative margin 5% for an immediate step change associated with the randomised intervention implementation date, assuming constant rates before and after this (interrupted time series design, see below)). Secondary endpoints included other non-inferiority clinical outcomes and specific categories of antibiotics, as described in(105).

4.1.2 ARK statistical methods

The immediate and sustained year-on-year intervention effect was estimated in each site using an interrupted time series analysis based on the randomised implementation date, following an ITT approach, that is, ignoring any differences in when the intervention was implemented and how well. Antibiotic endpoints were measured as monthly DDDs on wards that implemented ARK per adult (≥16 years) acute/general medical admission that month (henceforth described simply as DDDs/admission), modelled using negative binomial regression. Due to limitations posed by local pharmacy systems, 2 sites were only able to provide hospital-level antibiotic data, rather than ward-level data. Patient-level 30-day mortality was modelled using logit models, with adjustments for confounding as described in(105). Overall effects were derived by combining site-specific estimates through random-effects meta-analysis (Stata command: metan, random)(228), which assumes the intervention effect followed a distribution across sites, with between-site variance occurring due to both sampling error and (unexplained) sources of heterogeneity. Thus overall effects are a weighted mean of the site-specific effects(229). Non-inferiority was declared if the overall immediate implementation effect estimate had an upper 95% confidence limit lower than a 5% relative increase in 30-day mortality odds.

4.1.3 Summary of the impact of the ARK intervention on antibiotic outcomes

A summary of immediate and sustained implementation effects for all trial endpoints are in Figure 3 in (105). In brief, the ARK intervention was associated with sustained year-on-year reductions in total antibiotic DDDs/admission of -4.8% per year (95% CI: -9.1%, -0.2%, p=0.042) over a median 23 months (range: 14-37), compared with the 0.5%-1.7% year-on-year reduction targets outlined in NHS hospital contracts (Figure 3 in (105))(45). Sustained year-on-year reductions were also seen among antibiotics often used as first-line and for de-escalation, including narrow-spectrum (-5.2%, 95% CI: -9.4%,-0.9%), oral (-6.4%, 95% CI: -11.2%,-1.4%), a modified version(42) of the WHO-categorised “Access” (-5.3%, 95% CI: -10.0%,-0.4%) and “Watch” agents (-11.0%, 95% CI: -17.1%,-

4.5%), and quinolones (-13.8%, 95% CI: -22.5%,-4.0%). No evidence of reduction was observed for broad-spectrum, parenteral, or Reserve agents, with slight increases in year-on-year use for carbapenems (+12.3%, 95% CI: +2.3%,+23.2%) (appendix Figures S10-S13 in (105)), although absolute levels of and increases in carbapenem DDDs were very small and increases have been reported across English Trusts over the time frame of the trial(34).

4.1.4 Summary of the impact of the ARK intervention on clinical outcomes

Although there was weak evidence for sustained increases in 30-day mortality (odds ratio, OR: +3.0%, 95% CI: -0.1%, +6.2%, p=0.060) (Figure 5 in (105)), this was not observed when excluding admissions after the start of the COVID-19 pandemic (OR -0.9%, 95% CI: -4.0%, +2.4%, p=0.602), which began after implementation in all sites. There was no evidence that sites with greater antibiotic reductions had greater increases in mortality risk, nor was there evidence of impact on hospital length-of-stay, intensive care admission, or emergency readmission risk within 30 days of discharge. This suggests the ARK intervention was safe and the limited evidence for increasing mortality risk was due to residual confounding from the COVID-19 pandemic. This is discussed extensively in the main trial paper.

4.1.5 Remaining questions

Process evaluations are recommended to help identify the precise mechanisms and mediators of impact in complex interventions, including the role of contextual factors and implementation fidelity(230–233). This information is considered a key research priority to help optimise and adopt complex intervention in other settings(234,235), prompting calls for their widespread inclusion in trials(207).

Given the multifaceted nature of the ARK intervention and the geographic distribution of the participating sites (i.e. 32 in England, 4 in Northern Ireland, 2 in Wales, and 1 in Scotland), it is

perhaps unsurprising that heterogeneity in the site-level effect estimates was moderate to high, with a median I^2 heterogeneity statistic across antibiotic endpoints of 74% (IQR: 68-84%, **Table 4.1**).

Table 4.1 Heterogeneity in site-level implementation effect estimates

Study endpoint[1]	**Site-specific implementation effect estimates**	**I^2 from random-effects meta-analysis**	**p-value**
Total antibiotics (co-primary endpoint)	Immediate effect	66.9%	p<0.001
	Sustained year-on-year effect	88.5%	p<0.001
Narrow-spectrum	Immediate effect	67.6%	p<0.001
	Sustained year-on-year effect	83.3%	p<0.001
Broad-spectrum	Immediate effect	68.5%	p<0.001
	Sustained year-on-year effect	91.9%	p<0.001
Access	Immediate effect	71.4%	p<0.001
	Sustained year-on-year effect	84.8%	p<0.001
Watch	Immediate effect	71.3%	p<0.001
	Sustained year-on-year effect	91.8%	p<0.001
Reserve	Immediate effect	77.8%	p<0.001
	Sustained year-on-year effect	77.3%	p<0.001
Oral	Immediate effect	69.3%	p<0.001
	Sustained year-on-year effect	87.9%	p<0.001
Parenteral	Immediate effect	67.0%	p<0.001
	Sustained year-on-year effect	82.6%	p<0.001
Quinolones	Immediate effect	74.3%	p<0.001
	Sustained year-on-year effect	88.8%	p<0.001
Carbapenems	Immediate effect	66.4%	p<0.001
	Sustained year-on-year effect	74.3%	p<0.001
30-day mortality (co-primary endpoint)	Immediate effect	57.3%	p<0.001
	Sustained year-on-year effect	59.3%	p<0.001

[1] Antibiotic endpoints measured as monthly DDDs per adult general medical admission.

To understand sources of this between-site variation, in the main trial paper I used meta-regression techniques to explore how select factors (8 pre-specified, 12 identified by me) may be associated with the intervention effects (appendix Table S3 in (105)). However, this analysis was restricted to the ARK co-primary outcomes and did not consider how baseline (pre-intervention) antibiotic use, which varies widely across Trusts(106), may influence intervention effectiveness. Similarly, although there was a significant increase in antibiotic stop rates between baseline (median 12.7%, IQR 15.4-

21.4) and the first 12-weeks after implementation (median 16.2%, IQR 12.8-23.3) (Wilcoxon matched-pairs p=0.006), previous analyses did not consider potential effect modification from increasing prescription stop rates. Since sustained declines in antibiotic use were observed for select antibiotic categories (i.e. narrow-spectrum, oral, Access, Watch, quinolones) but not others (i.e. broad-spectrum, parenteral, Reserve, carbapenems), a better understanding of how contextual factors and intervention uptake could have influenced intervention effects across these secondary antibiotic endpoints is important for generalising the trial results moving forwards.

This chapter therefore provides an expanded assessment of potential effect modifiers and antibiotic outcomes to better understand how and why the introduction of the ARK intervention succeeded or failed to reduce hospital antibiotic use. The role of implementation fidelity, prescription stop rates, baseline antibiotic use, and other contextual factors are explored as possible barriers and facilitators of impact across the ARK co-primary and secondary antibiotic endpoints.

4.2 Methods

4.2.1 Effect modifiers considered

The 8 pre-specified site-level factors included in the Statistical Analysis Plan and the main trial paper were, as follows, together with the reasons they could have affected immediate and/or long-term effects of the ARK intervention:

- Randomisation block (categorical with pilot sites and 7 other blocks, versus block 1) – sites were recruited through professional networks and the Society for Acute Medicine, and recruitment continued after the pilot sites and earliest blocks were randomised to implement (as described in the protocol). As a result, it is possible sites volunteering early in the study differed in ways that impact intervention effects (e.g. in terms of clinical and stewardship practice). Also, the duration of follow-up is greatest among sites randomised to

implement early in the study, which may impact intervention effect estimates if some intervention components take time to fully adopt or are not sustained.

- Calendar period of randomised implementation (categorical based on quarter, versus January-March) – seasonal variation in antibiotic use and clinical outcomes reflect different NHS pressures during the year and may impact intervention effect estimates, as sites were randomised to introduce the intervention at different calendar dates.
- UK region (categorical, versus South of England) – reflecting possible regional variation in case-mix and the configuration of acute medical services and prescribing behaviour.
- Trust size (categorised into terciles based on beds available, versus small ≤550 beds) – staffing resources available to support implementation and stewardship may differ by site size, affecting implementation and hence effects.
- Prescribing systems (electronic versus paper-based) – reflecting differences in how the decision aid was embedded into existing prescribing systems, with ARK categories being added either to physical drug charts or incorporated into e-prescribing.
- Functional role of the local study lead (acute medicine, or pharmacist, versus microbiology or infectious diseases) – specialty culture may influence how the intervention is adopted by sites, including staff engagement and uptake of the different intervention components.
- Percentage of essential people (i.e. those considered by the Champion in each site as key to implementation, including opinion leaders in acute/general medicine) completing ARK training within 12 weeks of implementation (continuous, estimated per 10% increase) – reflecting how effectively local leaders were engaged with preparation for and implementation of the intervention.
- Total number of staff completing training by 12 weeks per 100 acute beds (continuous, estimated per increase of 10 people per 100 acute beds) – reflecting how effectively sites prepared for and implemented the online training in the first 12 weeks after implementation.

I chose to consider 12 additional factors in exploratory analyses, including 8 implementation fidelity criteria that were pre-defined in the protocol:

- Implementation fidelity criteria 1: Provision of a list of key essential people, by the implementation date (versus not achieved) – reflecting staff engagement and implementation preparation.
- Implementation fidelity criteria 2: Achieving at least 20 people per 100 acute beds having done the online learning, by 12 weeks after implementation (versus not achieved) – reflecting intervention uptake early in the implementation phase.
- Implementation fidelity criteria 3: Introduced ARK categories into the prescribing process, by implementation date (versus not achieved) – reflecting implementation preparation.
- Implementation fidelity criteria 4: Process in place for making patient leaflet available to acute medical patients, by implementation date (versus not achieved) – reflecting implementation preparation.
- Implementation fidelity criteria 5: Submission of baseline audit data, by implementation date (versus not achieved) – reflecting timely data submission and staff engagement.
- Implementation fidelity criteria 6: Process in place for ongoing audit and feedback, by implementation date (versus not achieved) – reflecting implementation preparation.
- Implementation fidelity criteria 7: Submission of post-implementation audit data, by week 4 (versus not achieved) – reflecting timely data submission and staff engagement.
- Implementation fidelity criteria 8: Submission of electronic patient data, by week 16 after implementation (versus not achieved) – predominantly reflecting Champion engagement since this was typically done through IT departments but required a lot of input.
- Total number of fidelity criteria achieved (out of 8, estimated per additional criteria) – providing an overall measure of implementation preparation, intervention uptake, staff engagement and timely submission of data.

- Implemented the decision aid as a “hard stop” to ensure initial prescriptions were stopped within 72h unless revised, or as a “soft stop” to reminding reviewers of the need to stop or revise prescriptions by 72h (versus implemented neither approach) – implementation of a prescription soft or hard stops may improve the timeliness of prescription stopping.
- Years between start of follow-up and implementation (continuous, estimated per 1 year later) – sites with greater pre-implementation follow-up duration will have less post-implementation follow-up, which may impact intervention effect estimates if some intervention components take time to fully adopt or are not sustained.
- Years between implementation and March 2020 (continuous, estimated per 1 year later) – sites randomised to implement early in the study will have greater follow-up time than sites that implemented later, and follow-up time may impact intervention effect estimates if some intervention components take time to fully adopt or are not sustained.

Six additional *post hoc* factors that I have considered in this chapter:

- Baseline prescription review stop rates (continuous, estimated per increase of 10%) – sites with higher pre-intervention stop rates may have limited scope to increase stop rates post-implementation, possibly limiting any stop rate-mediated impact of the intervention.
- Prescription stop rate at 72h clinical review increased between baseline and the first 12 weeks post-implementation (versus no) – increasing stop rates may reduce antibiotic DDDs/admission. Stop rates were defined as increasing if the lower 95% confidence bound of the post-implementation stop rate was above the baseline rate.
- Total number of essential people completing ARK training within 12 weeks of implementation (continuous, estimated per 100 increase) – reflecting how effectively sites prepared for and implemented the online training component of the intervention during the first 12 weeks after implementation.

- Baseline Access antibiotics as share of total antibiotic use – Access DDDs/admission varies widely across NHS Trusts(106), likely reflecting variation in local guidelines, stewardship, AMR rates and case mix. Since Access antibiotics are often used as first-line agents or for de-escalation, variation in this ratio may impact intervention effect estimates.
- Baseline Reserve antibiotics as share of total antibiotic use – Reserve DDDs/admission also varies widely across NHS Trusts(106)
- Baseline Access antibiotics as a share of Access + Watch antibiotics only – both Access and Watch antibiotics may be used as first-line agents, so greater baseline use of one may be expected to decrease use of the other, and variation in this ratio may impact intervention effect estimates.

4.2.2 Outcomes

To investigate the mechanisms by which the ARK intervention impacted each co-primary outcome (i.e. total antibiotic DDDs/admission and 30-day mortality), I derived estimates of effect modification for each of the 8 factors pre-specified in the Statistical Analysis Plan and a further 12 exploratory factors (as above) in the main paper(105). In this chapter, I have expanded this assessment by adding 6 further exploratory factors (as above) and estimated how all 26 factors are associated with immediate and sustained intervention effect estimates for the co-primary outcomes and each of the following secondary antibiotic endpoints, selected because there was at least some evidence of an overall intervention effect on them in the main trial:

- Narrow-spectrum and broad-spectrum agents
- The UKHSA's interpretations of the WHO's AWaRe antibiotics(42)
- Oral and parenteral antibiotics
- Quinolones and carbapenems

Although other secondary antibiotic outcomes were analysed in the trial, they were either not associated with the ARK intervention (e.g. piperacillin-tazobactam) or yielded very similar effect estimates to the co-primary outcome (e.g. total antibiotic DDDs per bed-day); given the number of outcomes I was already considering and the potential for false-positives due to multiple testing, I therefore did not consider these further.

The analysis of 30-day mortality included all 39 acute NHS Trusts. One site was excluded from the analysis of all antibiotic endpoints, as it rolled out an electronic prescribing system after implementation which led to a large reduction (-90%) in reported antibiotic DDDs, preventing valid estimation of intervention effects. A second hospital organisation reported no carbapenem use in 55/56 months and was also excluded from analyses of carbapenems. Estimates of effect modification for two factors (i.e. baseline antibiotic prescription stop rates and whether stop rates increased after baseline) relied on baseline audit data that was missing from 8 sites, so these analyses were limited to either 31 sites (30-day mortality endpoint), 30 sites (all antibiotic endpoints except carbapenems), or 29 sites (carbapenems).

4.2.3 Statistical methods

The objective was to identify whether specific characteristics of sites (e.g. whether processes for ongoing audit and feedback were introduced by implementation or not) were associated with greater or lesser impacts of the ARK intervention on the outcomes listed above. I explored this using univariable random-effects meta-regression which estimates the association between each of the 26 factors above and adjusted site-level intervention effect estimates from the ARK trial—measured as incidence rate ratios ((IRR) antibiotic endpoints) or odds ratios (30-day mortality). Sites were weighted by the precision (inverse of the within-hospital variance) of the effect estimate within that site (Stata command: metareg)(144) and analyses considered both immediate and sustained year-on-year intervention effect estimates separately. Associations are presented as percent changes in the intervention effect relative to each factor's reference group. Where there was evidence of effect

modification ($p<0.05$), predicted implementation effects were derived for each level of the effect modifier. Overall intervention effects are also presented for each study endpoint.

To understand whether the distribution of effect modification estimates differed from what might be expected assuming no true effect modification, I ordered the heterogeneity p-values for all antibiotic endpoints against the quantiles of a uniform distribution and visually assessed the plot for departures from the line of equality. If none of the factors acted as effect modifiers, the heterogeneity p-values should not skew away from the line of equality, while false null hypotheses would be expected to appear as deviations from this line (i.e. skewing towards smaller p-values)(236).

It is unclear whether characteristics of sites that are associated with greater immediate intervention effects are also associated with greater sustained year-on-year effects, or whether the magnitude and direction of short and long-term effects are driven by different effect modifiers. To assess this, I estimated the non-parametric Spearman rank coefficient (ρ) between estimates of effect modification on immediate intervention effects and estimates of effect modification on sustained intervention effects. The value of Spearman's ρ provides a measure of whether these pairs of ranked coefficients are consistently in the same order. Thus, ρ would be positive ($0 < \rho \leq 1$) if factors associated with greater immediate increases in the outcome were also associated with greater sustained increases, or negative ($-1 \leq \rho < 0$) if associated with lower sustained increases instead.

4.3 Results

The distribution of sites by potential effect modifiers is shown in **Table 4.2**; implementation fidelity was good overall (i.e. sites achieved a median 6/8 fidelity criteria, IQR 5-7), though 9 (23%) sites achieved ≤4/8 criteria.

Table 4.2 Twenty-six potential effect modifiers considered

Characteristic		NHS Trusts N=38 (%) (in analyses of antibiotic endpoints)
8 pre-specified factors		
Randomisation block[1]	Pilot sites (site 1-3)	2 (5%)
	Block 1 (site 4-8) (reference)	5 (13%)
	Block 2 (site 9-12)	4 (11%)
	Block 3 (site 13-18)	6 (16%)
	Block 4 (site 19-21)	3 (8%)
	Block 5 (site 22-27)	6 (16%)
	Block 6 (site 28-32)	5 (13%)
	Block 7 (site ≥33)	7 (18%)
Calendar period of randomised Implementation[2]	January-March	9 (24%)
	April-June	14 (37%)
	July-September	6 (16%)
	October-December	9 (24%)
UK region	South of England	12 (32%)
	North of England	10 (26%)
	Midlands / East of England	6 (16%)
	Northern Ireland	4 (11%)
	London	3 (8%)
	Wales	2 (5%)
	Scotland	1 (3%)
Trust size	Small (≤550 beds)	12 (32%)
	Medium (551-850 beds)	14 (37%)
	Large (>850 beds)	12 (32%)
Prescribing system	Paper	24 (63%)
	Electronic	14 (37%)
Functional role of the local study lead	Microbiology or infectious diseases	21 (55%)
	Acute medicine	7 (18%)
	Pharmacist	10 (26%)
Percentage of essential people completing training within 12 weeks of implementation	Median, IQR, range	76% (63-90) [0-100]
Total number completing training by 12 weeks per 100 acute beds	Median, IQR, range	24% (15-41) [1-102]
12 factors considered in exploratory analyses		
Implementation fidelity criteria		

Characteristic		NHS Trusts N=38 (%) (in analyses of antibiotic endpoints)
C1: Provision of a list of key essential people, by implementation date	Achieved	30 (79%)
C2: Achieving at least 20 people per 100 acute beds having done the online learning, by end of implementation phase (12 weeks)	Achieved	24 (63%)
C3: Introduced ARK categories into the prescribing process, by implementation date	Achieved	30 (79%)
C4: Process in place for making patient leaflet available to acute medical patients, by implementation date	Achieved	24 (63%)
C5: Submission of baseline audit data, by implementation date	Achieved	18 (47%)
C6: Process in place for ongoing audit and feedback, by implementation date	Achieved	36 (95%)
C7: Submission of post-implementation audit data, by week 4	Achieved	29 (76%)
C8: Submission of electronic patient data, by week 16 after implementation[3]	Achieved	21/36 (58%)
Total fidelity criteria achieved (out of 8)	Median, IQR, range	6 (5-7) [2-8]
How the decision aid was implemented	Hard stop	20 (53%)
	Soft stop	9 (24%)
	Neither	9 (24%)
Months between start of follow-up and implementation[2]	Median, IQR, range	33 (27-37) [20-41]
Months between implementation and March 2020[1]	Median, IQR, range	16 (10-20) [8-29]
6 *post hoc* factors considered in this chapter		
Audited 72h prescription stop rate at baseline, among reviewed prescriptions[4]	Median, IQR	13% (5-21)
72h prescription stop rate increased between baseline and 12 weeks post-implementation[4]	Yes	18/30 (60%)
Total number of essential people completing training within 12 weeks of implementation	Median, IQR, range	141 (84-203) [3-508]
Baseline Access antibiotics as a share of total antibiotic use[5]	Median, IQR, range	47% (37-58) [31-85]
Baseline Reserve antibiotics as a share of total antibiotic use[5]	Median, IQR, range	2.7% (1.6-3.6) [0.3-5.6]
Baseline Access antibiotics as a share of Access + Watch only[5]	Median, IQR, range	65% (53-76) [44-93]

[1] Main trial sites were sequentially randomised to implement in blocks of 6, with implementation staggered between sites by 1-2 weeks. However, since some sites withdrew after randomisation (but before

implementation), not all blocks contain 6 sites—see trial manuscript for further details(105).

[2] When deriving implementation effect estimates for antibiotic endpoints, which were measured monthly as DDDs/admission, a site was considered exposed to the intervention only if the implementation date was scheduled in the first half of the month (days 1-15 inclusive), or in a previous month. Thus, if the implementation date was scheduled in the second half of a month, that specific month was considered unexposed, and the subsequent month was modelled as the first month of implementation when estimating follow-up duration and implementation quarter.

[3] Not required for pilot sites (n=3); treated as achieved in meta-regression models.

[4] 'Baseline' refers to the 12 weeks pre-implementation. Since baseline prescription audit data that was missing in 8 sites, stop rates could only be assessed in 31 sites (30-day mortality endpoint), 30 sites (all antibiotic endpoints except carbapenems), or 29 sites (carbapenems) (see Methods).

[5] 'Baseline' refers to the 12 months pre-implementation.

4.3.1 Example: Whether sites introduced processes for ongoing audit and feedback by implementation

To illustrate how potential effect modifiers were analysed, I begin with the fidelity criteria with the strongest evidence for effect modification on the co-primary endpoint total antibiotic DDDs/admission, that is whether or not sites had introduced processes for ongoing audit and feedback by implementation (fidelity criteria 6).

First, I considered associations between this factor and immediate intervention effects, starting with total DDDs/admission followed by secondary endpoints (**Figure 4.1** panel A).

- The overall immediate effect of the ARK intervention on total DDDs/admission was -1.0% (95% CI: -4.0%, 2.1%, p=0.540); however in sites that introduced these processes (n=36) it was -1.8% (95% CI: -4.9%, 1.5%) and in those that did not (n=2) it was 17.8% (95% CI: 1.4, 36.9), a relative difference of -16.6% (95% CI: -28.5%, -2.8%) with moderate evidence of difference (heterogeneity-p=0.022).
- Considering all the other antibiotic secondary endpoints for this factor, I found sites that introduced these processes, versus did not, also had greater immediate reductions in DDDs/admission for oral (-24.4%, 95% CI: -36.5%, -10.0%, heterogeneity-p=0.003) and narrow-spectrum antibiotics (-16.9%, 95% CI: -29.9%, -1.4%, heterogeneity-p=0.035), with

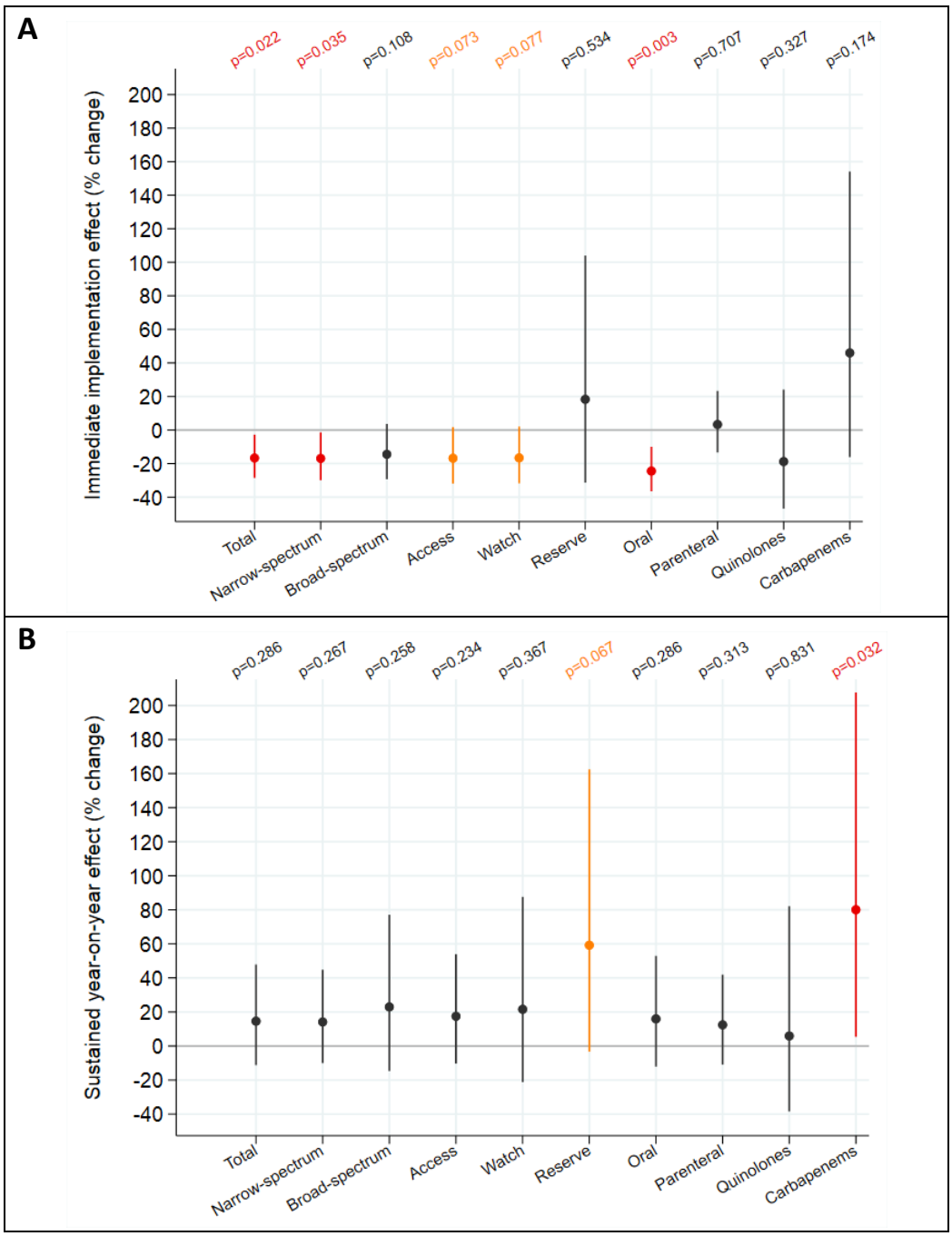
weaker evidence of difference for Access (-16.8% , 95% CI: -31.9%, 1.8%, heterogeneity-p=0.073) and Watch agents (-16.5%, 95% CI: -31.8%, 2.1%, heterogeneity-p=0.077).

Next, I consider sustained year-on-year implementation effects (**Figure 4.1** panel B).

- The overall year-on-year effect of the ARK intervention on total DDDs/admission was -4.8% (95% CI: -9.1%, -0.2%, p=0.042); however in sites that introduced these processes it was -4.1% (95% CI: -9.6%, 1.7%) and in those that had not, it was -16.3% (95% CI: -34.7%, 7.3%), a relative difference of 14.6% (95% CI: -11.2%, 47.9%) with no evidence of difference (heterogeneity-p=0.286).
- Introducing these processes was not associated with sustained year-on-year declines in intervention effects for any secondary antibiotic endpoint, though there was moderate evidence of greater sustained increases in carbapenems (80.1%, 95 CI: 5.4%, 207.7%, heterogeneity-p=0.032) and a similar trend for Reserve agents (59.2%, 95% CI: -3.4%, 162.5%, heterogeneity-p=0.067) in sites that did introduce them.

Where there is evidence of heterogeneity ($p<0.05$), subgroup implementation effects are presented for all antibiotic endpoints in appendix **Table 4.4**.

Figure 4.1 Association between sites that introduced processes for ongoing audit and feedback by implementation (versus did not) and site-level intervention effect estimates at implementation (A) and year-on-year (B) for all antibiotic endpoints



4.3.2 Antibiotic endpoints – immediate effects

Starting with immediate effects, among the 26 potential effect modifiers listed above there were 42 different comparisons versus the reference group for each endpoint. The magnitude and their direction of association with each antibiotic endpoint is shown in panel A of appendix **Figure 4.9–**

Figure 4.18. Eleven comparisons showed at least moderate evidence of heterogeneity (p<0.05) across one or more antibiotic endpoint (**Figure 4.2**), 6 of which showed evidence of association across ≥2 antibiotic endpoints, all for immediate declines in DDDs/admission, including:

- Sites that had a process in place for ongoing audit and feedback by implementation (described in the example above)
- Sites that submitted post-implementation audit data within 4 weeks following implementation (n=29), versus did not (n=9) (fidelity criteria 7)
    - The overall immediate intervention effect on total DDDs/admission was -1.0% (95% CI: -4.0%, 2.1%, p=0.540); however in sites that submitted this data it was -2.9% (95% CI: -6.3%, 0.7%) and in those that did not it was +5.9% (95% CI: -1.0%, 13.3%), a relative difference of -8.3% (95% CI: -15.1%, -1.0%) with moderate evidence of difference (heterogeneity-p=0.027) (appendix **Figure 4.9a**). However, this difference was not observed after adjusting for whether the site had a process in place for audit and feedback by implementation (heterogeneity-p=0.139).
    - Among secondary antibiotic endpoints, sites submitting this data also had greater immediate reductions in narrow-spectrum (-9.9%, 95% CI: -17.1%, -2.0%, heterogeneity-p=0.016) (appendix **Figure 4.10a**), oral (-9.2%, 95% CI: -17.3%, -0.5%, heterogeneity-p=0.040) (appendix **Figure 4.15a**), and Access antibiotics (-9.7%, 95% CI: -18.3%, -0.1%, heterogeneity-p=0.049) (appendix **Figure 4.12a**).
- Sites that increased prescription stop rates after baseline (n=18), vs sites that did not (n=12)
    - There was no evidence sites that increased stop rates after baseline had different immediate implementation effects on total DDDs/admission than those that did not (heterogeneity-p=0.505) (appendix **Figure 4.9a**).
    - Among secondary antibiotic endpoints, sites that increased stop rates had greater immediate declines in parenteral (-10.1%, 95% CI: -15.2%, -4.6%; p=0.001) (appendix

Figure 4.16a), carbapenem (-17.8%, 95% CI: -30.3%, -3.0%, p=0.022) (appendix **Figure 4.18a**), and Reserve antibiotics (-16.4%, 95% CI: -29.8%, -0.5%, p=0.044) (appendix **Figure 4.14a**).

- Medium-sized sites (n=14), versus small-sized sites (n=12)
    - Overall, the ARK intervention had no immediate effect on total DDDs/admission (-1.0%, 95% CI: -4.0%, 2.1%, p=0.540); however, in medium-sized sites total DDDs/admission reduced by -5.0% (95% CI: -10.0%, 0.2%) and in small-sized sites they increased by +2.5% (95% CI: -3.6%, 8.9%), a relative difference of -7.4% (95% CI: -14.6%, 0.5%) with weak evidence of difference (heterogeneity-p=0.064) (appendix **Figure 4.9a**).
    - Among secondary antibiotic endpoints, medium-sized sites also had greater immediate reductions in oral (-10.3%, 95% CI: -18.4%, -1.4%, heterogeneity-p=0.025) (appendix **Figure 4.15a**) and narrow-spectrum antibiotics (-9.2%, 95% CI: -17.0%, -0.7%, heterogeneity-p=0.035) (appendix **Figure 4.10a**).
- Sites in Wales (n=2), versus sites in the South of England (n=12)
    - There was no evidence sites in Wales (n=2) had different immediate implementation effects on total DDDs/admission (heterogeneity-p=0.116) compared with sites in the South of England (appendix **Figure 4.9a**).
    - Among secondary antibiotic endpoints, Welsh sites had greater immediate reductions in carbapenem (-39.8%, 95% CI: -56.6%, -16.6%, heterogeneity-p=0.003) (appendix **Figure 4.18a**) and Reserve antibiotics (-39.2%, 95% CI: -59.4%, -8.7%, heterogeneity-p=0.018) (appendix **Figure 4.14a**), with weak evidence of greater immediate reductions for quinolones (-33.5%, 95% CI: -56.0%, 0.6%, heterogeneity-p=0.053) (appendix **Figure 4.17a**), Access (-18.4%, 95% CI: -33.7%, 0.5%, heterogeneity-p=0.055) (appendix **Figure 4.12a**), and broad-spectrum agents (-16.3%, 95% CI: -31.5%, 2.3%, p=0.080) (appendix **Figure 4.11a**).

- Sites with a pharmacist as the study principal investigator (n=10), versus a specialty in microbiology or infectious diseases (n=21)
    - There was no evidence sites with pharmacist study leads had different immediate implementation effects on total DDDs/admission (heterogeneity-p=0.601) (appendix **Figure 4.9a**).
    - Among secondary antibiotic endpoints, sites with pharmacist study leads had greater immediate reductions for carbapenems (-19.3%, 95% CI: -33.2%, -2.4%, p=0.028) (appendix **Figure 4.18a**) and Reserve antibiotics (-21.0%, 95% CI: -35.7%, -2.9%, p=0.026) (appendix **Figure 4.14a**).

The other 5/11 factors had evidence of association with only one antibiotic endpoint, so may be more likely to reflect false-positive findings – specifically, (i) submission of baseline audit data by the implementation date (fidelity criteria 5), (ii) sites in London versus South of England, (iii) electronic versus paper prescribing, (iv) implemented the decision aid with a soft stop vs neither, and (v) baseline prescription stop rate. Also, the overall implementation fidelity score (per additional criteria achieved) was not associated with immediate implementation effects for any antibiotic endpoint (heterogeneity-p>0.4).

A visual assessment of the distribution of effect modification estimates showed slight skewness towards more significant p-values for some immediate implementation effects, relative to the quantiles of a uniform distribution (**Figure 4.3**), providing weak supporting evidence of heterogeneity, despite the number of tests performed, for greater immediate declines in prescribing following the intervention for (i) sites with a process in place for ongoing audit and feedback by the implementation date (versus not), (ii) sites that increased prescription stop rates after baseline (versus not), and (iii) sites in Wales (versus South of England).

Figure 4.2 Association between potential effect modifiers and the site-level immediate implementation effects for all antibiotic endpoints

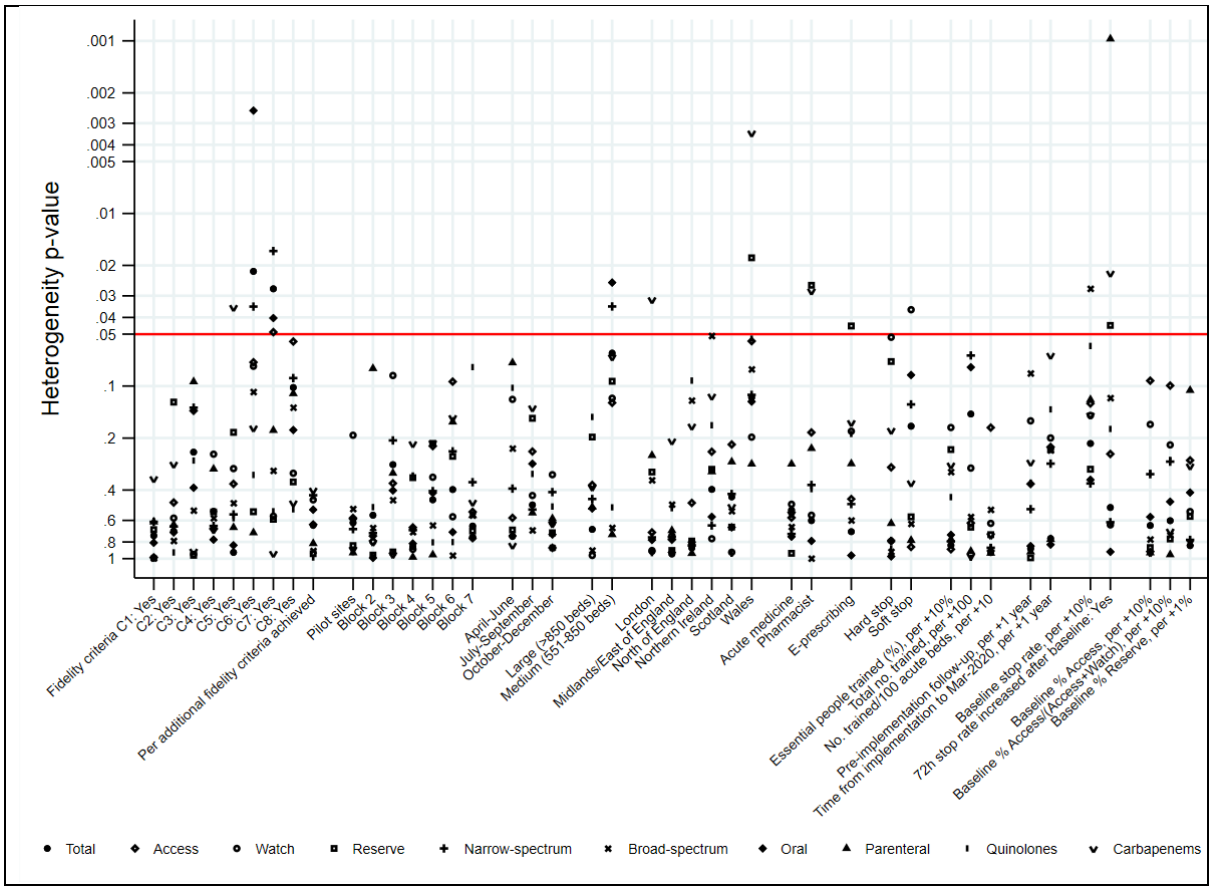
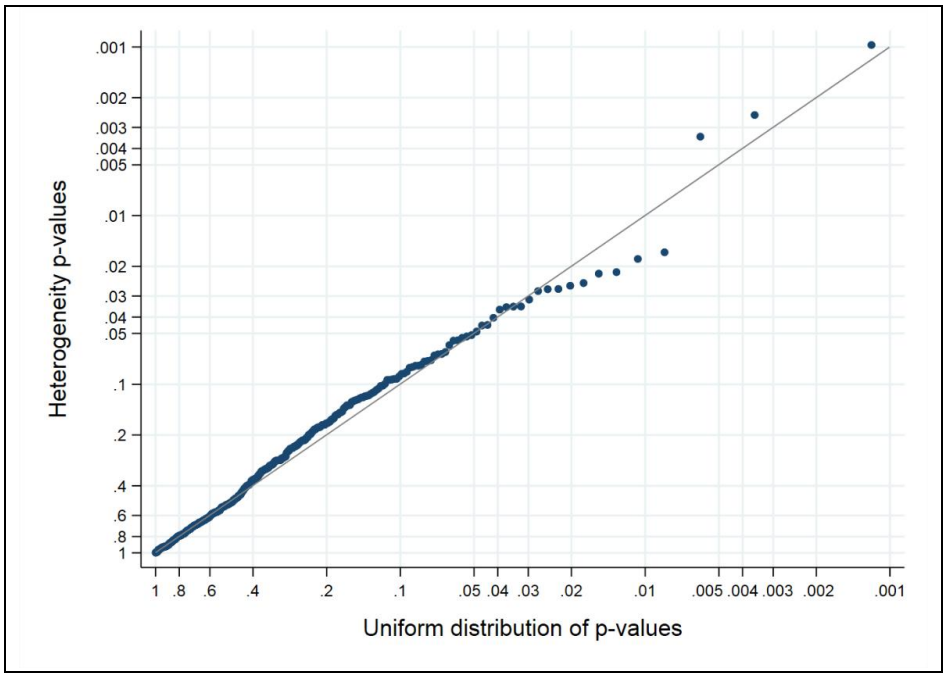


Figure 4.3 Estimates of effect modification for immediate intervention effects on all antibiotic endpoints ordered against the quantiles of a uniform distribution



4.3.3 Antibiotic endpoints – sustained year-on-year effects pre- vs post-implementation

Next, I consider effect modification on sustained year-on-year implementation effects. The magnitude and direction of association between all factors and each antibiotic endpoint is shown in panel B of appendix **Figure 4.9–Figure 4.18**. Among the 42 different comparisons vs reference group, 9 showed at least moderate evidence of heterogeneity ($p<0.05$) across one or more antibiotic endpoint, though only 4 showed evidence of association across ≥2 antibiotic endpoints, and 2 across 3 endpoints (**Figure 4.4**).

Among these 4 factors, one was associated with larger sustained declines in DDDs/admission:

- Sites that introduced the ARK decision aid categories into the prescribing process by the implementation date (n=30), versus did not (n=8) (fidelity criteria 3)
    - The overall sustained intervention effect on total DDDs/admission was -4.8% (95% CI: -9.1%, -0.2%, p=0.042); however in sites that introduced the decision aid it was -7.1% (95% CI: -12.7%, -1.1) and in those that did not it was +4.9% (95% CI: -7.3%, 18.8%), a relative difference of -11.5% (95% CI: -22.9%, 1.7%) with weak evidence of difference (heterogeneity-p=0.083) (appendix **Figure 4.9b**).
    - Among secondary antibiotic endpoints, sites that introduced the decision aid by the implementation also had greater sustained declines for Watch (-22.3%, 95% CI: -38.3%, -2.2%, heterogeneity-p=0.033) (appendix **Figure 4.13b**) and oral antibiotics (-13.9%, 95% CI: -25.9%, -0.1%, heterogeneity-p=0.049) (appendix **Figure 4.15b**).

The other 3/4 factors were associated with larger sustained increases in DDDs/admission, compared with the reference group.

- Medium-sized sites (n=14), versus small-sized sites (n=12)
    - The overall sustained intervention effect on total DDDs/admission was -4.8% (95% CI: -9.1%, -0.2%, p=0.042); however in medium-sized sites it was 1.6% (95% CI: -7.2%, 11.3%)

and in small sites it was -11.3% (95% CI: -19.9%, -1.9%), a relative increase of 14.6% (95% CI: 0.1%, 31.3%) with moderate evidence of difference (heterogeneity-p=0.049) (appendix **Figure 4.9b**).

  - Among secondary antibiotic endpoints, medium-sized sites had greater sustained increases in parenteral antibiotics of 14.3% (95% CI: 1.3%, 29.1%, heterogeneity-p=0.031), and there was weak evidence of greater sustained increases in broad-spectrum (19.4%, 95% CI: -1.8, 45.3%, heterogeneity-p=0.074) (appendix **Figure 4.11b**), narrow-spectrum (12.2%, 95% CI: -1.5%, 27.7%, heterogeneity-p=0.080) (appendix **Figure 4.10b**), and oral antibiotics (13.9%, 95% CI: -1.9%, 32.3%, heterogeneity-p=0.085) (appendix **Figure 4.15b**).

- Sites in Wales (n=2), versus sites in the South of England (n=12)
  - There was no evidence sites in Wales (n=2) had different sustained implementation effects on total DDDs/admission (heterogeneity-p=0.188) (appendix **Figure 4.9b**).
  - Among secondary antibiotic endpoints, Welsh sites had greater sustained increases in Watch (76.7%, 95% CI: 14.0%, 173.9%, heterogeneity-p=0.013) (appendix **Figure 4.13b**), broad-spectrum (50.9%, 95% CI: 5.0%, 116.9%, heterogeneity-p=0.028) (appendix **Figure 4.11b**), and quinolones (77.6%, 95% CI: 3.0%, 206.2%, heterogeneity-p=0.039) (appendix **Figure 4.17b**).
- Baseline Access antibiotics as a share of Access plus Watch (median 65.1%, IQR: 53.5-76.4), analysed per 10% increase
  - The overall sustained intervention effect on total DDDs/admission was -4.8% (95% CI: -9.1%, -0.2%, p=0.042). This effect increased from -9.0% (95% CI: -15.9%, -2.1%) at the 25th percentile of baseline % Access/(Access+Watch) (53.5%), to -4.6% (95% CI: -10.0, 0.7%) at the median (65.1%) and -0.4% (95% CI: -7.5%, 6.8%) at the 75th percentile (76.4%) (that is, the sustained intervention decrease was greater in sites with a the lowest Access antibiotics as a share of Access plus Watch), though there was only weak

evidence of difference (3.9% per 10+ increase, 95% CI: -0.3%, 8.2%, heterogeneity-p=0.071) (appendix **Figure 4.9b**).

- Among secondary antibiotic endpoints, a 10% higher baseline % Access/(Access+Watch) was associated with greater sustained increases in broad-spectrum (7.6%, 95% CI: 1.7%, 13.8%, heterogeneity-p=0.013) (appendix **Figure 4.11b**), oral (5.0%, 95% CI: 0.4%, 9.7%, heterogeneity-p=0.032) (appendix **Figure 4.15b**), and watch antibiotics (7.8%, 95% CI: 0.5%, 15.7%, heterogeneity-p=0.036) (appendix **Figure 4.13b**), with weak evidence of increases quinolones (8.6%, 95% CI: -0.9%, 18.9%, heterogeneity-p=0.077) (appendix **Figure 4.17b**).

The remaining 5/9 factors showed evidence of association with only one antibiotic endpoint, so may be more likely to reflect false-positive findings – specifically, (i) sites that had a process in place for ongoing audit and feedback by implementation (fidelity criteria 6), (ii) pilot sites versus randomisation block 1, (iii) Midlands/East of England versus South of England, (iv) local principal investigator specialises in acute medicine versus microbiology or infectious diseases, and (v) baseline Access use as a share of total antibiotics. Also, the overall implementation fidelity score (per additional criteria achieved) was not associated with sustained implementation effects for any antibiotic endpoint (heterogeneity-p>0.2).

Effect modification estimates did not show skewness towards more significant p-values for sustained implementation effects when compared with the quantiles of a uniform distribution (**Figure 4.5**), suggesting the modest evidence for heterogeneity (p=0.01-0.02) is consistent with what might be expected under the global null (i.e. assuming no true effect modification) given the number of tests performed. This may be because the ratio of true to non-true null hypotheses is relatively high or the analysis lacks the power required to detect weaker associations.

Figure 4.4 Association between potential effect modifiers and the site-level sustained year-on-year implementation effects for all antibiotic endpoints

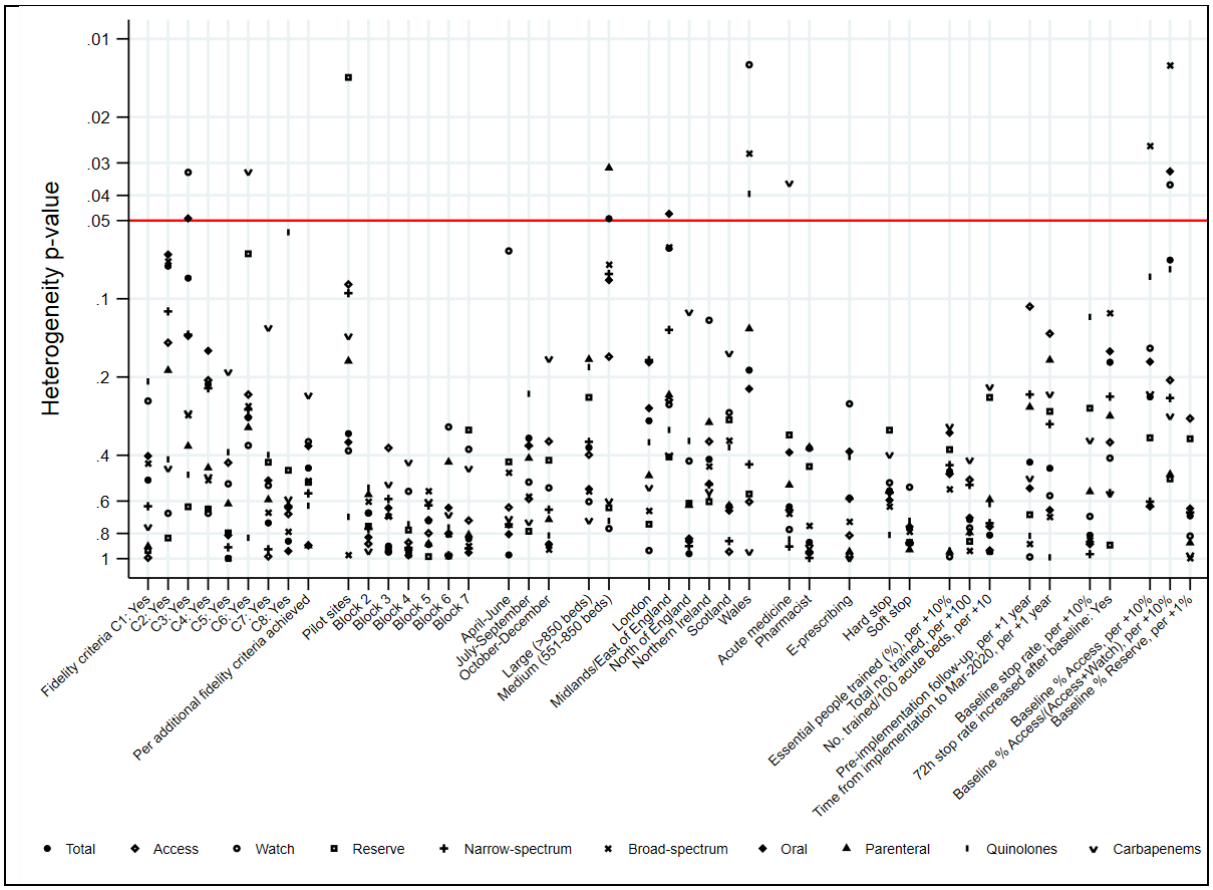
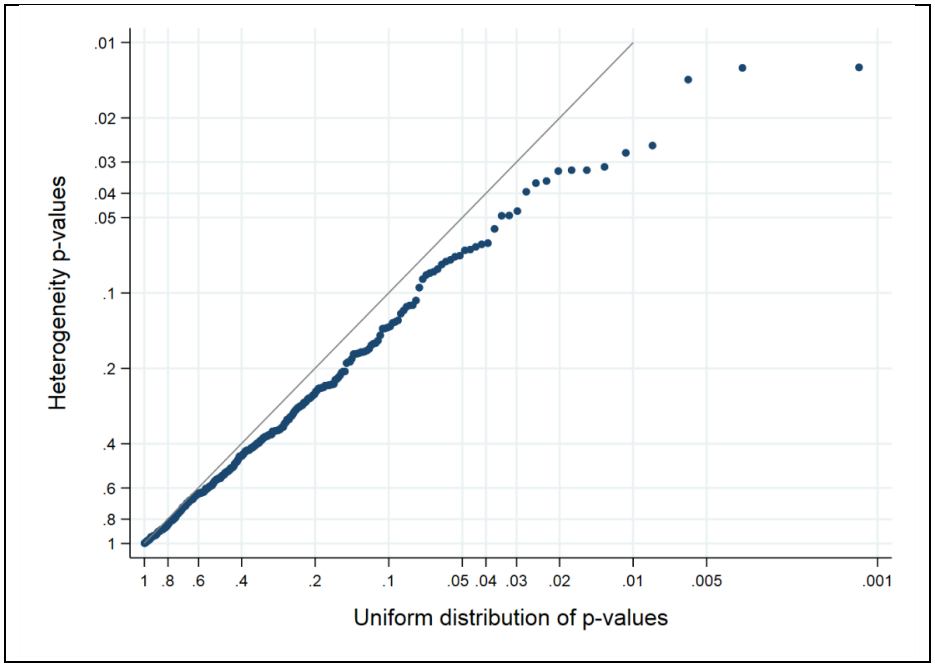


Figure 4.5 Estimates of effect modification for sustained year-on-year intervention effects for all antibiotic endpoints ordered against the quantiles of a uniform distribution



4.3.4 Antibiotic endpoints – comparing immediate and sustained implementation effect modifiers

I then considered how effect modification on immediate reductions in antibiotic use following the ARK intervention related to effect modification on sustained year-on-year reductions (summarised in **Table 4.3**), by comparing the Spearman rank correlation for the estimates of effect modification (shown for each outcome in panel C of appendix **Figure 4.9–Figure 4.18**).

For most (9/10) antibiotic outcomes there was no evidence that factors associated with the magnitude of immediate implementation effects were also associated with the magnitude of sustained year-on-year effects (Spearman's p>0.05 in **Figure 4.9C–Figure 4.17C**). There was moderate evidence of a correlation suggesting factors associated with greater increases in immediate intervention effects were also associated with greater increases in sustained effects for carbapenems (Spearman's ρ=0.356, p=0.021, appendix **Figure 4.18C**). However, no individual factor was associated (p<0.05) with both immediate and sustained implementation effects for a given antibiotic endpoint.

Factors in bold in **Table 4.3** retained evidence of heterogeneity (p<0.05) when included in a multivariable meta-regression model with all other factors that showed evidence of heterogeneity in univariate analyses. Those associated with greater reductions in DDDs/admission included:

- Welsh sites had greater immediate declines in carbapenem DDDs/admission (-34.9%, 95% CI: -56.4%, -3.0%, heterogeneity-p=0.036) than sites in the South of England
- Sites that introduced ARK categories into their prescribing process by implementation (fidelity criteria 3) had greater sustained declines in oral DDDs/admission (-15.3%, 95% CI: -26.7%, -2.0%, heterogeneity-p=0.027), as did sites in the Midlands/East of England versus the South of England (-17.1%, 95% CI: -30.1%, -1.7%, heterogeneity-p=0.033)

Table 4.3 Spearman rank correlation between estimates of effect modification on immediate and sustained intervention effects

Antibiotic endpoint (DDDs per admission)	**Spearman correlation between effect modification on immediate and sustained effects**	**Factors associated[1] with greater declines in both immediate and sustained effects**	**Factors associated[1] with greater declines on immediate but no impact on sustained effects**	**Factors associated[1] with greater declines on sustained but no impact on immediate effects**
Total	r=-0.063, p=0.690	None	Fidelity criteria 6 & 7 achieved	Small-sized sites (vs medium-size)
Narrow-spectrum	r=-0.152, p=0.337	None	Fidelity criteria 6 & 7 achieved, medium-sized sites (vs small size)	None
Broad-spectrum	r=-0.118, p=0.455	None	Lower baseline prescription stop rate	Lower baseline % Access/Total; lower baseline % Access/(Access + Watch); sites in South of England (vs Wales)
Access	r=-0.246, p=0.117	None	Fidelity criteria 7 achieved	None
Watch	r=-0.196, p=0.213	None	Did not implement a prescription hard/soft stop (vs implemented a soft stop)	Fidelity criteria 3 achieved; lower baseline % Access/(Access + Watch); **sites in South of England (vs Wales)**
Reserve	r=0.057, p=0.719	None	Pharmacist as local study lead (vs microbiology/infectious diseases); site in Wales (vs South of England); prescription stop rate increased after baseline; paper-based prescribing (vs electronic);	Pilot sites (vs randomisation block 1)
Oral	r=-0.030, p=0.849	None	Fidelity criteria 6 & 7 achieved, medium-sized sites (vs small-size)	**Fidelity criteria 3 achieved**; **sites in Midlands/East of England (vs South of England);** lower baseline % Access/(Access + Watch)

Antibiotic endpoint (DDDs per admission)	**Spearman correlation between effect modification on immediate and sustained effects**	**Factors associated[1] with greater declines in both immediate and sustained effects**	**Factors associated[1] with greater declines on immediate but no impact on sustained effects**	**Factors associated[1] with greater declines on sustained but no impact on immediate effects**
Parenteral	r=-0.266, p=0.089	None	Randomised implementation in January-March (vs April-June); increased prescription review stop rate after baseline	Small-sized sites (vs medium-size)
Quinolones	r=-0.024, p=0.878	None	None	Sites in South of England (vs Wales)
Carbapenems	r=0.356, **p=0.021**	None	Fidelity criteria 5 not achieved; Pharmacist as local study lead (vs microbiology / infectious diseases); sites in London and **Wales (vs South of England)**; increased prescription review stop rate after baseline	**Fidelity criteria 6 not achieved; Role of local study lead was microbiology / infectious diseases (vs acute medicine)**

[1] Heterogeneity p<0.05

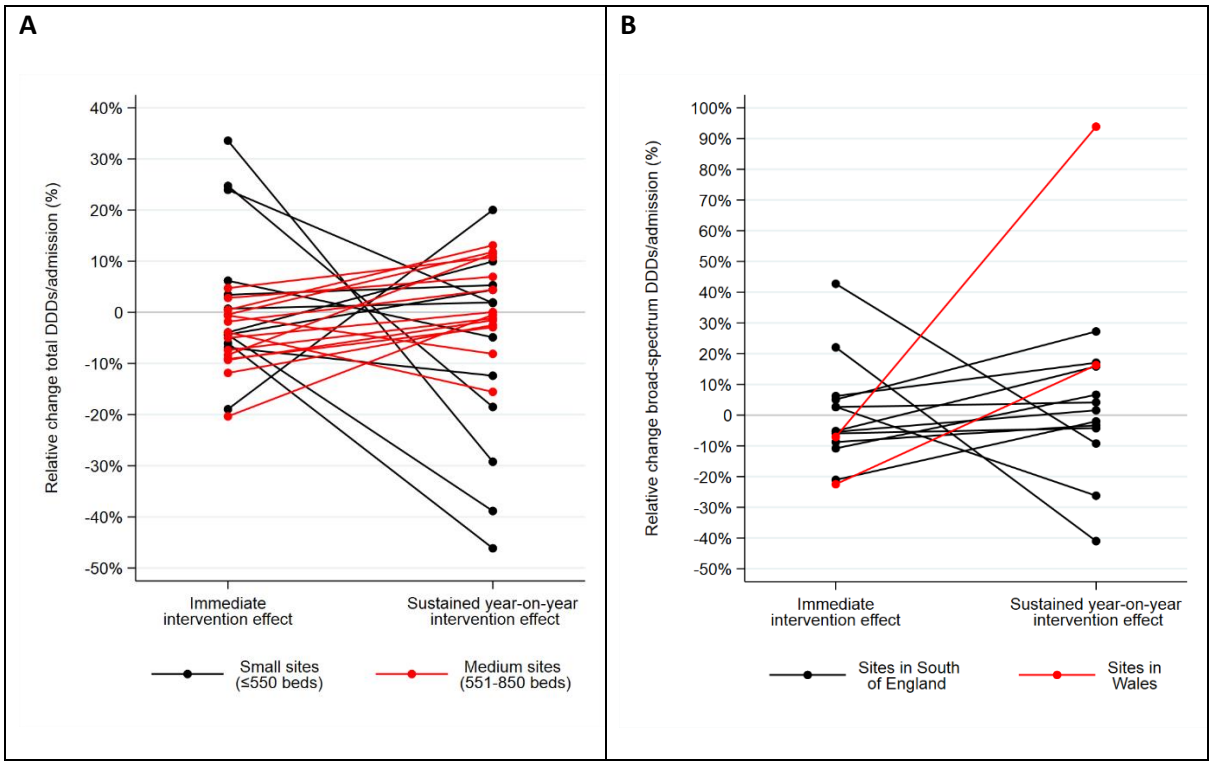
By contrast, those associated with greater increases in DDDs/admission included:

- Welsh sites had greater sustained increases in Watch DDDs/admission (71.8%, 95% CI: 8.2%, 172.9%, heterogeneity-p=0.023) than sites in the South of England
- Sites that had a process in place for ongoing audit and feedback by implementation (fidelity criteria 6) had greater sustained increases in carbapenem DDDs/admission (72.1%, 95% CI: 1.4%, 192.0%, heterogeneity-p=0.044), as did sites with an acute medicine specialist as their local study lead (vs microbiology/infectious diseases) (31.8%, 95% CI: 2.3%, 69.7%, heterogeneity-p=0.034)

Two factors were at least weakly associated ($p<0.1$) with larger initial reductions, followed by larger sustained increases in select antibiotic endpoints, suggesting a "push-pull" effect from possible overfitting in the site-level interrupted time series models. The first was medium-sized sites (vs small), where weak evidence of larger immediate declines in total DDDs/admission (heterogeneity-p=0.064) was followed by larger sustained increases (heterogeneity-p=0.049). The second was sites in Wales, where weak evidence of larger immediate declines in broad-spectrum DDDs/admission (heterogeneity-p=0.080) was followed by larger sustained increases (heterogeneity-p=0.028) (**Figure 4.6**), though the strength of association Welsh sites had with greater sustained increases was attenuated (heterogeneity-p=0.058) when site location was included in a multivariable meta-regression model containing all factors that with evidence of heterogeneity in univariate analyses.

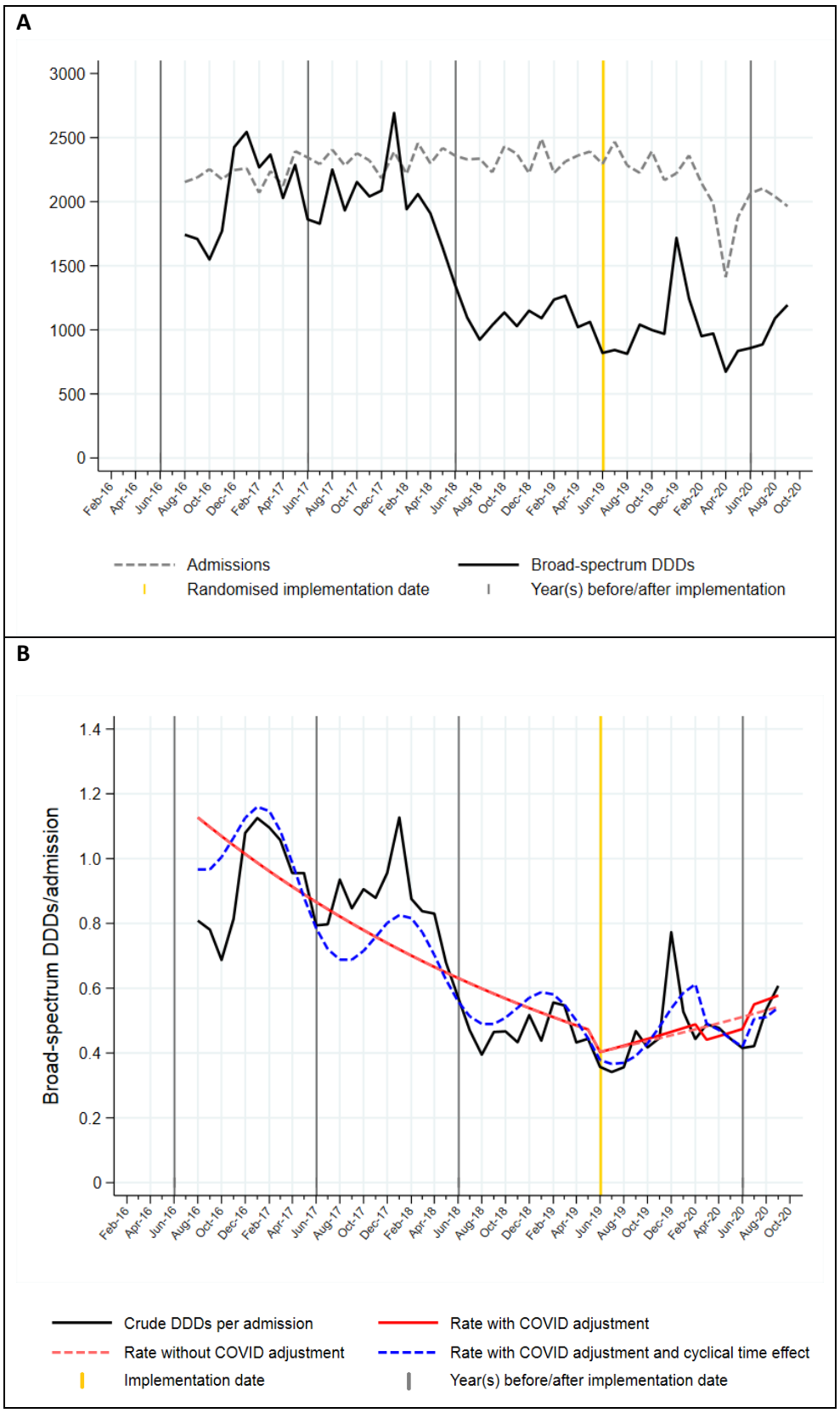
Figure 4.6B illustrates a limitation that occasionally can arise from the interrupted time series design used to estimate intervention effects. As shown, a site in Wales was associated with a large sustained increase in antibiotic broad-spectrum DDDs/admission of 93.9% per year (95% CI: 53.5%, 144.9%), but further investigation of monthly DDD and admission trends before and after implementation suggests this large increase may be misleading. Although monthly admissions in this site were relatively constant before implementation on 3 June 2019, monthly broad-spectrum DDDs

Figure 4.6 Immediate and sustained implementation effect estimates, by site size for total DDDs/admission (A) and by location for broad-spectrum DDDs per admission (B)



showed large (50%) declines between 2017 and 2018 (**Figure 4.7A**). This decrease resulted in an estimated pre-implementation trend showing broad-spectrum DDDs/admission declining by -28.8% per year (95% CI: -33.7%, -23.6%), even though DDD use had stabilised in the year before implementation (i.e. from August 2018 onwards). As one of the last sites to implement in the stepped wedge design, follow-up after implementation was just 16 months and included the onset the COVID-19 pandemic, which likely contributed (due to imperfect adjustment for COVID and seasonal effects, as per Discussion) to estimated post-implementation increases in broad-spectrum DDDs/admission of 38.0% per year (95% CI: 10.6%, 72.2%). The stark difference in the direction of pre- and post-implementation trends is then attributed to the sustained implementation effect in the interrupted time series analysis (**Figure 4.7B**), which given the limited number of sites in Wales (n=2) is likely driving the effect modification in opposite directions (i.e. for immediate and sustained intervention effects).

Figure 4.7 Monthly broad-spectrum DDDs and admissions at a site in Wales (A), with pre- and post-implementation trends in broad-spectrum DDDs/admission as estimated in an interrupted time series analysis (B)



4.3.5 30-day mortality – immediate effects

The magnitude and direction of association between all factors and intervention effect estimates at implementation for 30-day mortality is in appendix **Figure 4.19A**. Among the 42 different comparisons vs reference group, just 2 showed moderate evidence of heterogeneity (p<0.05):

- Sites that submitted electronic patient data within 16 weeks of implementation (fidelity criteria 8) (n=24), versus those that did not (n=15), had an immediate increase in mortality of 7.9% (95% CI: 1.4%, 14.7%, heterogeneity-p=0.018).
- Every 10% higher baseline Access antibiotics as a share of Access + Watch was associated with a greater immediate decrease in 30-day mortality (by -2.9%, 95% CI: -5.0%, -0.7%, heterogeneity-p=0.011). A similar trend was observed for each 10% higher baseline Access as a share of total DDDs/admission (-2.3%, 95% CI: -4.9%, 0.3%, heterogeneity-p=0.082).

These estimates showed a skewness towards more significant p-values when compared with the quantiles of a uniform distribution (**Figure 4.8A**), suggesting there remains moderate evidence of heterogeneity, despite the number of tests performed.

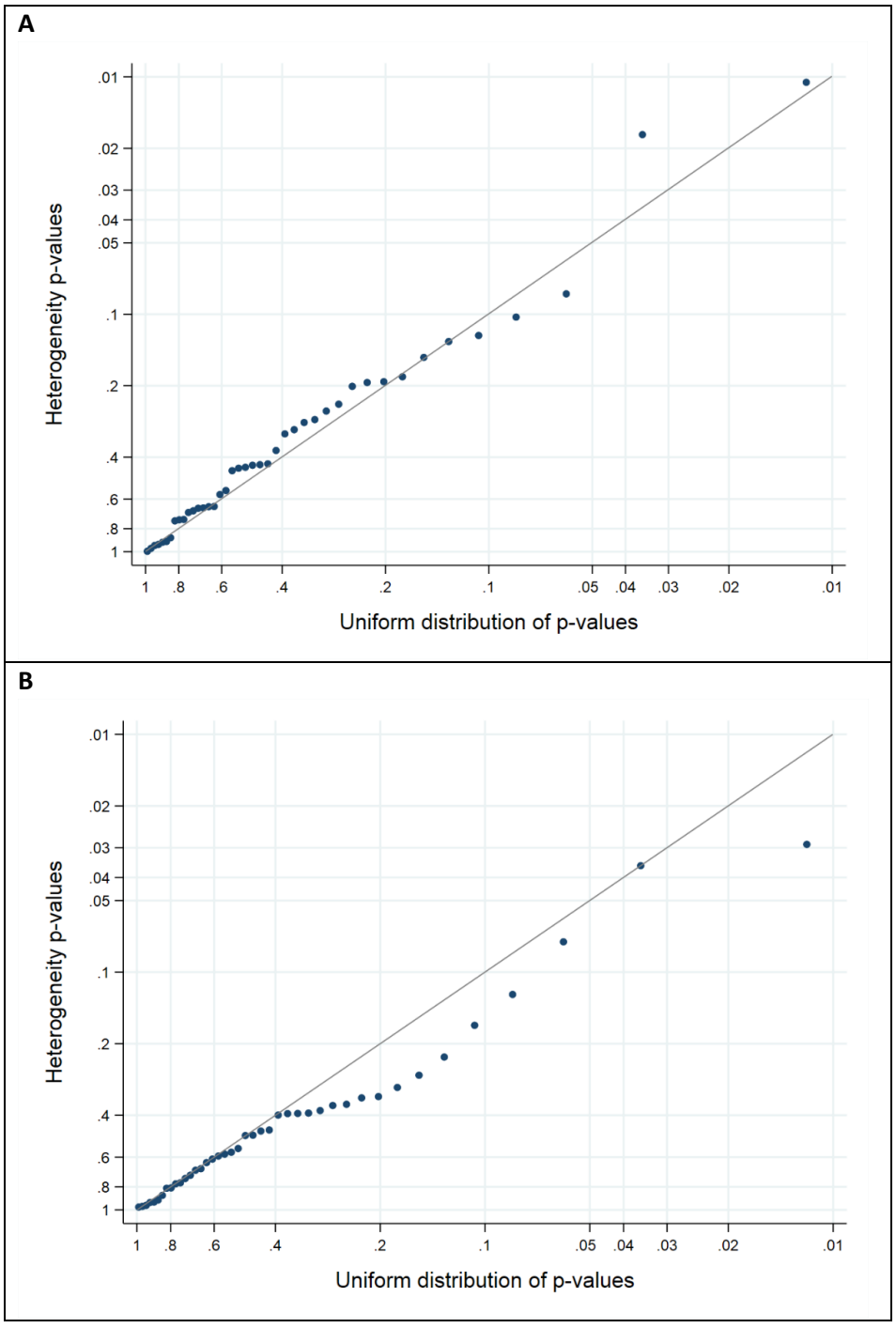
4.3.6 30-day mortality – sustained effects

Just 2 factors showed moderate evidence of heterogeneity (p<0.05) on sustained implementation effects (appendix **Figure 4.19B**), both for declines in 30-day mortality:

- Sites that introduced a prescription hard stop (n=21), versus sites that did not implement a soft or hard stop (n=9), had greater sustained decreases in 30-day mortality by -8.2% (95% CI: -15.0%, -0.9%, heterogeneity-p=0.029)
- Sites that were randomised to implement in July-September (n=6), versus January-March (n=10), had greater sustained decreases in 30-day mortality by -10.3% (95% CI: -19.0%, -0.8%, heterogeneity-p=0.036)

However, these estimates did not show skewness towards more significant p-values when compared with the quantiles of a uniform distribution (**Figure 4.8B**), suggesting the results are consistent with what might be expected under the global null given the number of tests performed.

Figure 4.8 Estimates of effect modification for immediate (A) and sustained year-on-year intervention effects (B) on 30-day mortality ordered against the quantiles of a uniform distribution



4.3.7 30-day mortality – comparing immediate and sustained implementation effect modifiers

Factors associated with greater immediate 30-day mortality declines were also associated with greater year-on-year increases (Spearman's rho: -0.401, p=0.009) (appendix **Figure 4.19C**), although no individual factor was associated with both immediate and sustained mortality effects.

4.4 Discussion

The ARK intervention achieved sustained, safe reductions in antibiotic use among adult acute/general medical admissions, including for total DDDs/admission, narrow-spectrum, Access, Watch, oral antibiotics, and quinolones, but heterogeneity in site-level effect estimates was moderate to high. In this chapter, I have explored possible sources of this between-site variation.

Overall fidelity of implementation, as assessed by a score summing eight individual intervention components, was not associated with changes in antibiotic use following introduction of the ARK intervention. However individual intervention components were associated with antibiotic reductions across multiple endpoints, suggesting overall fidelity may be averaging across elements, only some of which are key drivers of change. Of the eight pre-specified fidelity criteria considered, there was some evidence four were associated with the magnitude of the site-level intervention effect estimates, though none were associated with both immediate and sustained implementation effects. This suggests the drivers of short and long-term intervention effects may differ, perhaps because acceptance and adoption of some intervention components takes time, while others may not be sustained. For example, uptake of the online learning and integration of the decision aid categories into each hospital's prescribing process are likely to take time, which may explain why their association with declines in oral and total antibiotic use were observed year-on-year, rather than immediately. Conversely, sites that already had a process in place for ongoing audit and feedback at implementation, and those sharing post-implementation audit data by 4 weeks may have prepared most effectively in the weeks leading up to the randomised implementation date,

contributing to the greater immediate declines in total, narrow-spectrum, and oral antibiotics observed in these sites, but not sustained effects. Attempts to score other AMS programmes based on quality indicators related to structure and process have found wide variation between acute hospitals in England, though there is some evidence that hospitals with high process scores (i.e. demonstrating greater adherence to AMS process quality indicators) are associated with lower Reserve prescribing, while sites with low structure and process scores are associated higher total antimicrobial prescribing(214).

Surveys of antimicrobial stewardship leads at acute hospitals in England have expressed significant concerns about the safety of achieving antimicrobial reduction targets, presumably due to fears that inappropriate early stopping may lead to under-treatment and patient harm(46). Although prescription hard stops have been implicated in possible treatment delays(237), there is evidence hard stops can be implemented safely(238) and this analysis found hard stops (versus sites that implemented neither soft or hard stops) were associated with greater sustained decreases in 30-day mortality. This may suggest the perceived risk of under-treatment motivated sites implementing hard stops to engage in more careful prescription reviews and patient monitoring. Given the lack of evidence supporting overall changes in mortality risk following ARK implementation, and concerns about my ability to adjust for confounding from the COVID-19 pandemic (discussed below), the negative correlation between factors associated with immediate and sustained changes in 30-day mortality (appendix **Figure 4.19C**) may reflect overfitting in the meta-regression models, whereby factors estimated to be associated with greater immediate risk reductions are also estimated to be associated with smaller sustained reductions ("push-pull" in the interrupted time series models, as seen for medium-sized sites and sites in Wales for antibiotic endpoints in **Figure 4.6**). I also found no evidence that sites implementing the decision aid with a 'hard' or 'soft' stop had greater sustained changes in antibiotic use than sites implementing neither. While this finding aligns with several reports showing no overall reduction in antibiotic use after the introduction of prescription hard stops(239–241), there is some evidence hard stops can reduce overall treatment duration(238),

inappropriate therapy(239), treatment costs(240), and use of select agents(240,242), and may support improvements in antibiotic treatment plan documentation and de-escalation rates(241). As a result, prescription hard stops, also known as antibiotic 'time-outs' in the United States, have been recommended as supplemental, but not a substitute for prescription audit and feedback(243).

Higher baseline prescription stop rates might be expected to limit the scope for the stop rate to increase post-intervention, yet there was no evidence that baseline stop rates were associated with greater intervention effects on sustained changes in antibiotic use. Similarly, although it was hypothesised the impact of the intervention might be mediated by its impact on stop rates, greater increases in stop rates were only associated with immediate, rather than sustained, declines in DDDs/admission for carbapenem, Reserve, and parenteral antibiotics. Given the overall sustained declines noted for several antibiotic endpoints (appendix **Table 4.4**) and the fact that the decision aid was designed to support reductions in treatment duration, this finding is arguably surprising. However, as the audit data used to determine stop rates was only available from 31/39 sites (see Methods), this reduced the already limited power to detect associations, and the data was collected after randomisation and intervention planning had begun, so it is possible baseline estimates were already affected by the intervention. Similarly, since stop rates after implementation were estimated using data collected during the initial 12-week implementation phase, longer-term changes in stop rates are unknown, potentially obscuring the number of sites increasing stop rates during the trial (or even decreasing them again after an initial increase) and attenuating the impact of this on long-term trends in antibiotic use.

Of the remaining effect modifiers considered, there was some evidence geographic location was associated with the magnitude of year-on-year changes in antibiotic use, as sites in Wales, versus the South of England, were associated with sustained increases in quinolones, broad-spectrum, and Watch antibiotics. However, with only 2 sites in Wales the estimates are imprecise, and this effect may be partially explained by the fact that Welsh sites had especially high measures of baseline

Access antibiotic use as a share Access + Watch (79% and 86% respectively, versus median 65%, IQR: 53-76%, range: 44-93%), which had weak to moderate evidence for greater sustained reductions in the same antibiotic endpoints in sites with lower baseline percentages. Supporting this inference, evidence of heterogeneity of the Welsh sites attenuated ($p<0.05$) after adjusting for Access/(Access+Watch) in models of sustained intervention effects on broad-spectrum DDDs/admission (**Table 4.3**).

This chapter has several of important limitations. First, despite the large size of the trial, which randomised a quarter of the UK's acute NHS Trusts and included over 7 million acute/general medical admissions in analyses of mortality, the effect modifiers investigated through meta-regression rely on effect estimates from just 39 sites, providing limited power to detect weaker associations. Also, most (18/26) of the potential effect modifiers were not prespecified in the Statistical Analysis Plan, and inferences drawn from these analyses should therefore be considered hypothesis generating(139).

Second, false-positive associations (type I errors) will rise as the number of potential effect modifiers and outcomes assessed increases(244), regardless of whether the factors are prespecified or not(245). This occurs because, although the type I error rate in any single test is bounded by α (pre-specified as $\alpha=0.05$), the rejected hypothesis tests will be a selected subset in which type I errors are over-represented. The size of the family of tests over which this selection takes place will determine the probability of at least one false-positive (familywise error rate (FWER)), with the post-study probability any association is true being dependent on α, study power, bias, and the pre-study ratio of true to non-true null hypotheses(246). Although the latter ratio is unknown, the likelihood of true null hypotheses will be lower in hypothesis-driven research, and higher in hypothesis-generating research or settings with very large numbers hypothesis tests relative to true relationships, as in genome-wide association studies(247). With family defined as the set of hypotheses related to a single intervention effect estimate, the FWER in any given panel (i.e. panels A and B, in each of

appendix **Figure 4.9–Figure 4.19**) will be high, equalling $1-(1-\alpha)^n$, with n=42 and a=0.05, or 88.4% compared with the nominal 5%. With α=0.1, this probability increases to 98.8%, and with α=0.01 it declines to 34.4%. Only three comparisons had reasonable evidence of heterogeneity given the number of tests performed (p<0.01) (**Figure 4.2** and **Figure 4.3**), all for greater immediate declines in prescribing following the intervention in (i) sites with a process in place for ongoing audit and feedback by the implementation date (versus not), (ii) sites that increased prescription stop rates after baseline (versus not), and (iii) sites in Wales (versus South of England). For sustained year-on-year trends, the strongest effects had at most modest evidence of heterogeneity (p=0.01-0.02) (**Figure 4.4**), but these were consistent with what might be expected under the global null given the number of tests performed (**Figure 4.5**) – specifically, (i) pilot sites (versus randomisation block 1), (ii) sites in Wales (versus South of England), and (iii) sites with higher baseline percent Access antibiotic use as a share of Access+Watch.

However, these calculations assume the null hypothesis of each test *n* is true and are therefore conservative when the test statistics are not independent. Given the potential for clustering among individual fidelity criteria and correlations among effect modifiers and outcomes (e.g. narrow-spectrum and Access antibiotics), this rate is likely a conservative worst-case scenario. Also, the distribution of heterogeneity p-values for estimates of effect modification showed skewness towards less significant p-values for sustained implementation effects (**Figure 4.5**), suggesting the ratio of true to non-true null hypotheses is likely high and inferences from estimates in any single panel must be treated with great caution, which is why in this chapter, I have focused primarily on effect modifiers that showed significant associations across multiple antibiotic endpoints, as this would be less likely to occur by chance alone, assuming no true association.

To reduce the probability of type I errors, I could have used an even lower (though still arbitrary) α threshold(248) or made p-value adjustments using procedures that control either the FWER, the expected proportion of false positives among the rejected hypotheses (false discovery rate (FDR)), or

the probability that the proportion of false positives exceeds some threshold (false discovery proportion (FDP)). A simple example of strong FWER control is the Bonferroni correction, in which null hypotheses are rejected if $p<\alpha/n$ (i.e. in a given panel this would be 0.05/42=0.001), ensuring the FWER is no more than α when the global null holds, which is valid under any dependence structure of the underlying p-values, and for any combination of true and false hypotheses(247)(249). However, as indicated above, this is extremely conservative when tests are not independent. Other more powerful options for FWER control include the Holm procedure(250) or methods that bound the probability of making ≥*k* false rejections (*k*-FWER), rather than a single false rejection(251,252). However, while strict FWER control is appropriate and recommended by regulatory agencies in confirmatory contexts (e.g. efficacy trials for regulatory approval or labelling)(253,254), these methods are often inappropriately conservative in exploratory analyses(254) and settings with hypothesis-driven testing or correlations between p-values or outcomes, as is likely here, resulting in low power and increased false-negative (type II) errors(255,256). For this reason, methods of FWER control have not been applied here. Similarly, methods for FDR control have not been employed as they too may be inappropriately conservative in this context. For example, the popular Benjamini-Hochberg procedure for FDR control may improve power over strict FWER control methods(257), but it employs a sequential Bonferroni-type procedure, whereby the rank order of each p-value within a family of tests is divided by the number of tests within the family and then multiplied by the chosen FDR (i.e. for the lowest p-value in a given panel, this would be (1/42)*0.05=0.001, or for the second lowest p-value (2/42)*0.05=0.002, etc), and the product is then used, via a step-down procedure, to determine which test statistics provide evidence of effect modification. However, as unadjusted p-values in all panels were $p>0.001$ (with one exception in **Figure 4.16a**), setting α at 0.05 would ensure only one of the test statistics would be rejected, limiting the potential of this exploratory analyses to identify possibly relevant effect modifiers.

There is widespread interest in the potential of electronic health records for data collection in clinical trials(258,259) and AMS research(260,261), yet several limitations in this analysis stem from the routinely collected electronic data used to estimate the intervention effects in the meta-regression models. For example, the ARK intervention targeted antibiotic prescribers in acute/general medicine, but patients under their care needed to be identified from electronic health records using consultant specialty codes. Since coding practices may vary across hospital organisations, a broad list of specialty codes was chosen to reflect the specialties that most frequently admit medical inpatients (appendix Figure S1 in (105)). While this definition is likely highly sensitive, specificity will vary by site depending on local coding practices, and as a result the study population will include patients whose care providers were not exposed to the ARK intervention (i.e. in whom exposure status is misclassified), diluting effect estimates for clinical outcomes in some sites more than others. The broad definition of medical inpatients may also contribute to some outcome misclassification, as antibiotic endpoints were measured in monthly DDDs/admission, with the denominator including patients whose antibiotic consumption may not be reflected in the numerator (and vice-versa). This is because most (35/39) sites lacked the electronic prescribing systems and informatics resources required to provide patient-level antibiotic data, so the numerator was measured at an organisation-level based on overall dispensing to general medical wards. While this was appropriate for an intervention designed to be implemented at an organisational-level, it meant patient-level mediators of impact could not be assessed. The dilution of intervention effects due to misclassification errors has been flagged as one possible explanation for why trials using routine data often show smaller treatment benefits than trials using traditional patient-level data collection(262), and it may limit the ability of the meta-regression models to detect valid factor-outcome associations.

Intervention effect estimates may also be biased by residual time-dependent confounding from the COVID-19 pandemic, which began after implementation in all sites, affecting hospital case-mix, mortality rates, and antibiotic DDDs/admission. Although I attempted to adjust for COVID-19 by

including a binary factor for March-June 2020 in the interrupted time series analyses, this adjustment is likely to be imperfect, as intervention effects showed slight sustained increases in mortality that disappeared in sensitivity analyses excluding data from March 2020 onwards (appendix Table S3 in (105)). Similarly, time-dependent confounding may exist due to changes stewardship, clinical practice, or AMR rates (which were unknown but may vary by site and over time, remaining unbalanced despite the randomised stepped wedge design). Since participating sites may not be representative of all UK secondary care organisations, it is also unclear if associations identified in the meta-regression models have good external validity, though the heterogeneity of site-level effect estimates and variation in implementation fidelity suggests there was substantial diversity among the participating sites, which accounted for nearly a quarter of all acute hospitals in the UK.

Finally, although the ARK intervention evaluation and design process avoided many of the pitfalls that often undermine the translation of AMS research into practice (i.e. non-experimental, underpowered, or single-centred study designs, without clinical endpoints)(47,215–219), the multifaceted nature of the ARK intervention means there may be complex interactions between several mediators of effect, not all of which can be investigated quantitatively or detected with sufficient precision given the limited power. Interpersonal factors, professional hierarchy, perceptions of risk and specialty culture all contribute to hospital prescribing behaviour and the reluctance prescribers might feel when deciding whether to change prescribing decisions made by their peers(50–54). For example, specialty culture is known to shape the tendency to seek input from colleagues, openness to feedback, perceptions of accountability for the outcomes of prescription choices, and willingness to delegate decision-making to junior staff and other specialties, suggesting AMS interventions may need to be carefully optimised with end-users and specialty setting in mind(51). Hospital prescribers also report prescription review decisions are influenced by perceptions of external pressure to continue(52) and understandings of who is responsible for review decisions, suggesting interventions to improve stop rates must promote the

distribution of responsibility, with feedback for junior trainees to facilitate learning(263). While the ARK intervention included online training and audit tools to support prescribers with light-touch feedback, the complex interaction between the different intervention components and these contextual factors may be better explored through complimentary qualitative and mixed-methods approaches. For this reason, ARK collaborators conducted interviews with study leads and focus groups with front-line staff at sites to understand their experiences implementing the intervention, perceptions of barriers and facilitators of impact, and how the online training materials were used. Once analysed, this information could be usefully triangulated with the results of this chapter to compliment our understanding of mediators of impact and long-term sustainability, so the ARK intervention may be safely adapted to support antimicrobial reduction efforts in other acute care settings.

4.5 Appendix

Figure 4.9 Total DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

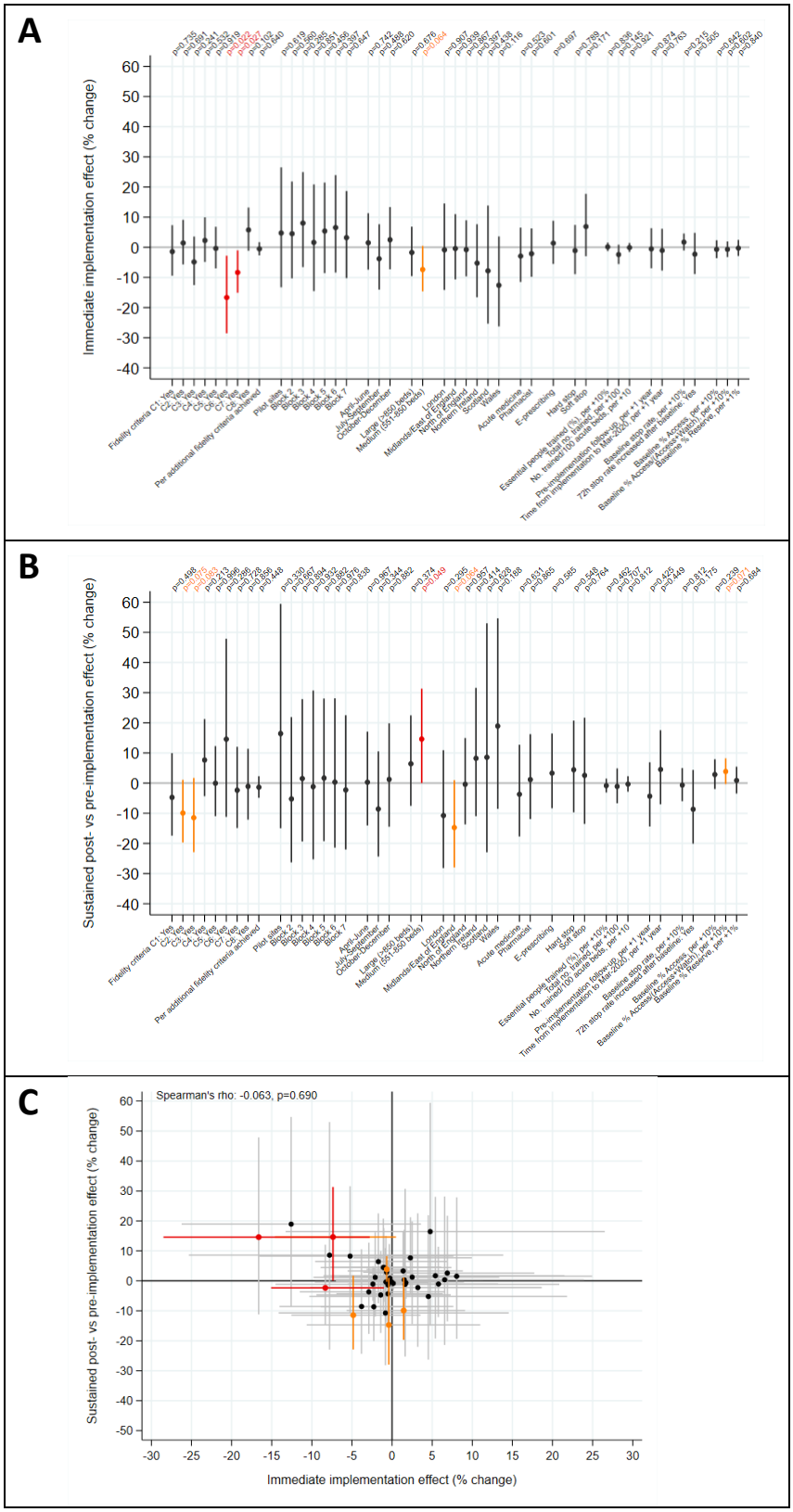


Figure 4.10 Narrow-spectrum DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

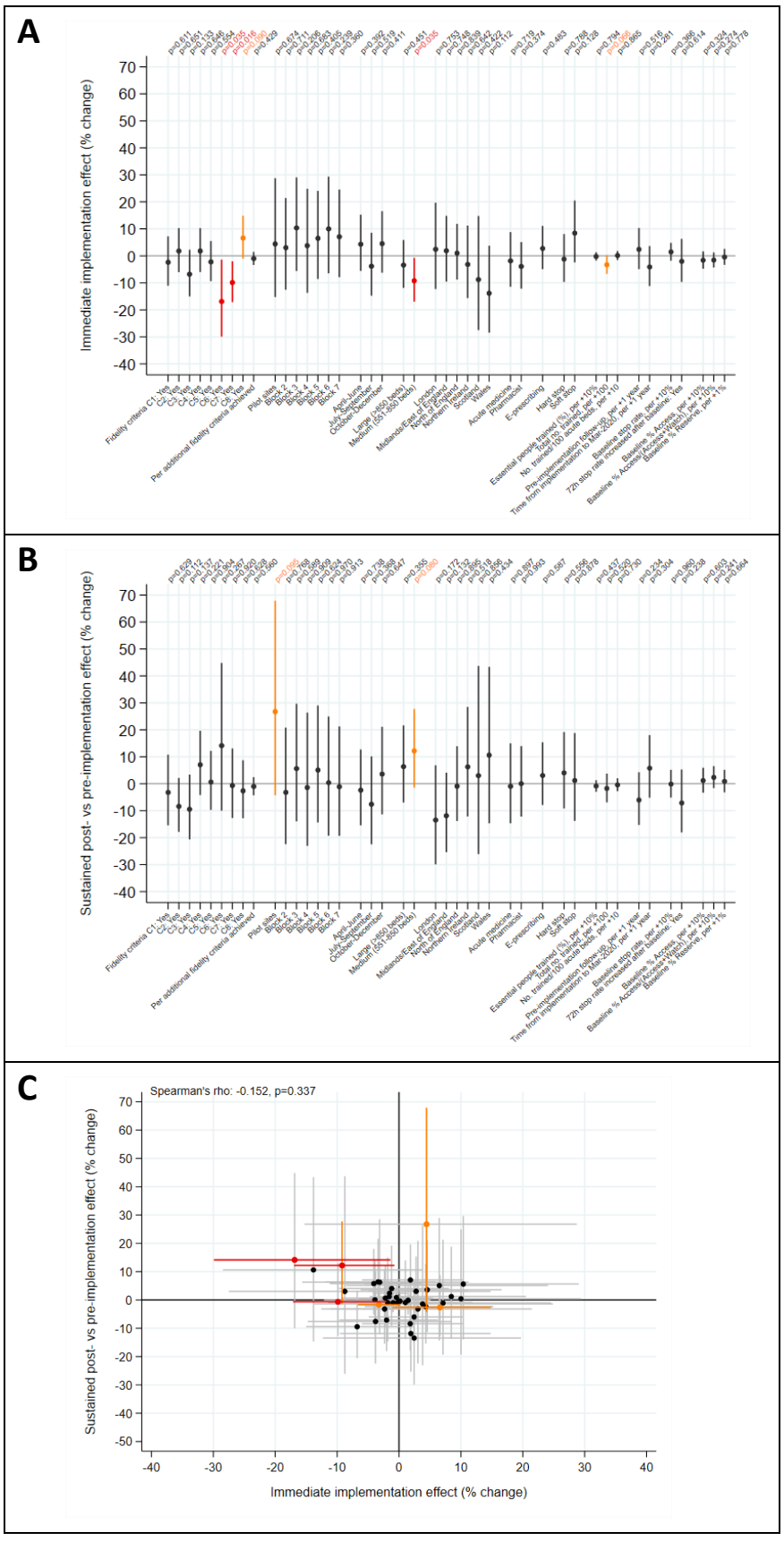


Figure 4.11 Broad-spectrum DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

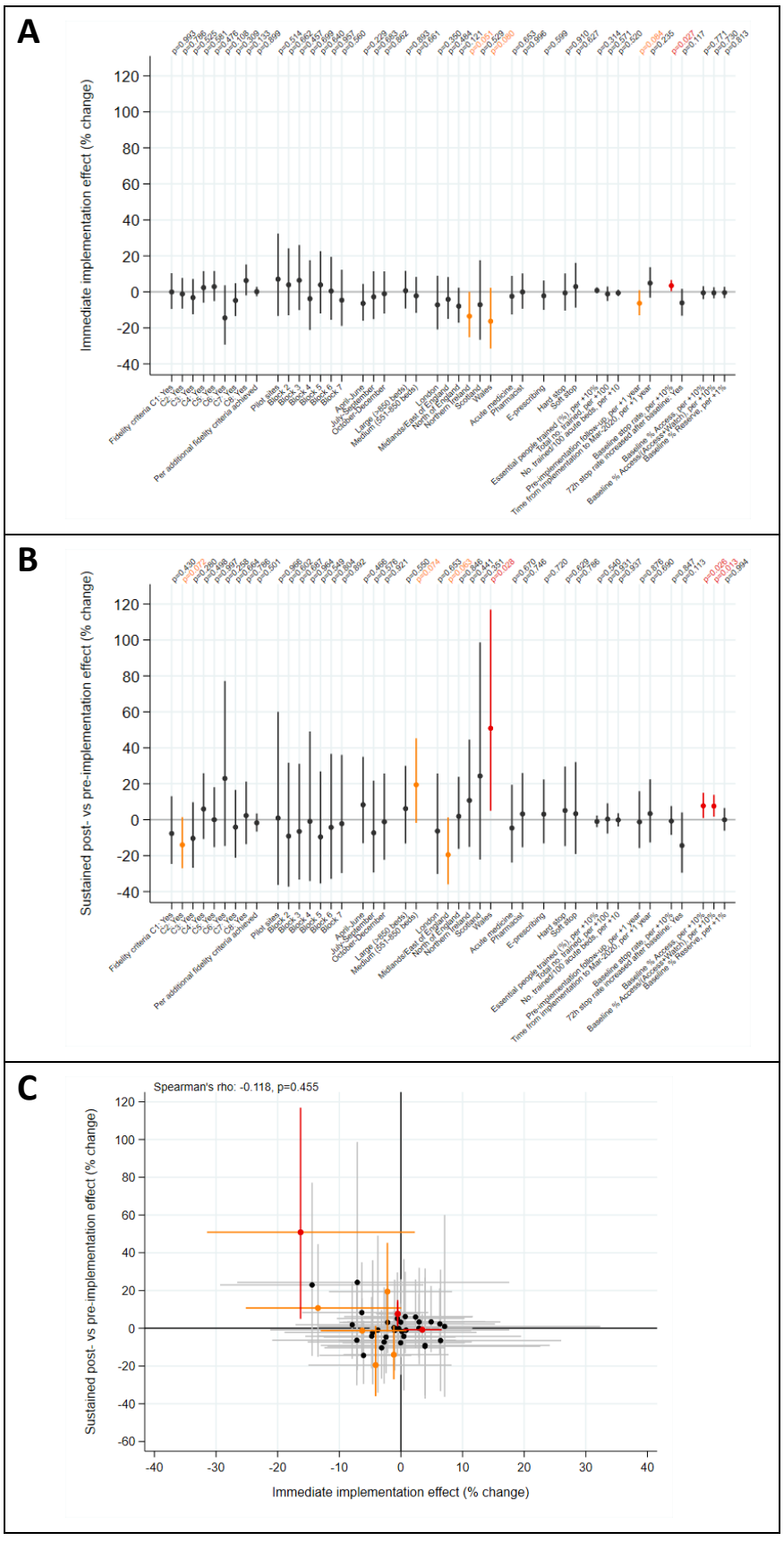


Figure 4.12 Access antibiotic DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

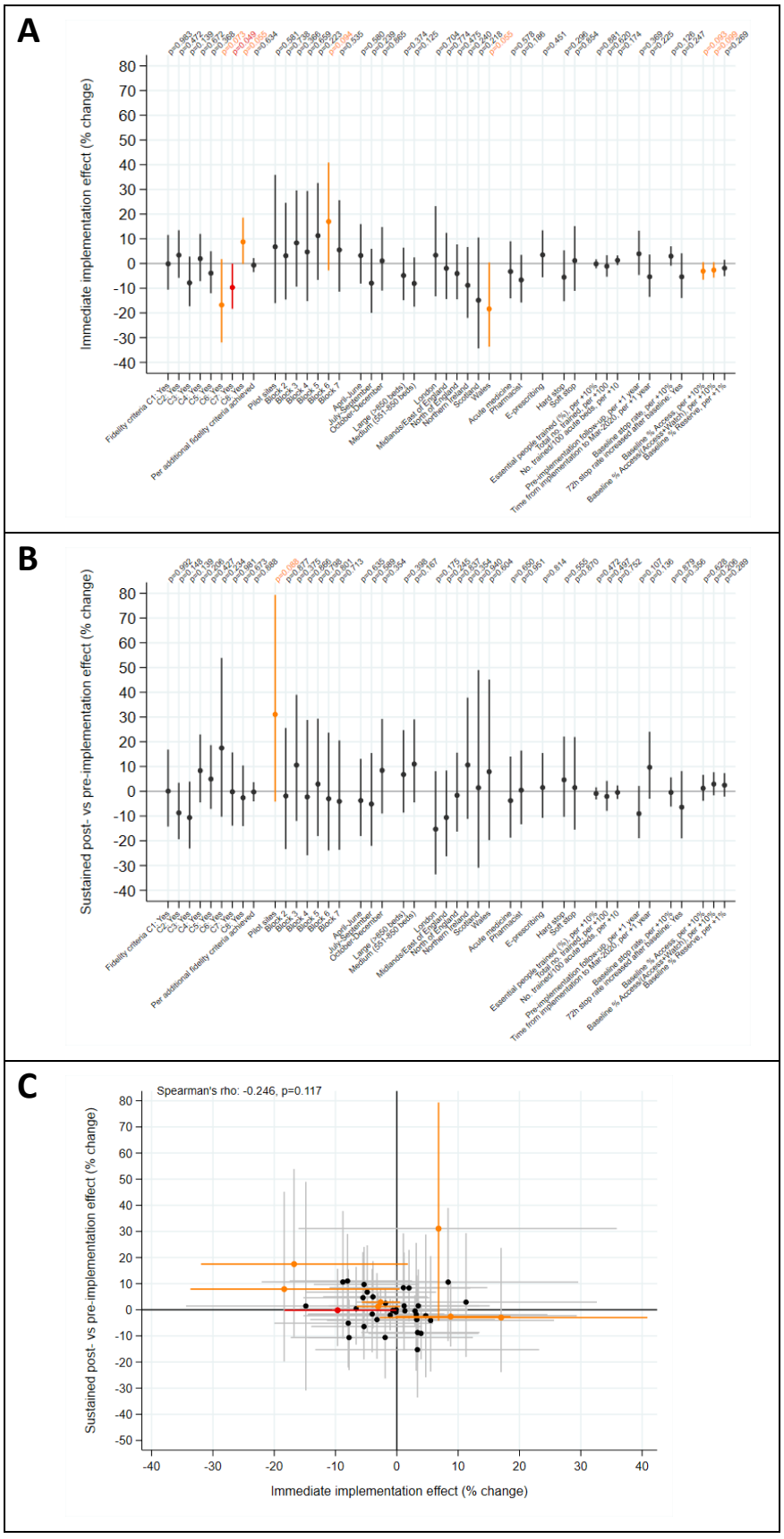


Figure 4.13 Watch antibiotic DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

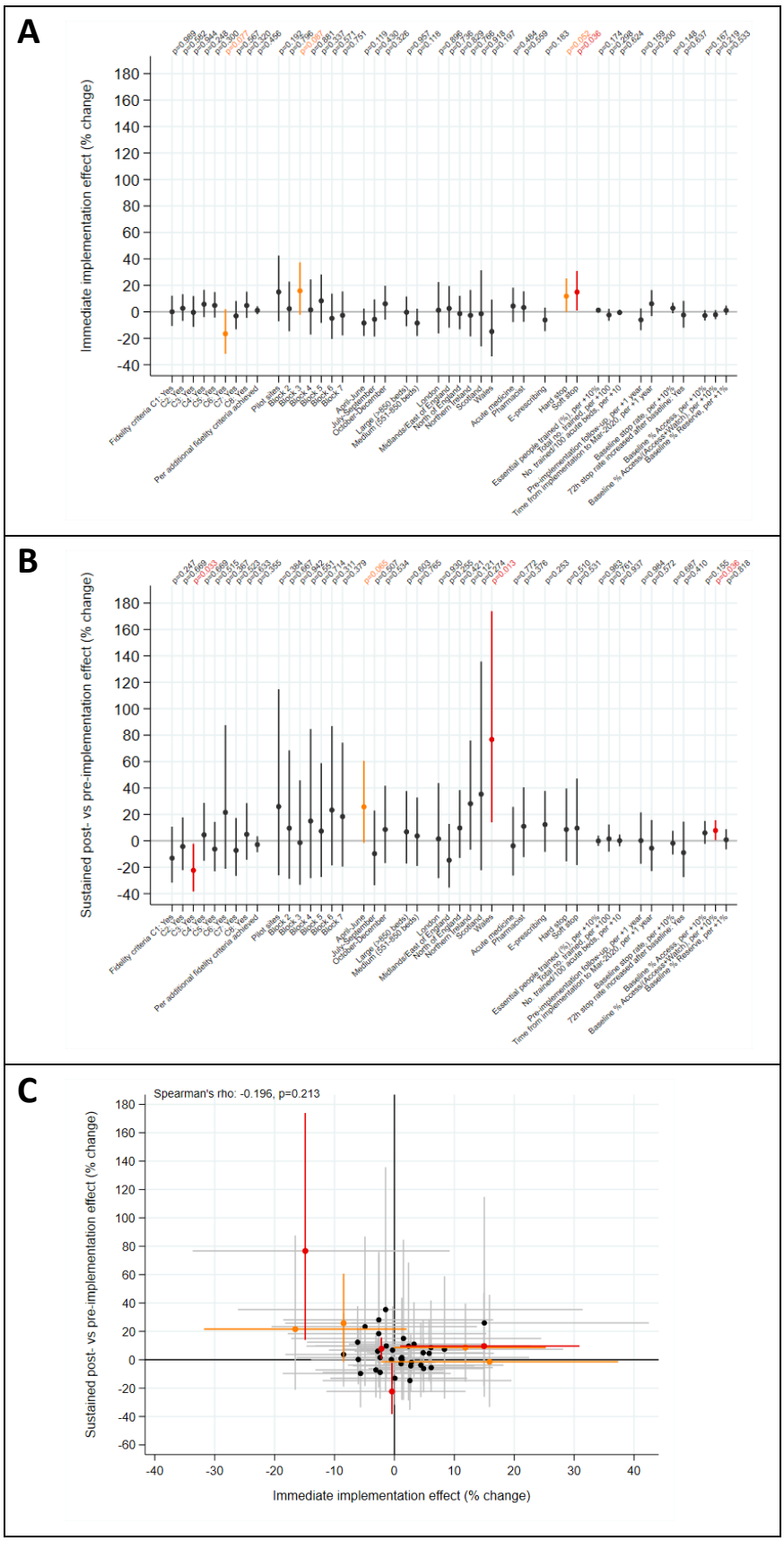


Figure 4.14 Reserve antibiotic DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

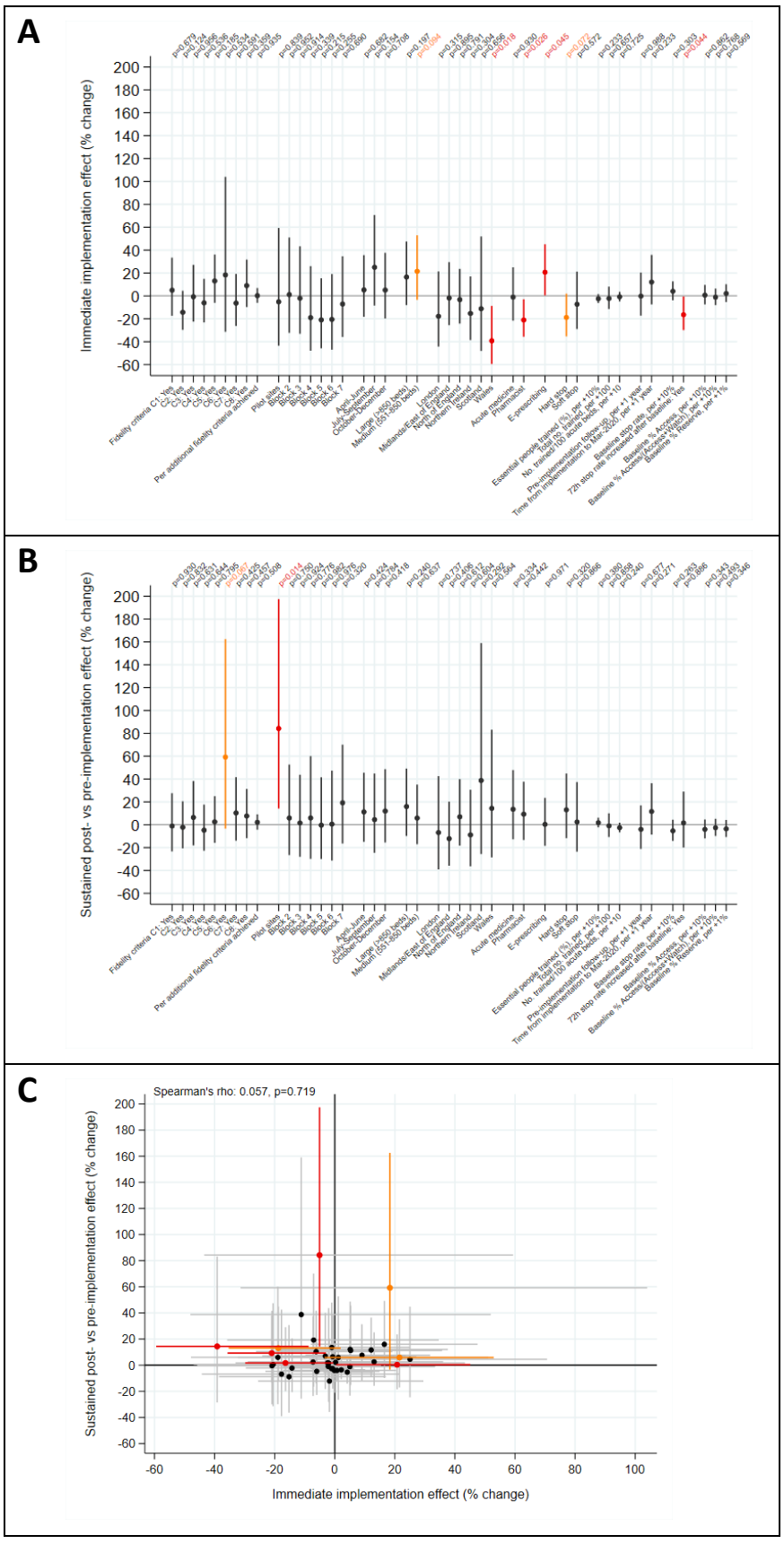


Figure 4.15 Oral antibiotic DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

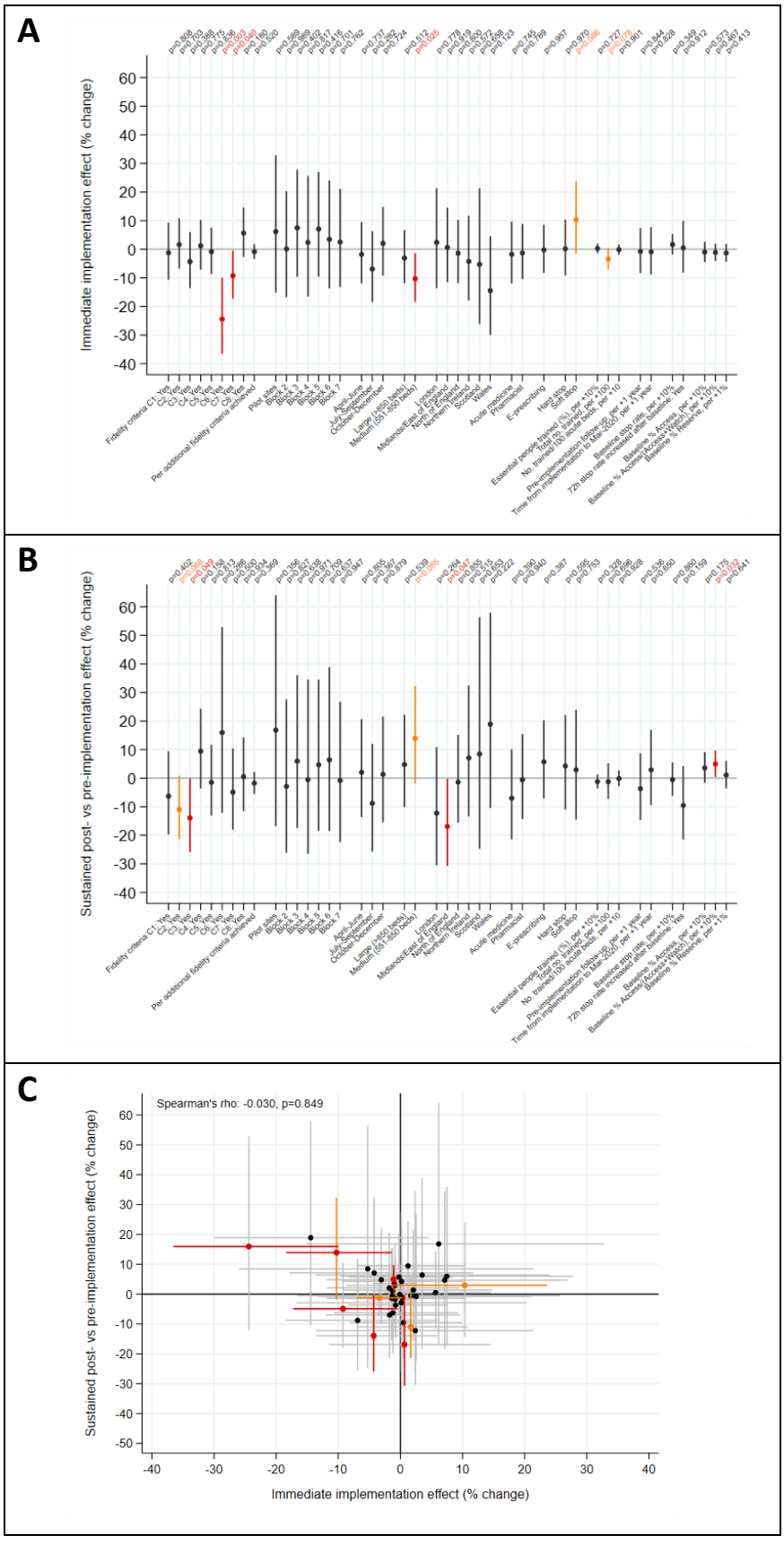


Figure 4.16 Parenteral DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

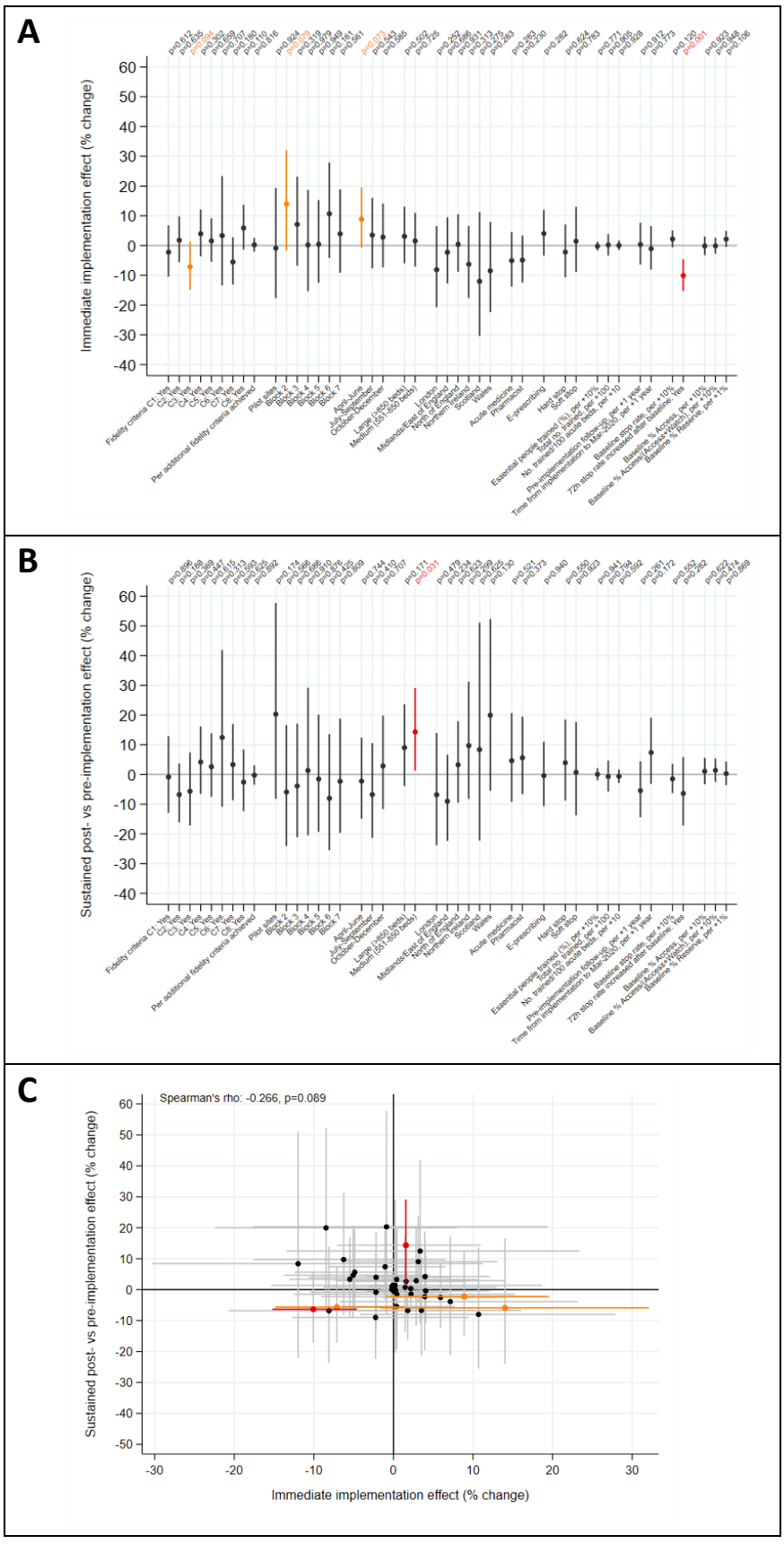


Figure 4.17 Quinolone DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

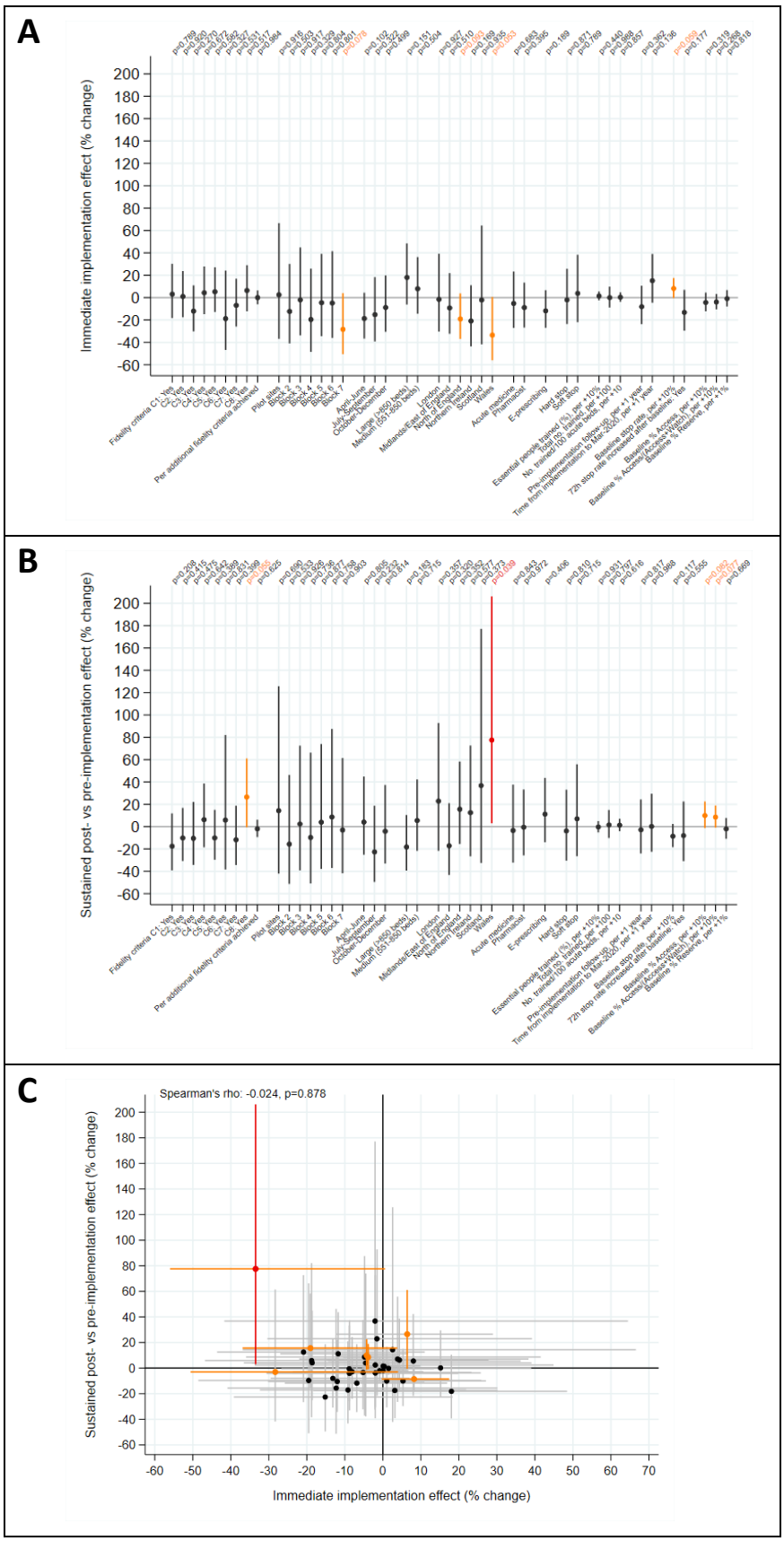


Figure 4.18 Carbapenem DDDs per admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

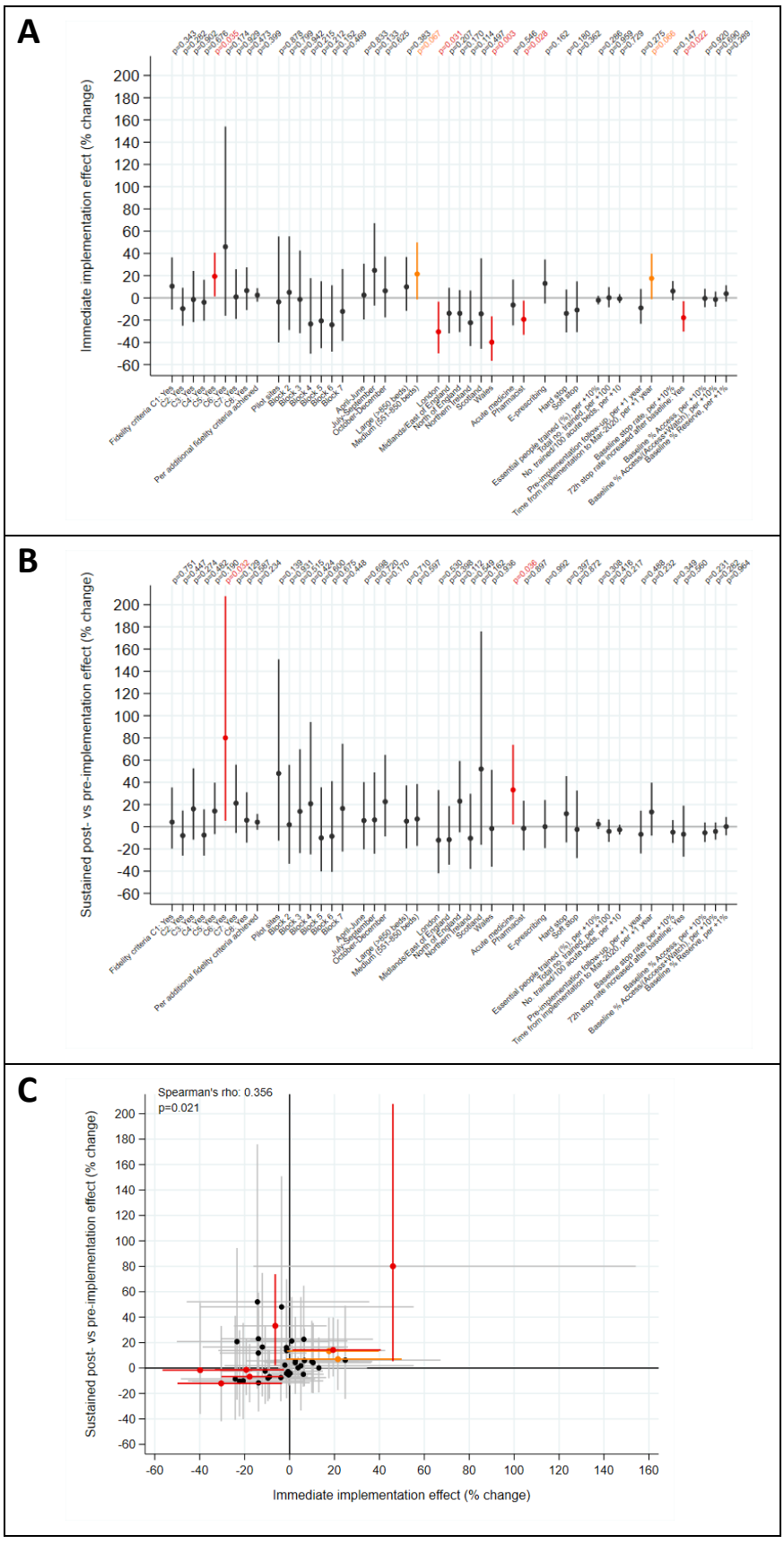


Figure 4.19 Died within 30 days of admission: Association between potential effect modifiers and site-level intervention effect estimates at implementation (A) and year-on-year (B), with an assessment of whether ranked coefficients from panels A-B are consistently in the same order (C)

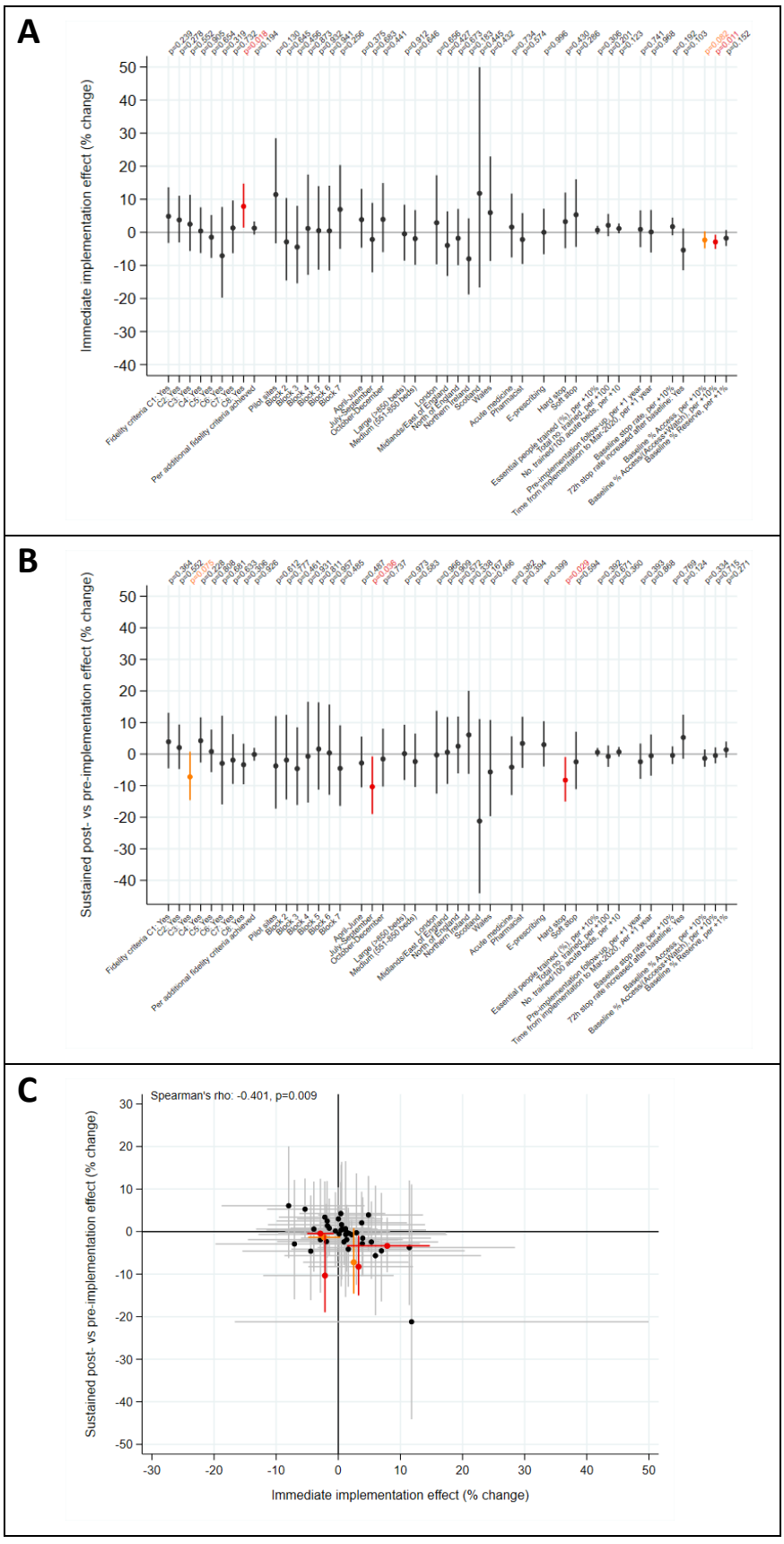


Table 4.4 Implementation effects overall and for factors with a significant test of heterogeneity in the intervention effect (p<0.05)

Outcome (DDDs per admission)	Effect type	Effect modifier considered	p-value[1]	Implementation effect (95% CI)
Total	Immediate effect	Overall	p=0.540	-1.0% (-4.0, 2.1)
		C6 not achieved	Reference	17.8% (1.4, 36.9)
		C6 achieved	**p=0.022**	-1.8% (-4.9, 1.5)
		C7 not achieved	Reference	5.9% (-1.0, 13.3)
		C7 achieved	**p=0.027**	-2.9% (-6.3, 0.7)
	Sustained effect	Overall	**p=0.042**	-4.8% (-9.1, -0.2)
		Small site (≤550 beds)	Reference	-11.3% (-19.9, -1.9)
		Medium site (551-850 beds)	**p=0.049**	1.6% (-7.2, 11.3)
		Large site (>850 beds)	p=0.374	-5.6% (-14.4, 4.0)
Narrow-spectrum	Immediate effect	Overall	p=0.488	1.3% (-2.3, 4.9)
		C6 not achieved	Reference	20.8% (2.2, 42.7)
		C6 achieved	**p=0.035**	0.4% (-3.2, 4.1)
		C7 not achieved	Reference	9.7% (1.9, 18.1)
		C7 achieved	**p=0.016**	-1.1% (-5.0, 2.9)
		Small site (≤550 beds)	Reference	6.3% (-0.7, 13.7)
		Medium site (551-850 beds)	**p=0.035**	-3.5% (-9.0, 2.3)
		Large site (>850 beds)	p=0.451	2.7% (-3.6, 9.3)
	Sustained effect	Overall	**p=0.019**	-5.2% (-9.4, -0.9)
Broad-spectrum	Immediate effect	Overall	**p=0.0003**	-6.3% (-9.5, -3.0)
		Baseline prescription stop rate, 25th percentile: 5.4%	**p=0.027**	-9.5% (-13.5, -5.2)
		Baseline prescription stop rate, median: 12.8%		-7.1% (-10.4, -3.8)
		Baseline prescription stop rate, 75th percentile: 21.4%		-4.4% (-8.3, -0.3)
	Sustained effect	Overall	p=0.395	-2.6% (-8.5, 3.6)
		Region: South of England	Reference	-2.8% (-14.9, 11.0)
		Region: North of England	p=0.846	-1.0% (-14.2, 14.3)
		Region: Midlands/East England	p=0.063	-21.8% (-35.1, -5.6)
		Region: London	p=0.653	-9.0% (-30.0, 18.4)
		Region: Scotland	p=0.351	20.8% (-23.0, 89.5)
		Region: Northern Ireland	p=0.441	7.6% (-14.6, 35.6)
		Region: Wales	**p=0.028**	46.6% (4.6, 105.6)
		Baseline % Access, 25th percentile: 37.0%	**p=0.026**	-9.8% (-18.6, -0.1)

Outcome (DDDs per admission)	Effect type	Effect modifier considered	p-value[1]	Implementation effect (95% CI)
		Baseline % Access, median: 47.4%		-2.6% (-9.6, 4.9)
		Baseline % Access, 75th percentile: 58.2%		5.5% (-4.9, 17.2)
		Baseline % Access/(Access + Watch), 25th percentile: 53.5%	**p=0.013**	-10.0% (-18.4, -0.8)
		Baseline % Access/(Access + Watch), median: 65.1%		-2.0% (-9.0, 5.4)
		Baseline % Access/(Access + Watch), 75th percentile: 76.4%		6.4% (-3.9, 17.8)
Access	Immediate effect	Overall	**p=0.037**	4.4% (0.3, 8.8)
		C7 not achieved	Reference	13.0% (3.4, 23.5)
		C7 achieved	**p=0.049**	2.1% (-2.7, 7.1)
	Sustained effect	Overall	**p=0.033**	-5.3% (-10.0, -0.4)
Watch	Immediate effect	Overall	**p<0.0001**	-9.5% (-13.2, -5.6)
		Did not implement a prescription hard/soft stop	Reference	-17.7% (-25.3, -9.4)
		Implemented a hard stop	p=0.052	-8.0% (-13.3, -2.3)
		Implemented a soft stop	**p=0.036**	-5.4% (-13.3, 3.2)
	Sustained effect	Overall	**p=0.001**	-11.0% (-17.1, -4.5)
		C3 not achieved	Reference	8.9% (-11.4, 33.8)
		C3 achieved	**p=0.033**	-15.4% (-23.8, -6.1)
		Region: South of England	Reference	-16.5% (-28.7, -2.1)
		Region: North of England	p=0.421	-8.4% (-22.7, 8.6)
		Region: Midlands/East England	p=0.255	-28.7% (-43.3, -10.3)
		Region: London	p=0.930	-15.2% (-37.8, 15.6)
		Region: Scotland	p=0.274	13.0% (-33.6, 92.5)
		Region: Northern Ireland	p=0.121	7.0% (-18.7, 40.9)
		Region: Wales	**p=0.013**	47.6% (-2.0, 122.2)
		Baseline % Access/(Access + Watch), 25th percentile: 53.5%	**p=0.036**	-17.9% (-27.1, -7.4)
		Baseline % Access/(Access + Watch), median: 65.1%		-10.3% (-18.1, -1.9)
		Baseline % Access/(Access + Watch), 75th percentile: 76.4%		-2.4% (-14.0, 10.8)
Reserve	Immediate effect	Overall	p=0.543	3.0% (-6.4, 13.3)
		Region: South of England	Reference	11.5% (-6.7, 33.2)
		Region: North of England	p=0.791	8.0% (-8.7, 27.6)
		Region: Midlands/East England	p=0.895	9.5% (-11.4, 35.3)
		Region: London	p=0.315	-8.3% (-35.2, 29.7)

Outcome (DDDs per admission)	Effect type	Effect modifier considered	p-value[1]	Implementation effect (95% CI)
		Region: Scotland	p=0.656	-1.0% (-40.4, 64.4)
		Region: Northern Ireland	p=0.304	-5.5% (-27.8, 23.6)
		Region: Wales	**p=0.018**	-32.2% (-52.9, -2.3)
		Role of site study lead: Microbiology / Infectious diseases	Reference	10.5% (-2.3, 24.9)
		Role of site study lead: Acute Medicine	p=0.930	9.3% (-10.4, 33.4)
		Role of site study lead: Pharmacist	**p=0.026**	-12.7% (-26.0, 3.0)
		Paper prescribing system	Reference	-3.8% (-13.9, 7.5)
		Electronic prescribing system	**p=0.045**	16.2% (0.3, 34.6)
		Did not increase in prescription review stop rate after baseline	Reference	17.3% (2.0, 34.9)
		Increased prescription review stop rate after baseline	**p=0.044**	-2.0% (-11.8, 8.9)
	Sustained effect	Overall	p=0.107	7.3% (-1.5, 17.0)
		Randomisation block 1	Reference	83.4% (22.4, 174.6)
		Randomisation block: pilot sites	**p=0.014**	-0.5% (-23.1, 28.8)
		Randomisation block 2	p=0.750	5.4% (-18.7, 36.7)
		Randomisation block 3	p=0.924	1.1% (-19.7, 27.5)
		Randomisation block 4	p=0.776	5.5% (-23.6, 45.5)
		Randomisation block 5	p=0.982	-0.9% (-22.0, 25.9)
		Randomisation block 6	p=0.976	0.1% (-24.6, 32.7)
		Randomisation block 7	p=0.320	18.7% (-7.1, 51.5)
Oral	Immediate effect	Overall	p=0.458	-1.4% (-4.8, 2.3)
		C6 not achieved	Reference	28.9% (8.6, 52.9)
		C6 achieved	**p=0.003**	-2.6% (-6.0, 1.0)
		C7 not achieved	Reference	6.4% (-2.0, 15.5)
		C7 achieved	**p=0.040**	-3.4% (-7.5, 0.8)
		Small site (≤550 beds)	Reference	3.7% (-3.3, 11.2)
		Medium site (551-850 beds)	**p=0.025**	-7.0% (-12.7, -0.9)
		Large site (>850 beds)	p=0.512	0.5% (-5.8, 7.3)
	Sustained effect	Overall	**p=0.012**	-6.4% (-11.2, -1.4)
		C3 not achieved	Reference	5.6% (-7.7, 20.7)
		C3 achieved	**p=0.049**	-9.1% (-15.0, -2.9)
		Region: South of England	Reference	-4.0% (-13.7, 6.7)
		Region: North of England	p=0.855	-5.4% (-15.5, 6.0)
		Region: Midlands/East England	**p=0.047**	-20.2% (-31.2, -7.4)
		Region: London	p=0.264	-15.8% (-31.6, 3.8)

Outcome (DDDs per admission)	Effect type	Effect modifier considered	p-value[1]	Implementation effect (95% CI)
		Region: Scotland	p=0.653	4.1% (-26.7, 47.8)
		Region: Northern Ireland	p=0.515	2.8% (-14.5, 23.6)
		Region: Wales	p=0.222	14.1% (-12.3, 48.5)
		Baseline % Access/(Access + Watch), 25th percentile: 53.5%	**p=0.032**	-11.2% (-17.6, -4.2)
		Baseline % Access/(Access + Watch), median: 65.1%		-6.0% (-11.2, -0.5)
		Baseline % Access/(Access + Watch), 75th percentile: 76.4%		-0.7% (-8.3, 7.5)
Parenteral	Immediate effect	Overall	p=0.655	-0.8% (-4.2, 2.8)
		Randomised implementation in January-March	Reference	-5.1% (-11.4, 1.5)
		Randomised implementation in April-June	**p=0.040**	4.0% (-1.6, 9.8)
		Randomised implementation in July-September	p=0.883	-4.4% (-12.3, 4.2)
		Randomised implementation in October-December	p=0.460	-1.4% (-9.1, 7.0)
		Did not increase in prescription review stop rate after baseline	Reference	5.9% (1.1, 10.8)
		Increased prescription review stop rate after baseline	**p=0.001**	-4.8% (-8.3, -1.2)
	Sustained effect	Overall	p=0.703	-0.9% (-5.2, 3.6)
		Small site (≤550 beds)	Reference	-8.4% (-16.4, 0.3)
		Medium site (551-850 beds)	**p=0.031**	4.7% (-3.4, 13.4)
		Large site (>850 beds)	p=0.171	-0.2% (-8.5, 8.9)
Quinolone	Immediate effect	Overall	**p<0.0001**	-15.5% (-21.9, -8.5)
	Sustained effect	Overall	**p=0.007**	-13.8% (-22.5, -4.0)
		Region: South of England	Reference	-21.0% (-36.8, -1.3)
		Region: North of England	p=0.352	-8.7% (-26.9, 14.0)
		Region: Midlands/East England	p=0.320	-34.6% (-51.9, -11.1)
		Region: London	p=0.357	-2.9% (-34.4, 43.6)
		Region: Scotland	p=0.373	8.0% (-44.8, 111.3)
		Region: Northern Ireland	p=0.577	-11.1% (-38.4, 28.1)
		Region: Wales	**p=0.039**	40.2% (-14.8, 130.7)
Carbapenem	Immediate effect	Overall	p=0.376	4.1% (-4.7, 13.7)
		C5 not achieved	Reference	-4.0% (-14.0, 7.3)
		C5 achieved	**p=0.035**	14.7% (1.6, 29.4)
		Region: South of England	Reference	22.9% (4.6, 44.3)

Outcome (DDDs per admission)	Effect type	Effect modifier considered	p-value[1]	Implementation effect (95% CI)
		Region: North of England	p=0.170	5.8% (-8.6, 22.5)
		Region: Midlands/East England	p=0.207	5.9% (-10.9, 25.8)
		Region: London	**p=0.031**	-14.5% (-35.8, 13.9)
		Region: Scotland	p=0.497	5.3% (-31.5, 61.9)
		Region: Northern Ireland	p=0.114	-4.5% (-27.3, 25.4)
		Region: Wales	**p=0.003**	-26.1% (-44.3, -1.7)
		Role of site study lead: Microbiology / Infectious diseases	Reference	12.4% (0.2, 26.1)
		Role of site study lead: Acute Medicine	p=0.546	5.2% (-12.7, 26.8)
		Role of site study lead: Pharmacist	**p=0.028**	-9.3% (-22.0, 5.5)
		Did not increase in prescription review stop rate after baseline	Reference	21.0% (6.1, 38.0)
		Increased prescription review stop rate after baseline	**p=0.022**	-0.5% (-10.1, 10.1)
	Sustained effect	Overall	**p=0.015**	12.3% (2.3, 23.2)
		C6 not achieved	Reference	-36.4% (-62.5, 7.7)
		C6 achieved	**p=0.032**	14.5% (3.8, 26.3)
		Role of site study lead: Microbiology / Infectious diseases	Reference	7.2% (-6.1, 22.4)
		Role of site study lead: Acute Medicine	**p=0.036**	42.8% (13.3, 79.9)
		Role of site study lead: Pharmacist	p=0.897	5.7% (-11.9, 26.7)
30-day mortality	Immediate effect	Overall	p=0.079	-2.7% (-5.7, 0.3)
		C8 not achieved	Reference	-7.0% (-11.3, -2.5)
		C8 achieved	**p=0.018**	0.3% (-3.6, 4.4)
		Baseline % Access/(Access + Watch), 25th percentile: 53.5%	**p=0.011**	0.1% (-3.6, 4.1)
		Baseline % Access/(Access + Watch), median: 65.1%		-3.2% (-6.0, -0.3)
		Baseline % Access/(Access + Watch), 75th percentile: 76.4%		-6.4% (-10.1, -2.5)
	Sustained effect	Overall	p=0.060	3.0% (-0.1, 6.2)
		Randomised implementation in January-March	Reference	6.3% (-0.3, 13.4)
		Randomised implementation in April-June	p=0.487	3.3% (-1.9, 8.8)

Outcome (DDDs per admission)	Effect type	Effect modifier considered	p-value[1]	Implementation effect (95% CI)
		Randomised implementation in July-September	**p=0.036**	-4.7% (-11.9, 3.1)
		Randomised implementation in October-December	p=0.737	4.7% (-2.1, 11.9)
		Did not implement a prescription hard/soft stop	Reference	8.6% (1.8, 15.8)
		Implemented a hard stop	**p=0.029**	-0.4% (-4.4, 3.9)
		Implemented a soft stop	p=0.594	5.9% (-0.9, 13.3)

[1] For potential effect modifiers the p-value reflects a test of heterogeneity in the intervention effect (i.e. derived through univariate random-effects meta-regression). For overall implementation effects the p-value is from the estimated weighted mean of the site-specific effects (i.e. derived through random-effects meta-analysis).

Chapter 5: An observational analysis of hospital-level antibiotic use from all acute hospitals in England covering the period over which the ARK trial was conducted

5.1 Introduction

Cluster randomised trial (CRT) designs are widely used to evaluate interventions that modify how services are delivered, including complex behaviour change and policy interventions(264). In the ARK-Hospital trial(105), NHS Trusts (or equivalent healthcare organisations in the Devolved Administrations) were the preferred unit of randomisation because the intervention targeted healthcare professionals who regularly move between teams and/or work in different areas of one organisation. Had the trial randomised individual patients or clinical teams instead, this frequent movement would have introduced substantial contamination (exposure misclassification) and attenuated the estimand. Also, instead of randomising Trusts using a conventional parallel group design (i.e. where sites are randomised to either implement or act as controls, with controls possibly waitlisted to implement at the end of the evaluation), the ARK intervention was implemented using a stepped wedge design, whereby sites transitioned from control to intervention states at sequential, randomised timepoints (as depicted in appendix Figure S2 in (105)).

The decision to adopt a stepped wedge design had several advantages. For example, by ensuring all sites had the chance to implement, the stepped wedge design facilitated cluster recruitment, as sites had financial CQUIN incentives to reduce antibiotic use(41) and meet reduction targets mandated in hospital contracts(45). Also, while a parallel CRT would have been more efficient in a setting with a large number of similar clusters (i.e. where intra-cluster correlation is small), in this particular context, the number of available NHS Trusts was limited and sites were large and heterogenous. Since prescribing behaviour is influenced by local Trust guidelines and policies, which vary by Trust(60), outcome variance was likely to be much greater between-Trusts than within-Trusts (i.e. intra-cluster correlation was expected to be high). This meant a stepped wedge design was more

efficient, achieving the desired power with fewer clusters than would have been required in a parallel CRT, as each Trust acted as its own control, reducing the potential for chance imbalance (despite randomisation) in Trust-level factors across the control and exposed conditions(265–267).

Despite these advantages, there are challenges that arise from the staggered nature of the stepped wedge design, which can complicate interpretation of results. Since the pre-implementation (control) stage always occurs before the post-implementation stage, calendar time is a potential confounder and the analytic approach needs to differentiate changes in outcomes due to the intervention from any underlying changes in the outcome over time (called secular trends). In the ARK trial, intervention effects were assessed using a separate interrupted time series analysis in each Trust, which allowed for the estimation of both an immediate implementation effect and a sustained year-on-year change in trend post- vs pre-implementation, with an overall weighted mean of the site-specific effects derived using meta-analytic techniques, as described in Chapter 4. However, estimates of the change in trend implicitly assume that the pre-intervention secular trends would have continued had ARK not been implemented. If this assumption does not hold and the pre-implementation calendar trends later increased (or decreased) instead (e.g. due to other changes in antimicrobial stewardship, clinical practice, or AMR rates), then the modest overall impact of the ARK intervention – a 4.8% per year reduction in total antibiotic DDDs/admission in general medicine – may have been even greater (or smaller) than the comparison of trends post- vs pre-implementation in ARK Trusts suggests.

Given the risk of bias from mis-specified secular trends, as has been flagged by others(268,269), this chapter therefore examined whether trends in the use of antibiotics differ between ARK sites and all other acute (non-ARK) Trusts in England, with the latter sites serving as a counterfactual to estimate what may have occurred in ARK sites had they not implemented. This comparison considered both trends before the ARK trial began in February 2018 (i.e. as differences would suggest the ARK sites are not representative, and hence findings may not be generalisable), and also trends after the onset

of COVID (post-intervention in ARK sites) to help interpret the ARK trial findings in the context of national background trends in the acute hospital setting. Finally, since the ARK intervention effects may have changed over time and follow-up in the ARK trial ended in October 2020 – just 7 months after the onset of COVID in March 2020 and only 14 months after implementation in some sites (median 23 months, IQR: 18–28, range: 14–37) – this chapter also examined how ARK implementation effects change with access to more longitudinal data post-COVID, shedding light on the longer-term impact of the ARK intervention.

5.2 Methods

5.2.1 Data sources

Antibiotic DDDs were sourced from IQVIA (formerly Quintiles/IMS Health) through UKHSA and reflect hospital-level prescribing (inpatient and outpatient combined), by month, in 159 secondary care organisations in England. This database is also used by the English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR) to monitor antibiotic consumption in acute NHS Trusts(34). As summarised in **Table 5.1**, the IQVIA data differs from the antibiotic data used in the ARK trial in some important respects, namely:

- The IQVIA data includes all acute NHS Trusts in England, enabling estimation of:
    a) Trends in antibiotic use in English ARK vs non-ARK sites, including pre-intervention in ARK sites to assess generalisability, and post-intervention to assess whether the intervention had a greater (or lesser) impact than suggested by the trial-site specific analysis alone
    b) The longer-term impact of the ARK intervention in English ARK sites, using data with longer follow-up before and after implementation than was available in the ARK trial

- The IQVIA data reports prescribing by hospital specialty and can therefore be used to compare antibiotic use in general medicine vs overall (at the Trust-level), facilitating an

assessment of, and adjustment for, possible temporal changes in specialty coding that was not possible in the ARK trial (see Methods below)

Antibiotic data from the ARK trial was also analysed to compare with the IQVIA data.

Table 5.1 Differences between antibiotic data sourced from IQVIA vs the ARK-Hospital trial

Characteristic	Antibiotic data source		Comment
	ARK trial	IQVIA	
How data was obtained	Through data sharing agreements with ARK trial sites	From IQVIA (via UKHSA), which obtained the data directly from acute NHS Trusts in England	IQVIA data access was granted through an ARK trial protocol amendment, approved by the South Central Oxford C Research Ethics Committee (reference: 17/SC/0034) and the Health Research Authority (HRA) (IRAS Project ID 217773) in December 2020
Number of NHS Trusts	39 in the UK, of which 32 were in England	159 in England, including all 32 English ARK trial sites	The ARK trial analysis excluded 1 site in England (Frimley Health NHS Foundation Trust) due to data quality issues, leaving 38 sites (31 in England) in analyses of antibiotic endpoints(105). Since the cleaning of IQVIA data excluded 2 other ARK trial sites (Royal United Hospitals Bath and Wye Valley, **Figure 5.1**) due to data quality issues (see below), the analytic datasets shared 29 ARK sites in common.
Time span covered	February 2016 – October 2020	May 2015 – June 2022[1]	ARK trial sites were randomised to implement the intervention between 25 September 2017 – 1 July 2019; thus, the IQVIA data provided greater follow-up in these sites, spanning a minimum of 29 (vs 20) months pre-implementation and at least 35 (vs 15) months post-implementation
Metric of antibiotic use	Hospital-level DDDs by month, drug, ATC code, and route of administration	Hospital-level DDDs by month, drug, ATC code, route of administration, and specialty	Both data sources reflected routinely collected inpatient and outpatient prescribing (combined). Since the ARK trial data excluded non-systemic and anti-tuberculosis medications, these cleaning steps were also applied to the IQVIA data (see Methods).
Hospital specialties included	General medicine only[2]	All hospital specialties, including	Since the IQVIA data reported DDDs for each specialty separately, it can be used to

Characteristic	Antibiotic data source		Comment
	ARK trial	IQVIA	
		general medicine	analyse DDDs in general medicine and overall at the Trust-level.
Definition of general medicine	Defined by ward or consultant specialty depending on local patient pathways, at the discretion of each site	Prescribing specialties were defined by each secondary care organisation according to local routine coding practices	To understand whether the algorithms used to define general medicine differed between the two data sources, I compared differences in DDDs reported among the ARK trial sites (see Methods).
WHO ATC / DDD index	2020	2022	No DDD alterations occurred between 2020-2022 among included antibiotics(199)

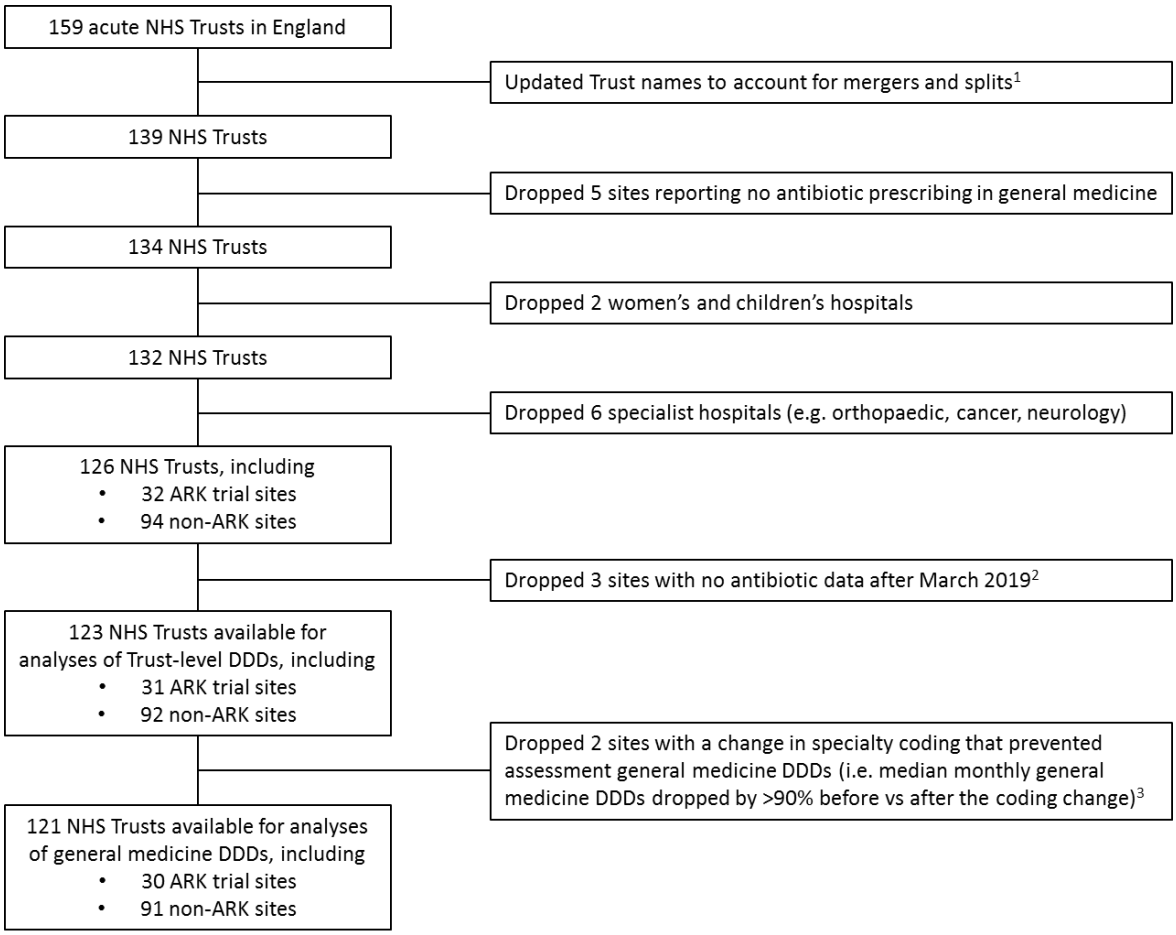
[1] Although IQVIA data was available back to January 2014, analyses were restricted to data from May 2015 onwards due to data quality concerns before May 2015 expressed by UKHSA.

[2] 2/39 sites only provided Trust-level antibiotic data for the ARK trial analysis due to limitations posed by local pharmacy systems. However, as these sites were in Scotland and Wales, they were not included in the analyses in this chapter.

5.2.2 IQVIA data cleaning

To facilitate the comparison of estimates derived from each data source, I compared the antibiotics in IQVIA vs ARK trial data and excluded the same non-systemic agents (e.g. neomycin), as per ARK trial methods(105). Anti-tuberculosis medications (excluded in the ARK trial) had already been dropped from the IQVIA data in advance (before receipt of data). Next, since the ARK trial sites admitted unplanned adult medical inpatients (i.e. the trial's target patient population), I restricted the IQVIA data to the subset of acute Trusts providing emergency adult medical services. Lastly, to prevent temporal changes in antibiotic DDDs from being impacted by Trust mergers or splitting, I updated Trust names to ensure they were current as of June 2022 (based on records maintained by NHS Digital(131)), as described in **Figure 5.1** footnote. An exception was made for mergers involving ARK trial sites (n=2); in this case the merger was not accounted for and the sites were analysed separately, up to the merger date, to match the trial analysis.

Figure 5.1 IQVIA data cleaning steps



[1] If Trusts A and B merged to create Trust C, the data from Trusts A and B was analysed as belonging to Trust C. Only one Trust split—on 1 November 2014 (before the study timeframe), Mid Staffordshire NHS Foundation Trust split into The Royal Wolverhampton Hospitals NHS Trust and the University Hospitals of North Midlands NHS Trust. Despite this split, the raw IQVIA data reported DDDs from Mid Staffordshire NHS Foundation Trust through June 2022 (i.e. these DDDs should have been reported as belonging to either The Royal Wolverhampton Hospitals or the University Hospitals of North Midlands). It was not possible to determine what fraction of monthly DDDs from Mid Staffordshire belonged to each of these two sites, however since University Hospitals of North Midlands NHS Trust had more Trust-level admissions per year than The Royal Wolverhampton Hospitals NHS Trust(270), DDDs from Mid Staffordshire NHS Foundation Trust were counted as belonging to University Hospitals of North Midlands.

[2] Excluded as data from these sites could not be used to estimate trends in antibiotic use during or after the ARK trial, as the last ARK site was randomised to implement in July 2019. One of these three sites was in the ARK trial (Royal United Hospitals Bath NHS Foundation Trust).

[3] One of these two sites was in the ARK trial (Wye Valley NHS Trust). See details below.

5.2.3 Detecting possible changes in specialty coding in IQVIA data

Changes in specialty coding have the potential to confound estimated antibiotic prescribing trends by erroneously increasing or decreasing the magnitude of monthly antibiotic DDDs in general

medicine. To detect possible changes in specialty coding in the IQVIA data, monthly variations in general medicine DDDs were visually assessed for abrupt, sustained shifts ('step changes') in prescribing trends using spaghetti plots and charts comparing general medical DDDs as a share of Trust-level DDDs by month and site. When concerns with specialty coding were identified, I applied fixed rules to limit this source of potential confounding bias:

1. If a suspected change in specialty coding occurred within 12 months of the planned analysis start date (May 2015), the start date was pushed forward to exclude data before the change.

2. If a suspected change in specialty coding occurred within 12 months of the planned analysis end date (June 2022), the analysis end date was pulled back to exclude data after the change.

3. If a suspected change in specialty coding occurred between May 2016 (planned start date +1 year) and June 2021 (planned end date -1 year) and the median monthly general medicine DDDs changed substantially (by a factor of 10 or more, before vs after the step change), then the site was excluded, as this was considered likely to prevent valid estimation of antibiotic trends.

4. If applying the rules above did not remove this source of confounding, the step change was adjusted for by including a binary factor (=0 before the step change, and =1 afterwards) in the negative binomial models used to estimate antibiotic use trends (see statistical Methods below).

5.2.4 Comparing ARK vs IQVIA data

Since both data sources were restricted to hospital-level prescribing of systemic antibiotics, any differences in DDDs may be due to differences in the algorithm used to define general medicine. To assess the similarity of these datasets, I compared the absolute and relative magnitude of DDDs in general medicine in IQVIA vs ARK data. Among ARK sites, this comparison provided a secondary source of supporting evidence to detect possible changes in specialty coding in IQVIA data.

5.2.5 Outcomes

Since the ARK trial's co-primary antibiotic outcome was monthly total DDDs/admission, I initially planned to normalise DDDs for admissions in all analyses. However, while I applied to NHS Digital in August 2021 for an extract of general medicine admissions to all acute Trusts in England, their Independent Group Advising on the Release of Data (IGARD) did not schedule a review of my application until 18 months later, in January 2023. Feedback in the lead-up to this meeting suggested that if my application was approved, a framework Data Sharing Agreement would need to be signed and then their data production team would require further 3-4 months to deliver the extract, suggesting the data would not be available until May/June 2023 at the earliest. Given this delay, I withdrew the application and instead obtained admissions from ARK trial data. Since these admissions were only available for ARK sites from February/2016-October/2020, I also analysed DDDs without normalising for admissions as this could be assessed in both ARK and non-ARK sites and from May/2015-June/2022. Analyses therefore considered up to 3 separate outcomes, depending on the data source and estimates derived (**Table 5.2**).

Table 5.2 Antibiotic endpoints considered when estimating antibiotic prescribing trends and ARK implementation effects

Outcome (measured monthly, at the hospital-level)	**Used to estimate prescribing trends in non-ARK sites?**	**Used to estimate prescribing trends and ARK implementation effects in ARK sites?**	
	IQVIA DDDs	**ARK trial DDDs**	**IQVIA DDDs**
1. Total DDDs in general medicine per admission[1]	No, admissions not available in non-ARK sites	Yes	Yes
2. Total DDDs in general medicine	Yes	Yes	Yes
3. Total DDDs across all specialties[2]	Yes	No, DDDs not available outside general medicine	Yes

[1] The admissions denominator was obtained from the ARK trial and reflected adult (age ≥16 years) general medical admissions, defined using consultant specialty codes (as per appendix Figure S1 in (105)), available from February/2016-October/2020 in ARK sites only.

[2] Included as analyses identified potential challenges with specialty coding (see Results)

5.2.6 Statistical methods

Estimating prescribing trends in ARK vs non-ARK sites using IQVIA data, May 2015-June 2022 (Model 1) (Figure 5.2)

To understand whether antibiotic prescribing trends differed between ARK and non-ARK sites, I started by examining DDDs in general medicine using a 3-step approach:

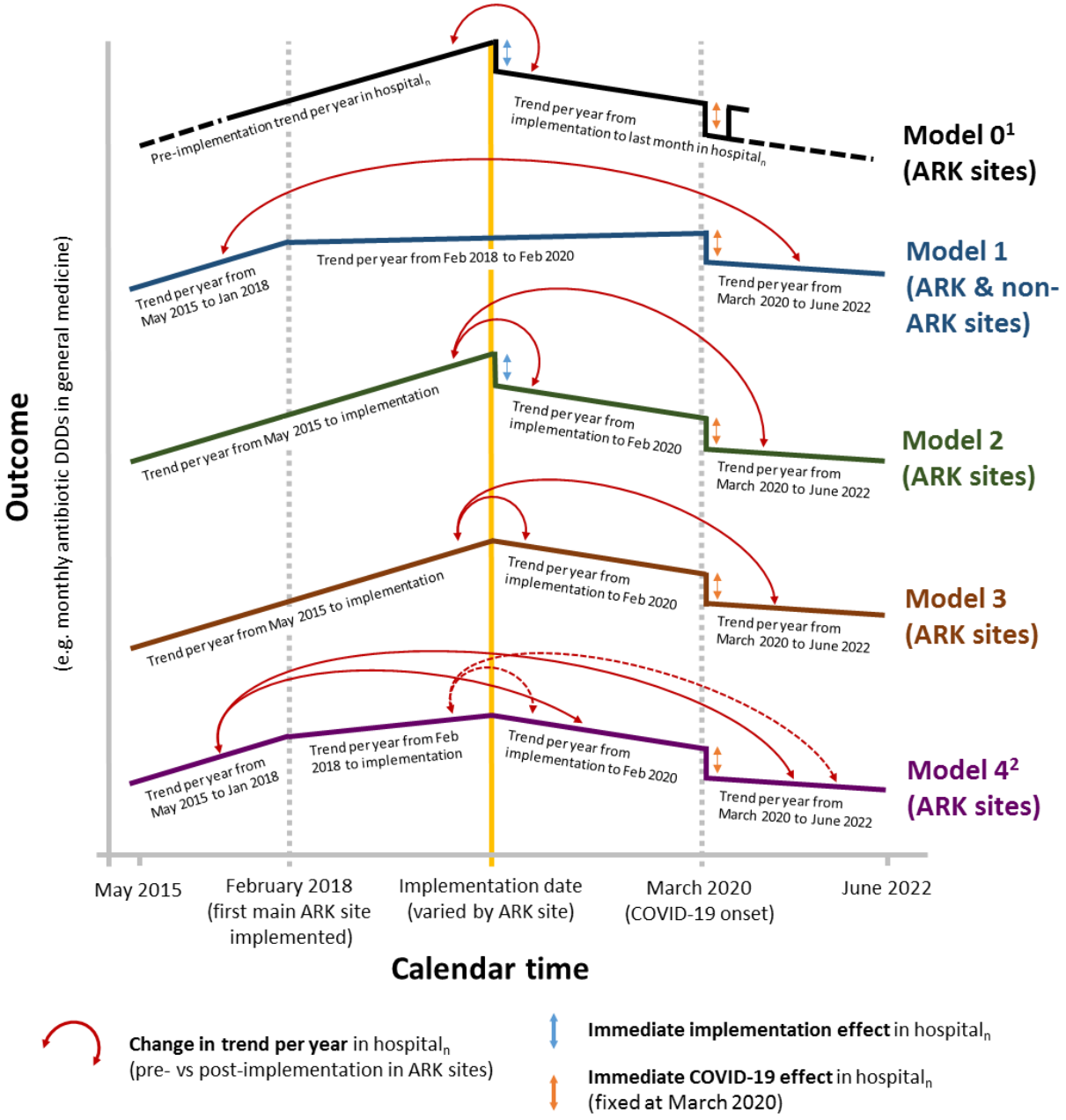
- Step 1: I fit a negative binomial model separately in each ARK and non-ARK site. Since antibiotic use showed seasonal variation (reflecting variation in hospital case-mix and presenting infections), as in the main ARK trial analysis(105), this model was adjusted for a cyclical time effect between May/2015-February/2020, with month of year fit as a sin()+cos() function to ensure smooth changes in estimated prescribing trends from year to year (set =0 from March 2020 onwards as seasonality was not observed after COVID onset). Similarly, since COVID greatly affected antibiotic prescribing, this model adjusted for COVID as binary factor (=1 from March 2020 onwards and =0 otherwise). In 4 ARK sites and 8 non-ARK sites an additional post-COVID adjustment was added to models (e.g. =1 from July 2020 onwards, otherwise =0) as the immediate change in the outcome due to COVID was especially pronounced but temporary, leading to what would have otherwise been poorly fitting post-COVID trend estimates.
- Step 2: I used an interrupted time series analysis in each site (**Figure 5.2**) to estimate:
    - A prescribing trend before the first ARK trial site implemented in February 2018 (i.e. change in incidence rate ratio (IRR) per year, from May 2015-January 2018)
    - A prescribing trend during the ARK trial up to the start of COVID (i.e. change in IRR per year, from February 2018-February 2020)
    - A prescribing trend post-COVID (i.e. change in IRR per year, March 2020-June 2022)
    - The immediate COVID effect (i.e. change in IRR associated with March 2020)

    - The change in trend after the start of COVID (from March 2020 onwards) vs before the start of the ARK trial (before February 2018), calculated from the effects above
- Step 3: I derived a weighted mean of the site-specific estimates using random-effects meta-analysis.

For the remainder of this chapter, the estimates derived in Step 2 above are described as "model 1". I chose this model as the primary analysis as it can be fit in both ARK and non-ARK sites, since trend estimates are based on fixed dates rather than an implementation date (which does not exist in non-ARK sites). Analyses included antibiotic DDDs from May/2015-June/2022 in all but 19 sites, where alternate start or end dates were used for reasons outlined in appendix **Table 5.8**.

Since analyses identified potential challenges with specialty coding (see Results), model 1 was also fit using Trust-level DDDs as an outcome. It was not possible to assess DDDs/admission as an outcome since admissions data was not available in non-ARK sites (**Table 5.2**).

Figure 5.2 Interrupted time series analyses used to estimate antibiotic prescribing trends and ARK implementation effects



[1] In model 0, the solid black line depicts the model that was fit in the ARK trial (data spanning February 2016-October 2020, with a COVID adjustment from March-June 2022 only), while the dotted black line depicts how this model was modified using IQVIA data (data spanning May 2015-June 2022, with a COVID adjustment from March 2022 onwards).

[2] Model 4 enabled the estimation of 4 separate changes in trend post- vs pre-implementation (solid and dotted red arrows). However, follow-up between February 2018 and implementation was limited (≥12 months in just 12 sites), as was follow-up between implementation and February 2020 (≥12 months in just 7 sites), and therefore this model was fitted only to a subset of sites with ≥8 months between implementation and Feb 2018, since residual confounding will be more likely when fewer months are available to derive prescribing trend estimates (i.e. due to inadequate adjustment for cyclical time effects). I also did not estimate the changes in trend depicted by the dotted red arrows.

Checking the stability of prescribing trends and ARK implementation effects in ARK sites using IQVIA data, May 2015-June 2022 (Models 2, 3, 4)

In ARK sites, prescribing trends were also estimated before and after the implementation date, rather than using fixed dates for all sites (as in model 1). Although these models could not be fit in non-ARK sites, they allowed the stability of trend estimates, the impact of COVID, and changes in trends pre- and post-implementation to be assessed in ARK sites and compared with model 1.

I considered 3 separate models, with estimates derived using the same 3-step analytic approach outlined above (i.e. fitting negative binomial models in each site, using an interrupted time series analysis to estimate prescribing trends, and combining site-specific effects using random-effects meta-analysis). As in model 1, these alternate models adjusted for COVID (from March 2020) and a cyclical time effect (up to February 2020) and used DDDs in general medicine and Trust-level DDDs as outcomes. Total DDDs/admission were not modelled as an outcome because admissions were only available from February 2016-October 2020 and these models used DDDs from IQVIA spanning May 2015-June 2022. The first of these models ("model 2") included an ARK implementation effect and estimated:

- A prescribing trend from May 2015 to implementation
- A prescribing trend from implementation to February 2020 (pre-COVID)
- A prescribing trend from March 2020-June 2022 (post-COVID)
- The immediate effect of COVID (change in IRR associated with March 2020)
- The immediate effect of ARK implementation (change in IRR associated with the month of implementation)
- The sustained effect of ARK implementation, calculated in two ways from the estimates above, including as a change in trend:
    - Post-implementation to February 2020 (pre-COVID) vs pre-implementation, per year
    - Post-implementation from March 2020 (post-COVID) vs pre-implementation, per year

The next alternative model ("model 3") differed from model 2 only in that it did not adjust for the immediate effect of ARK implementation. This forced the value of prescribing trend estimates immediately before and after implementation to join (with no step change).

The final alternative model ("model 4") differed from model 2 in two respects. First, it did not adjust for the immediate effect of the ARK implementation. Second, it estimated 2 separate pre-implementation prescribing trends, with one spanning May 2015-January 2018 (before the first main ARK trial site implemented) and the other spanning February 2018 to implementation.

All models are depicted in **Figure 5.2**, alongside the model used in the ARK trial ("model 0", described below). Since it was unclear which (if any) model fit the data best, models 1-3 were compared using the Akaike and Bayesian information criterion from each site, with a model considered "best" at $p<0.05$ only if AIC or BIC values were at least 3.84 below the next lowest value. Model 4 was not included in this comparison since it excluded 13 sites that lacked sufficient data (<8 months) to estimate trends between February 2018 and implementation with reasonable precision.

Estimating prescribing trends and ARK implementation effects using ARK trial methods with greater follow-up than was available in the ARK trial (Model 0)

To investigate how prescribing trends and ARK implementation effects varied with greater follow-up than was available in the ARK trial, I started by re-analysing the ARK trial data, making modifications to the trial methods as needed until the same model could be fit using IQVIA data. Since the adjusted and unadjusted ARK intervention effects were similar (Figure 3 in (1)), analyses were restricted to adjusted models.

As a first step, I analysed the co-primary outcome from the ARK trial (i.e. monthly total antibiotic DDDs in general medicine per adult general medical admission) using data from the trial (n=38 sites, including 7 sites outside England). As in the trial, this negative binomial model adjusted for COVID (March-June 2020 only) and a cyclical time effect up to October 2020, and this was used to estimate:

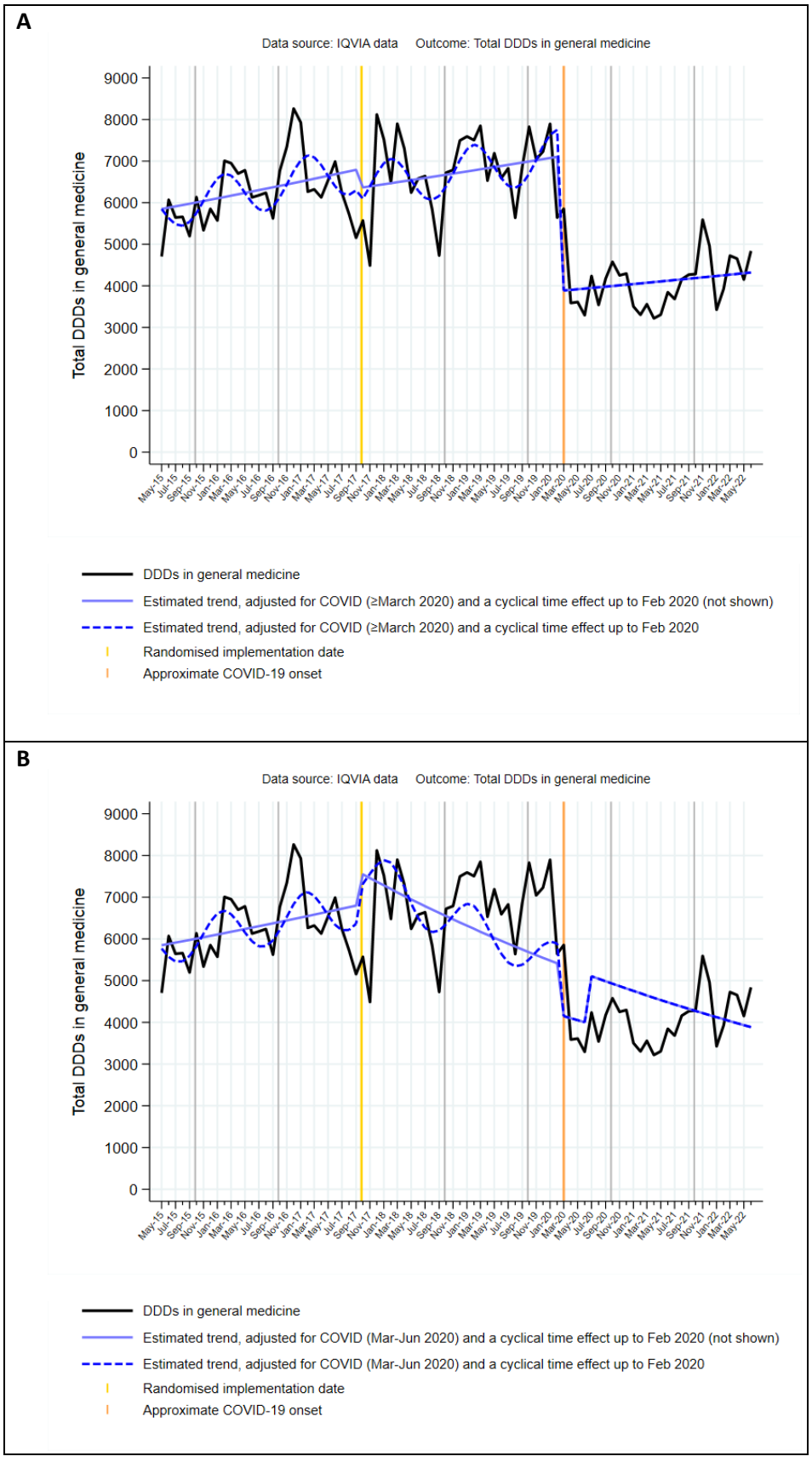
- The pre-implementation trend (change in IRR per year, from February 2016 to implementation)
- The post-implementation trend (change in IRR per year, from implementation to October 2020)
- The immediate effect of COVID
- The immediate effect of ARK implementation
- The sustained effect of ARK implementation, calculated from the estimates as above as a change in trend post- vs pre-implementation (to October 2020), per year

For the remainder of this chapter, these estimates are described as "model 0" (**Figure 5.2**). Unlike models 1-4, model 0 only estimated a single post-implementation trend, as the trial follow-up ended in October 2020, leaving too little data to estimate a separate post-COVID trend.

Next, since there was no consistent seasonal variation in antibiotic prescribing after the onset of COVID, this model was fit without a cyclical time effect after February 2020, as per models 1-4 above. I then restricted the analysis of the ARK trial data to just 29 sites in England (i.e. sites with available IQVIA data, as described in **Table 5.1**). Finally, I modelled DDDs in general medicine without normalising for admissions.

I then analysed IQVIA DDDs among ARK sites in England, first as DDDs/admission (using admissions from the ARK trial, which were only available from February 2016-October 2020) and then without normalising for admissions (initially restricted to February 2016-October 2020, and then expanded to May 2015-June 2022). Models of DDDs in general medicine adjusted for COVID and included a cyclical time effect up to February 2020, however the COVID adjustment was modified slightly and fit as a binary factor =1 from March 2020 onwards (as in models 1-4) rather than =1 from March-June 2020 (as in the ARK trial), as this produced better fitting models (based on a comparison of AIC and BIC values, example shown in **Figure 5.3**).

Figure 5.3 Example: COVID adjustment from March 2020 onwards (used in Models 1-4) (A) vs March-June 2020 only (used in the ARK trial) (B)



5.3 Results

5.3.1 Detecting possible changes in specialty coding in IQVIA data

Possible changes in specialty coding were detected in nearly one-third of sites (37/123), including 9/31 ARK sites and 28/92 non-ARK sites (**Table 5.3**). **Figure 5.4** shows a representative example, where DDDs in general medicine drop suddenly in July 2018 (panel A). Importantly, this drop coincides with a comparable opposite increase in DDDs in specialties outside general medicine, without any overall change in Trust-level DDDs, suggesting an administrative change in specialty coding, rather than a genuine drop in DDDs within general medicine. When plotted as a fraction of DDDs across all specialties, the discontinuity is particularly apparent (panel B). In most sites with this problem (29/37), I adjusted for this likely change in specialty coding when deriving prescribing trend estimates (panel C), as per the rules described in **Table 5.3**. For comparison, **Figure 5.5** shows a site with no such change identified. Although flagging a potential change in specialty coding is ultimately a subjective decision requiring judgement, it is reassuring that the sites in which a change was suspected often had the largest standard deviation in the monthly distribution of general medical DDDs as a proportion of Trust-level DDDs (**Figure 5.6**).

Table 5.3 Rules applied to limit potential confounding from changes in specialty coding, IQVIA data

Rule	Number of ARK sites (N=31)	Number of non-ARK sites (N=92)	All sites (N=123)
1. If a suspected change in specialty coding occurred within 12 months of the planned analysis start date (May 2015), the start date was pushed forward to exclude data before the change.	0	7[2]	7
2. If a suspected change in specialty coding occurred within 12 months of the planned analysis end date (June 2022), the analysis end date was pulled back to exclude data after the change.	1	2[2]	3
3. If a suspected change in specialty coding occurred between May 2016 (planned start date + 1 year) and June 2021 (planned end date – 1 year) and the median monthly general medicine DDDs changed substantially (i.e. by a factor of 10 or more, before vs after the step change), then the site was excluded, as this was considered likely to prevent valid estimation of antibiotic trends.	1	1	2
4. If applying the rules above did not remove the suspected change in specialty coding, the step change was adjusted for by including a binary factor (=0 before the step change, and =1 afterwards) in the negative binomial models used to estimate antibiotic use trends (see Statistical Methods below), as in **Figure 5.2** panel C.[1]	7	22[2]	29
Total number of sites with possible changes in specialty coding	9 (29%)	28 (33%)	37 (30%)[2]

[1] In 1/7 ARK sites and 3/22 non-ARK sites the shift in specialty coding was likely due to a reorganisation of services after the onset of the COVID pandemic, as the drop in DDDs in March 2020 was sustained thereafter, while Trust-level DDDs recovered to pre-COVID levels.

[2] In 3 non-ARK sites both rule #1 and rule #4 were applied, and in 1 non-ARK site both rule #2 and rule #4 were applied.

Figure 5.4 Shift in specialty coding in July 2018 at University Hospitals Sussex, seen for DDDs in general medicine (A) and in general medicine as a percentage of Trust-level DDDs (B), with trend estimates from Model 1 estimated on DDDs in general medicine (C)

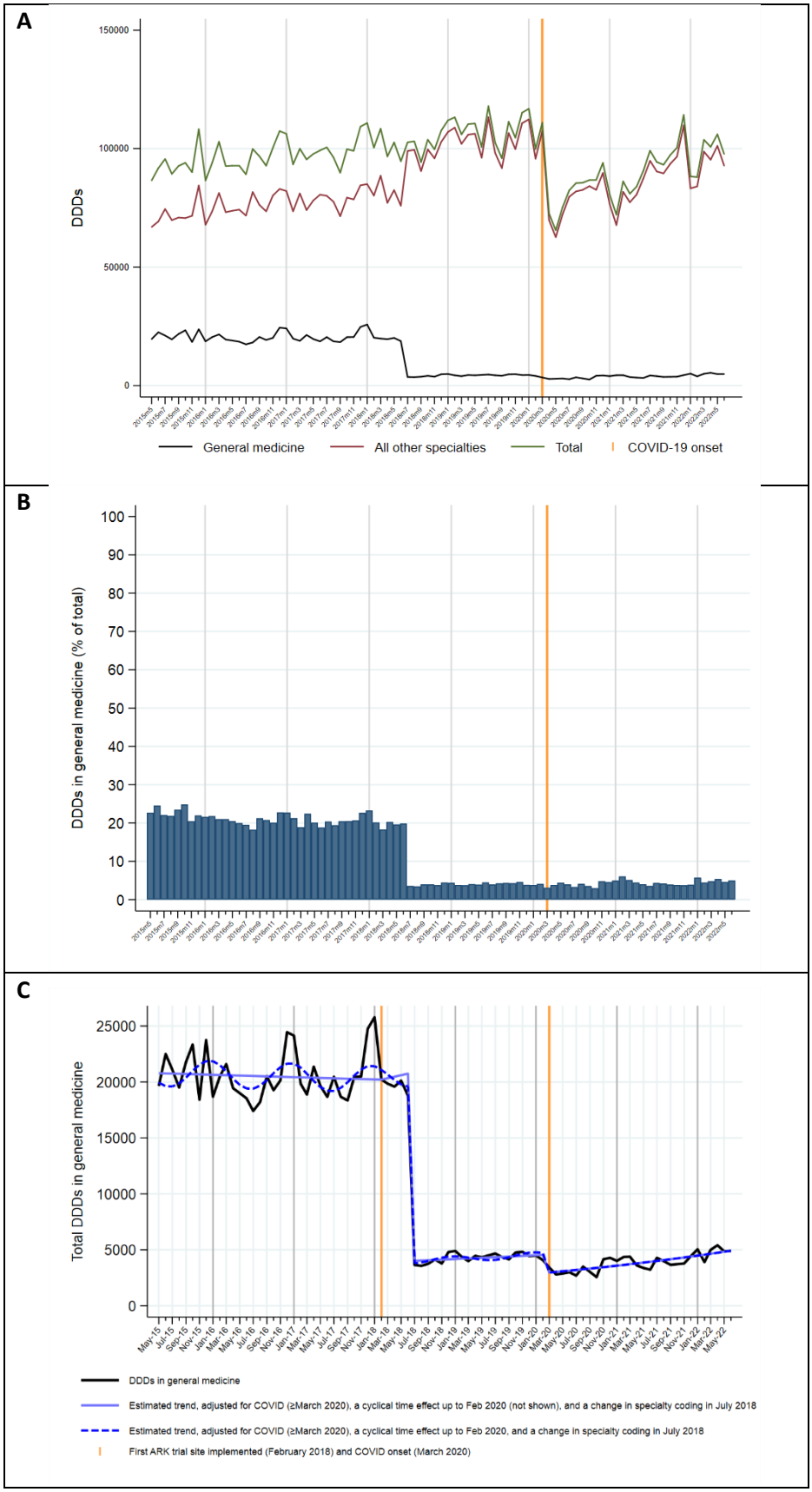


Figure 5.5 Prescribing trends at North Tees and Hartlepool, among DDDs general medicine (A) and in general medicine as a percentage of Trust-level DDDs (B), with trend estimates from Model 1 estimated on DDDs in general medicine (C)

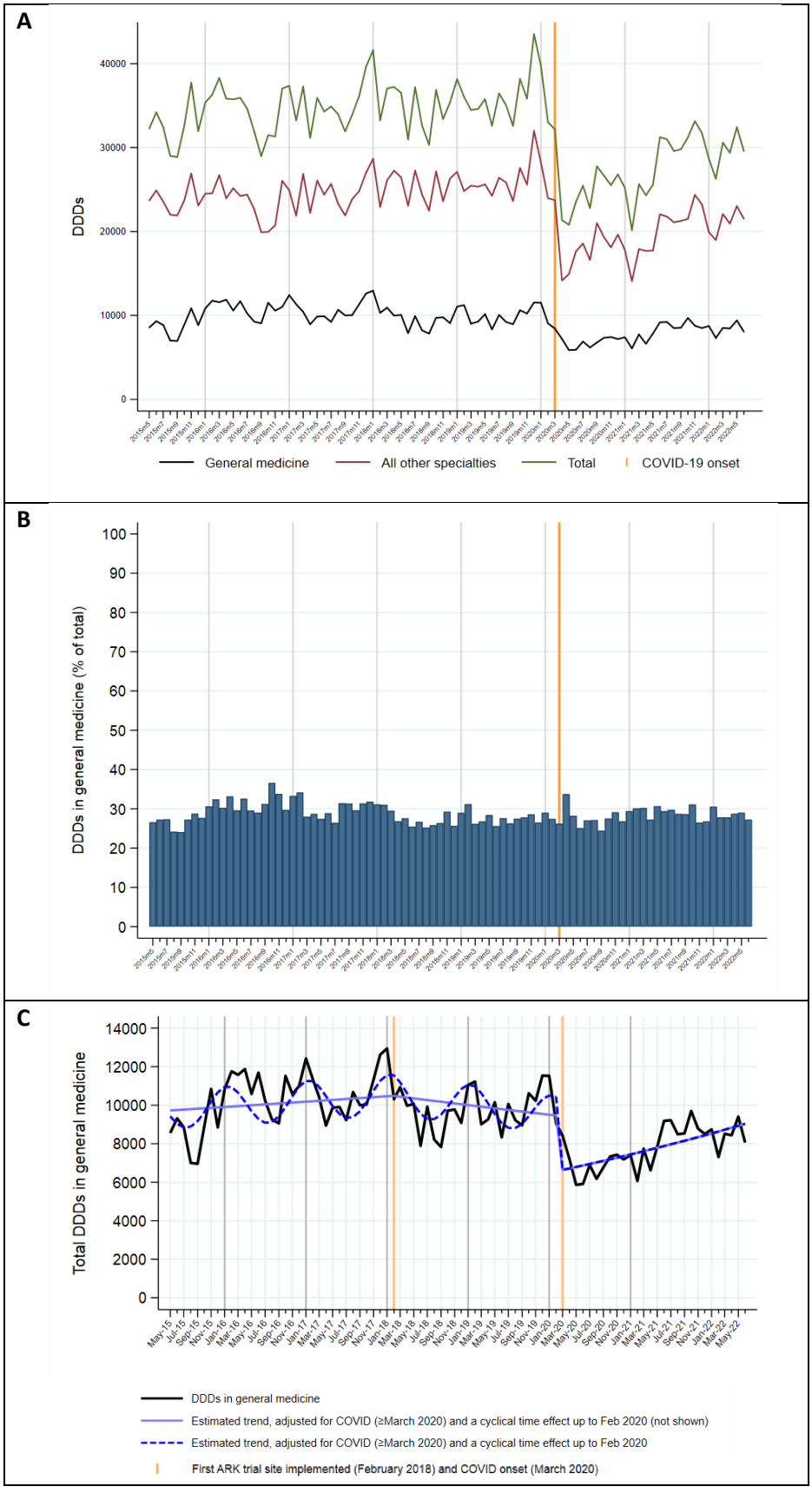
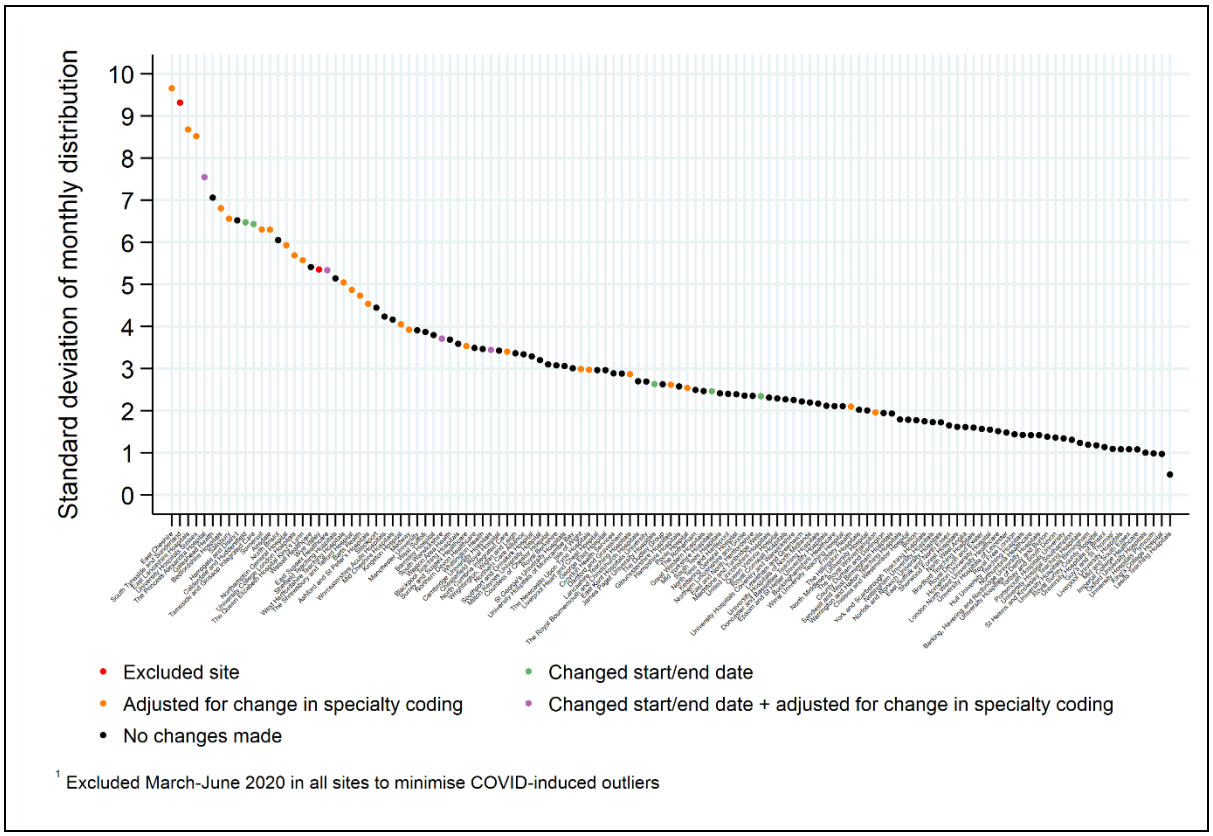
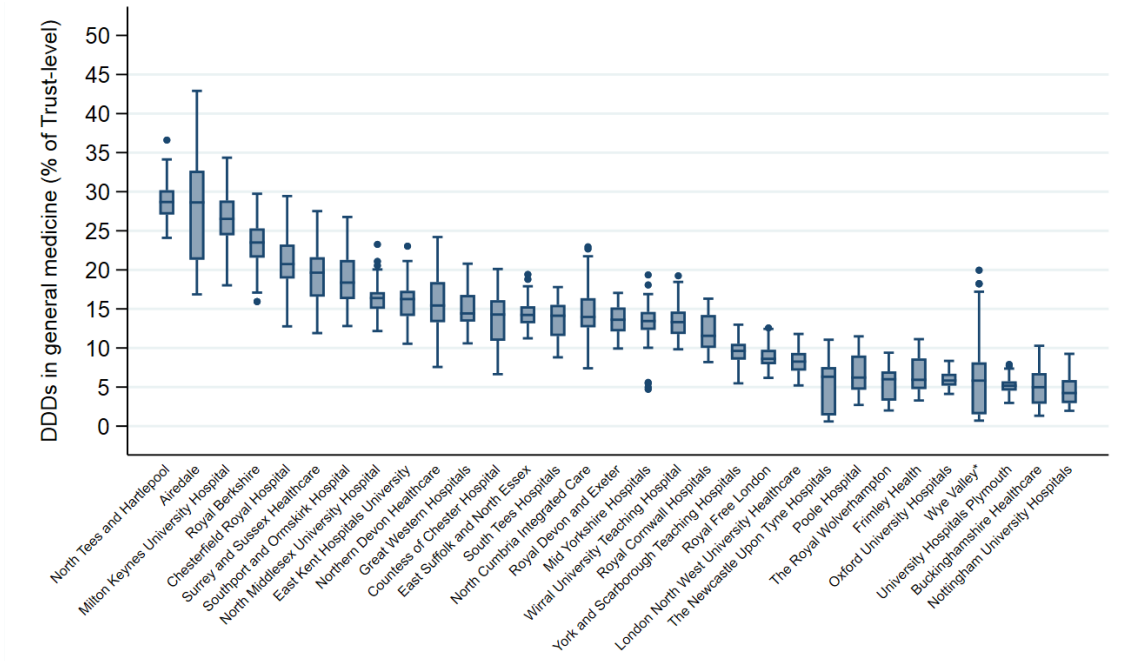


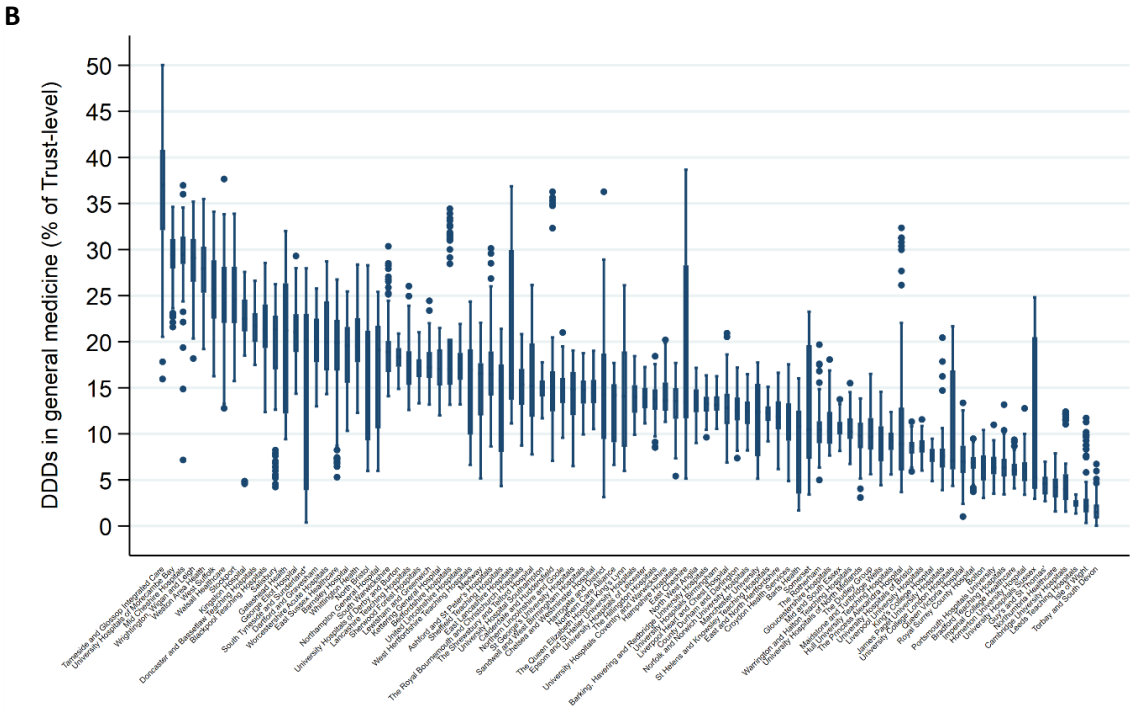
Figure 5.6 Sites with possible changes in specialty coding, ordered by the standard deviation of the monthly distribution of DDDs in general medicine as a percentage of Trust-level DDDs



Having observed these changes in specialty coding, and reflecting on the somewhat subjective nature of the rules I developed to address them, I also analysed Trust-level DDDs (rather than DDDs assigned to general medicine in the IQVIA dataset) as an outcome in sensitivity analyses, though in most sites DDDs in general medicine accounted for a relatively small share of Trust-level DDDs (monthly median ranged from 4.2%-28.7% in ARK sites and 1.5%-37.0% in non-ARK sites) (**Figure 5.7**).

Figure 5.7 Monthly total DDDs in general medicine as a percentage of Trust-level DDDs among 31 ARK sites (A) and 92 non-ARK sites (B), IQVIA data





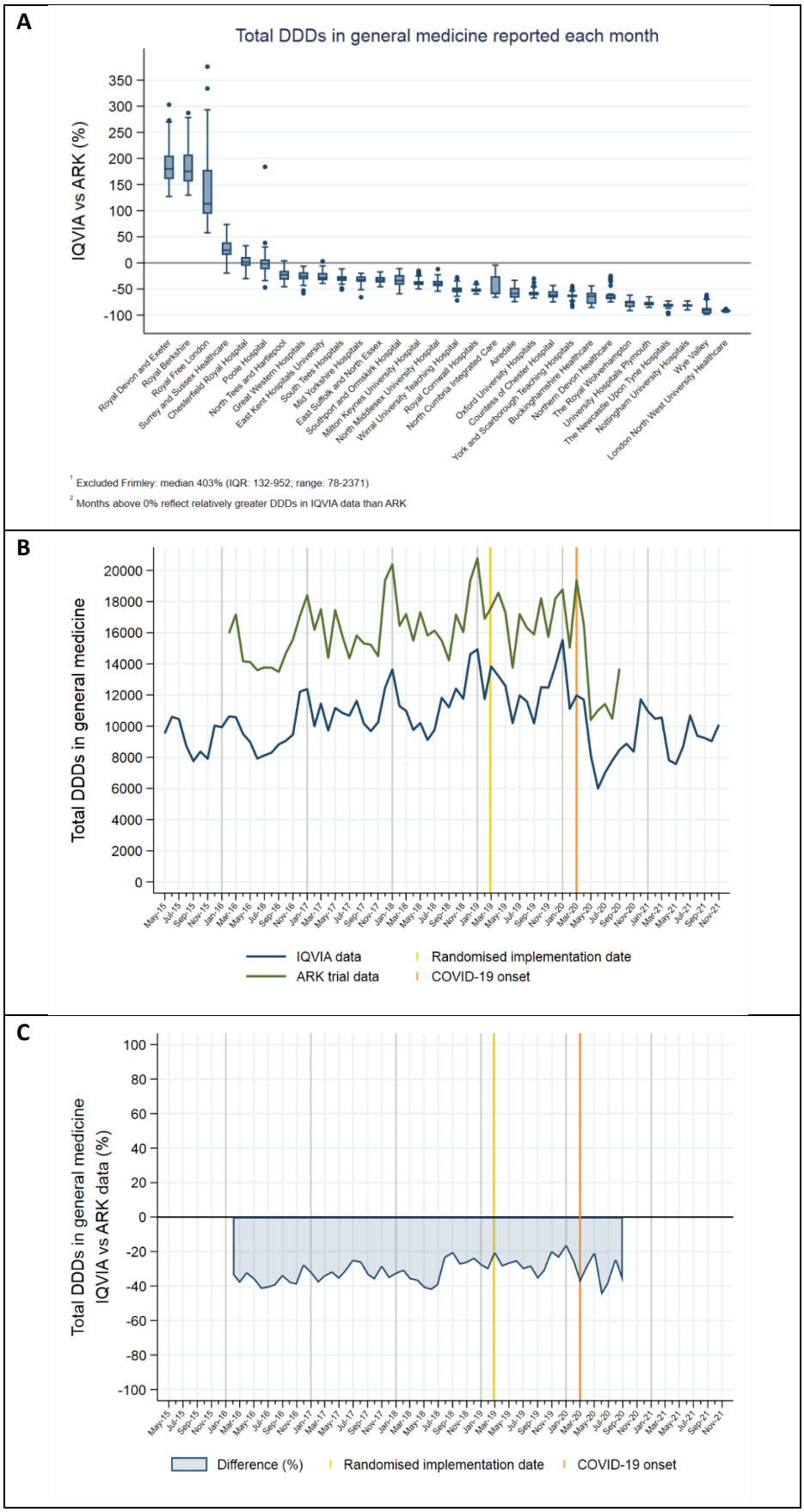
*Excluded South Tyneside and Sunderland from analyses of DDDs in general medicine due to a change in specialty coding

Note: excluded March-June 2020 in all sites to minimise COVID-induced outliers.

5.3.2 Comparing ARK vs IQVIA data

The absolute magnitude of total DDDs reported to have been used in general medicine was lower in the IQVIA data than in the ARK trial data in most sites (23/30, 77%) (**Figure 5.8**), suggesting the algorithms used to define general medicine likely varied between the two data sources. In 4 sites, relative differences in some months exceeded a factor of 10. However the relative magnitude was fairly constant over time in most sites (i.e. showed the same spikes and dips, as shown for one example site in panels B and C), with exceptions (n=7/30) being due to likely changes in specialty coding in the IQVIA data as described above.

Figure 5.8 Differences in the magnitude of DDDs reported as being used in general medicine in IQVIA vs ARK trial data in all ARK sites (A) and in East Suffolk and North Essex specifically (B and C)



5.3.3 Prescribing trends in ARK vs non-ARK sites (Model 1)

Prescribing trend estimates for models 1-4 are summarised in **Table 5.4** and **Figure 5.9** (general medicine) and **Table 5.5** and **Figure 5.10** (Trust-level).

In model 1 (rows shaded blue), analyses of IQVIA DDDs in general medicine showed similar estimates between ARK and non-ARK sites for the following:

- Prescribing trends between May 2015-January 2018, which increased 5-6% per year
- Prescribing trends between February 2018 and February 2020, where there was no evidence of change
- Immediate COVID effects, which showed DDDs declining by 30-32% in March 2020

Differences in estimates were observed between ARK and non-ARK sites for the following:

- Prescribing trends from March 2020 onwards showed slightly greater year-on-year increases in DDDs among ARK sites, resulting in a change in trend after March 2020 vs before February 2018 (pre-implementation in all ARK sites) of +5.7% per year (95% CI: +0.1, +11.6), compared with no evidence of change in trend among non-ARK sites (-1.4% per year, 95% CI: -4.6, +1.9) (appendix **Figure 5.12**).

Analyses of Trust-level DDDs (**Table 5.5**, **Figure 5.10**) yielded similar results, except an increasing change in year-on-year trend after March 2020 vs before February 2018 was observed in both ARK (+7.1%, 95% CI: 4.3, 9.9) and non-ARK sites (+4.0%, 95% CI: +2.5, +5.6).

5.3.4 Stability of prescribing trends and ARK implementation effects in ARK sites (Model 1-4)

Estimates for the effect of ARK implementation on IQVIA DDDs in general medicine from models 2, 3, and 4 in comparison to model 1 (the only model which could be directly compared with non-ARK

sites) are summarised in **Table 5.4** and **Figure 5.9** and shown by Trust in appendix **Figure 5.13**, **Figure 5.14**, and **Figure 5.15** respectively.

Broadly similar estimates were observed across models 1-4 for the following:

- Prescribing trends before implementation (models 2, 3) and before February 2018 (model 1, 4) all showed small increases in DDDs between 3-8% per year
- Immediate COVID effects showed DDDs declining between 31-33% in March 2020
- Prescribing trends from March 2020 onwards showed similar year-on-year increases in DDDs in general medicine between +8 to +11% per year (models 2-4) (5% per year in model 1)

Differences in estimates were observed for the following:

- Model 4 (restricted to only 17 ARK sites with sufficient data from February 2018 to implementation and from implementation to February 2020 to fit separate trends) showed year-on-year declines in DDDs in general medicine between February 2018 and implementation (IRR per year: 0.91, 95% CI; 0.85, 0.98), which were not seen in model 1 (trends from February 2018-February 2020 IRR per year: 0.98, 95% CI: 0.94, 1.02). However, 3 sites had just 8 months available to derive this estimate, spanning February-September 2018 (i.e. resulting in trend estimates that begin in winter when antibiotic use is generally at a peak, and ending in the late summer when antibiotic use is at a low). Since confounding adjustment for cyclical time effects is more likely to be inadequate when trend estimates are based on fewer months, the model was also fit excluding these 3 sites and this resulted in a trend estimate that was consistent with no decline (IRR per year: 0.95, 95% CI: 0.90, 1.01, p=0.121).
- Models 2 and 4 showed no evidence of a change in trend between implementation and February 2020, nor was there evidence of a change in trend post-implementation to February 2020 vs pre-implementation. However, model 3 estimated a small non-significant

decrease in DDDs in general medicine between implementation and February 2020 (IRR per year: 0.95, 95% CI; 0.90, 1.01), and this contributed to a change in year-on-year trend post-implementation to February 2020 vs before February 2018 of -7.7% per year (95% CI: -13.5, -1.5). This change in trend may be because model 3 did not adjust for the immediate implementation effect, which was associated with a -7.0% (95% CI: -11.8, -1.8) decline in DDDs in general medicine in model 1 (i.e. by not adjusting for the immediate implementation effect, the pre- and post-implementation trends were forced to join at implementation, without accounting for the -7.0% step change at implementation, and this consequently led to the trend estimate from implementation to February 2020 showing a greater year-on-year decline relative to model 2).

- There was a comparable change in year-on-year trend after COVID (after March 2020) vs pre-implementation (models 2, 3) or before February 2018 (model 1), varying between an increase of 6-8% per year. However, model 4 did not find any evidence of a change in year-on-year trend after COVID (March 2020) vs before February 2018 (+0.3%, 95% CI; -7.4, +8.6).

Trend estimates from models of Trust-level DDDs were broadly similar (**Table 5.5**, **Figure 5.10**), though there was no evidence of an immediate implementation effect in model 2 (Δ IRR: -0.9%, 95% CI: -3.2, +1.4), and the changes in trend estimates were consistent across all 4 models, with DDDs increasing 7-8% after March 2020 vs before February 2018 (models 1,4) or pre-implementation (models 2,3).

Site-specific AIC and BIC estimates from models 1, 2 and 3 for DDDs in general medicine were very similar (**Table 5.6**). In 23/30 sites, all three AIC values were within 3.84 of each other (chi-squared p=0.05 on 1 df), suggesting they were not distinguishable statistically. Based on AIC, model 1 was considered “best” at $p<0.05$ in just 4/30 sites, model 2 was considered “best” in 3/30 sites, and model 3 was not considered best in any site.

Table 5.4 Impact of the ARK intervention and prescribing trends in ARK vs non-ARK sites in England, using DDDs in general medicine (Models 0-4)

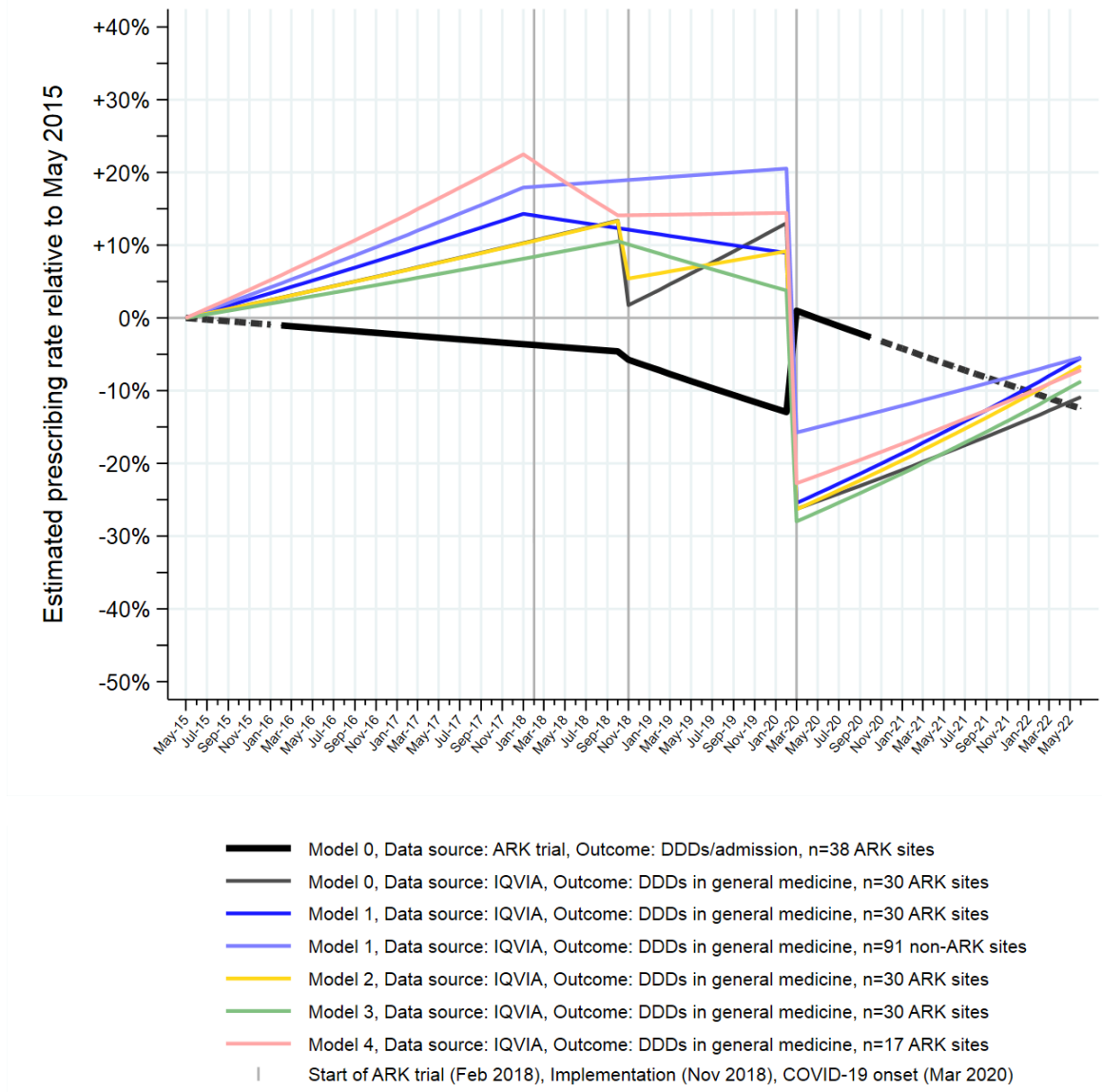
Outcome	Model and no. ARK or non-ARK sites in IQVIA data	Trend per year				Immediate effect		Change in trend per year	
		Before Feb-18 (model 1,4) or pre-implementation (model 0,2,3)	From Feb-18 to Feb-20 (model 1) or to implementation (model 4)	Post-implementation to Feb-20 (model 2-4) or end of follow-up (model 0)	From Mar-20 onwards (post-COVID)	COVID onset in Mar-20	Implementation	Post-implementation to Feb-20 (model 2-4) or onwards (model 0) vs pre-implementation[3]	After Mar-2020 vs pre-Feb-18 (model 1,4) or pre-implementation (model 2,3)
		IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	% Δ IRR (95% CI)	% Δ IRR (95% CI)	% Δ IRR (95% CI)
ARK	**0.** 38 ARK	0.99 (0.97,1.01) p=0.205		0.94 (0.90,0.97) **p=0.001**		1.16 (1.10,1.22) **p<0.001**	-1.0% (-4.0,+2.1) p=0.540	-4.8% (-9.1,-0.2) p=**0.042**	
IQVIA DDDs in general medicine[2]	**0.** 30 ARK	1.04 (1.01,1.07) p=**0.008**		1.09 (1.05,1.13) p<**0.001**		0.65 (0.58,0.73) p<**0.001**	-10.3% (-15.0,-5.2) p<**0.001**	+4.7% (-0.1,+9.7) p=0.054	
	1. 30 ARK	1.05 (1.02,1.09) p=**0.003**	0.98 (0.94,1.02) p=0.238		1.11 (1.06,1.17) p<**0.001**	0.68 (0.61,0.76) p<**0.001**			+5.7% (+0.1,+11.6) p=**0.045**
	1. 91 non-ARK	1.06 (1.05,1.08) p<**0.001**	1.01 (0.99,1.03) p=0.349		1.05 (1.02,1.08) p<**0.001**	0.70 (0.67,0.73) p<**0.001**			-1.4% (-4.6,+1.9) p=0.395
	2. 30 ARK	1.04 (1.01,1.07) p=**0.009**		1.03 (0.98,1.08) p=0.282	1.11 (1.06,1.16) p<**0.001**	0.67 (0.61,0.74) p<**0.001**	-7.0% (-11.8,-1.8) p=**0.009**	-0.5% (-6.3,+5.7) p=0.873	+7.0% (+1.1,+13.3) p=**0.020**
	3. 30 ARK	1.03 (1.00,1.06) p=**0.041**		0.95 (0.90,1.01) p=0.088	1.11 (1.06,1.16) p<**0.001**	0.69 (0.63,0.77) p<**0.001**		-7.7% (-13.5,-1.5) p=**0.016**	+7.9% (+2.1,+14.0) p=**0.007**
	4. 17 ARK	1.08 (1.04,1.12) p<**0.001**	0.91 (0.85,0.98) p=**0.011**	1.00 (0.89,1.13) p=0.972	1.08 (1.02,1.16) p=**0.015**	0.68 (0.58,0.78) p<**0.001**		-7.0% (-16.9,+4.1) p=0.209	+0.3% (-7.4,+8.6) p=0.945

[1] Results in top row are from the ARK trial(105). The outcome is DDDs/admission, assessed from Feb 2016-Oct 2020, with site-level estimates adjusted for COVID (Mar-Jun 2020) and a cyclical time effect (up to Oct 2020). All other estimates are IQVIA DDDs (not normalised for admissions as data not available).

[2] Outcomes measured monthly from May 2015-June 2022, with site-level estimates adjusted for COVID (from March 2020 onwards) and a cyclical time effect (up to February 2020). The model 0 results are also shown in **Table 5.7**.

[3] In this column, model 4 "pre-implementation" refers to before February 2018. IRR=incidence rate ratio per year.

Figure 5.9 Estimated antibiotic prescribing rates derived using DDDs in general medicine (Models 0-4)



Note: These predictions are from **Table 5.4**. The implementation date was fixed at Nov 2018 to facilitate model comparisons. The dotted black line has been extrapolated from trend estimates derived using ARK trial data spanning Feb 2016-Oct 2020.

Table 5.5 Impact of the ARK intervention and prescribing trends in ARK vs non-ARK sites in England, using IQVIA DDDs across all specialties (Models 0-4)

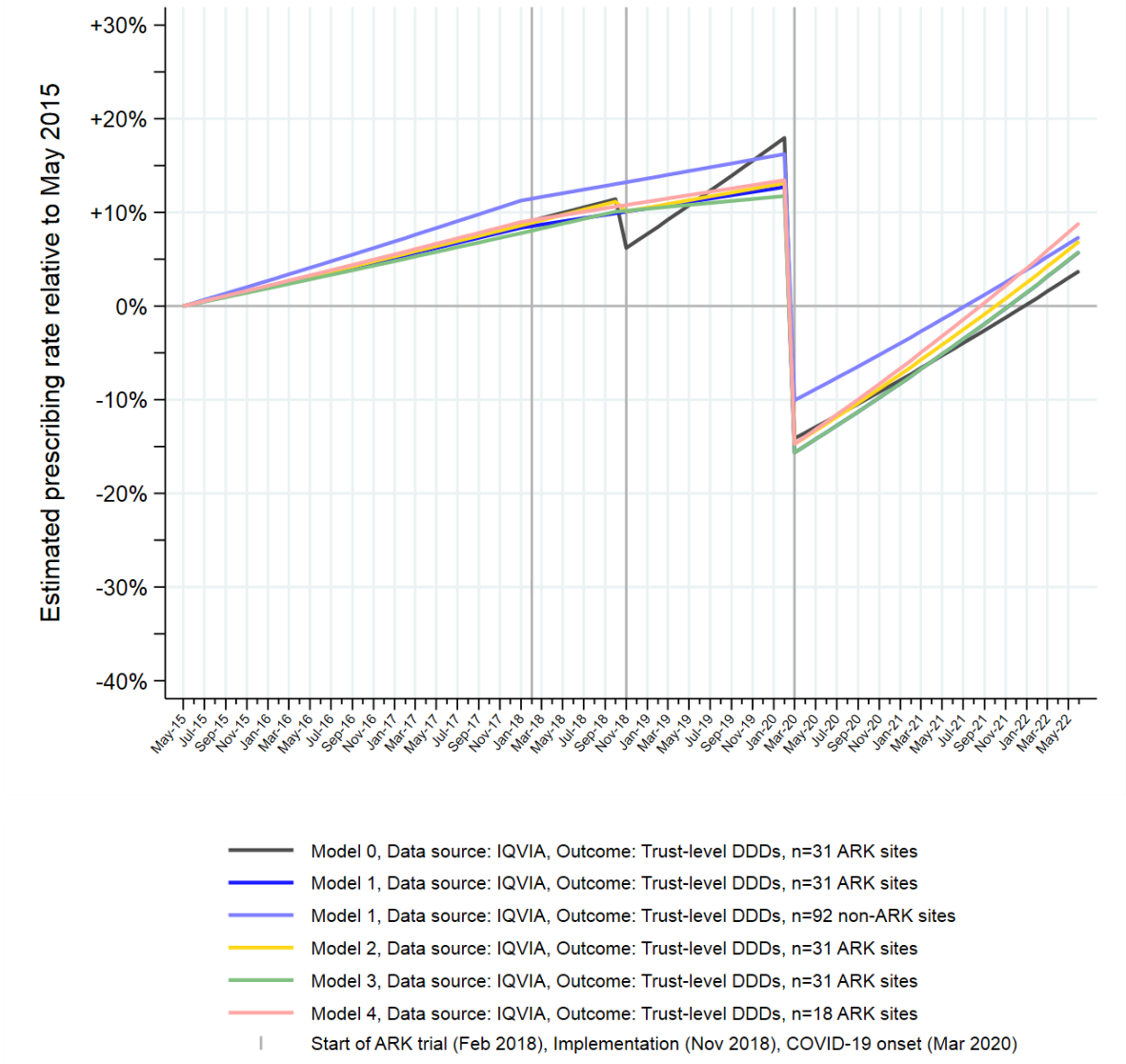
Outcome	Model and no. ARK or non-ARK sites in IQVIA data	Trend per year				Immediate effect		Change in trend per year	
		Before Feb-18 (model 1,4) or pre-implementation (model 0,2,3)	From Feb-18 to Feb-20 (model 1) or to implementation (model 4)	Post-implementation to Feb-20 (model 2-4) or Jun-22 (model 0)	From Mar-20 onwards (post-COVID)	COVID onset in Mar-20	Implementation	Post-implementation to Feb-20 (model 2-4) or Jun-22 (model 0) vs pre-implementation[2]	After Mar-2020 vs pre-Feb-18 (model 1,4) or pre-implementation (model 2,3)
		IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	% Δ IRR (95% CI)	% Δ IRR (95% CI)	% Δ IRR (95% CI)
IQVIA DDDs across all specialties[1]	**0.** 31 ARK	1.03 (1.02,1.05) p<**0.001**		1.09 (1.07,1.10) p<**0.001**		0.73 (0.71,0.75) p<**0.001**	-4.7% (-6.8,-2.5) p<**0.001**	+5.2% (+2.7,+7.7) p<**0.001**	
	1. 31 ARK	1.03 (1.01,1.05) p=**0.001**	1.02 (1.00,1.04) p=**0.042**		1.11 (1.09,1.12) p<**0.001**	0.75 (0.73,0.77) p<**0.001**			+7.1% (+4.3,+9.9) p<**0.001**
	1. 92 non-ARK	1.04 (1.03,1.05) p<**0.001**	1.02 (1.01,1.03) p<**0.001**		1.08 (1.07,1.09) p<**0.001**	0.77 (0.76,0.79) p<**0.001**			+4.0% (+2.5,+5.6) p<**0.001**
	2. 31 ARK	1.03 (1.02,1.05) p<**0.001**		1.02 (0.99,1.05) p=0.117	1.11 (1.09,1.12) p<**0.001**	0.75 (0.73,0.77) p<**0.001**	-0.9% (-3.2,+1.4) p=0.425	-1.1% (-3.9,+1.8) p=0.462	+6.9% (+4.1,+9.7) p<**0.001**
	3. 31 ARK	1.03 (1.01,1.04) p<**0.001**		1.01 (0.98,1.04) p=0.421	1.11 (1.09,1.12) p<**0.001**	0.75 (0.73,0.78) p<**0.001**		-1.7% (-4.6,+1.4) p=0.278	+7.3% (+4.6,+10.0) p<**0.001**
	4. 18 ARK	1.03 (1.01,1.05) p=**0.002**	1.02 (0.99,1.05) p=0.209	1.02 (0.98,1.06) p=0.316	1.11 (1.09,1.14) p<**0.001**	0.75 (0.73,0.78) p<**0.001**		-1.9% (-6.3,+2.8) p=0.427	+7.7% (+4.6,+10.8) p<**0.001**

[1] Outcome (Trust-level IQVIA DDDs, not normalised for admissions as this data was not available) measured monthly from May 2015-June 2022, with site-level estimates adjusted for COVID (from March 2020 onwards) and a cyclical time effect (up to February 2020). The model 0 results are also shown in **Table 5.7**.

[2] In this column, model 4 "pre-implementation" refers to before February 2018.

IRR=incidence rate ratio per year

Figure 5.10 Estimated antibiotic prescribing rates derived using Trust-level DDDs (Models 0-4)



Note: These predictions are from **Table 5.5**. The implementation date was fixed at Nov 2018 to facilitate model comparisons.

Table 5.6 IQVIA DDDs in general medicine: Akaike and Bayesian information criterion from models 1-3 in 30 ARK sites in England

ARK trial site	Akaike information criterion (AIC)					Bayesian information criterion (BIC)				
	Model 1	Model 2	Model 3	Lowest – next lowest	Best model[1]	Model 1	Model 2	Model 3	Lowest – next lowest	Best model[1]
Great Western Hospitals	1330.37	1350.27	1354.00	-19.89	Model 2	1350.01	1372.36	1373.64	-22.35	Model 2
North Cumbria Integrated Care	1372.98	1356.46	1369.93	-13.47	Model 1	1392.61	1378.55	1389.56	-11.02	Model 1
Milton Keynes University Hospital	1422.68	1431.46	1431.09	-8.42	Model 2	1442.31	1453.55	1450.73	-8.42	Model 2
Royal Cornwall Hospitals	1326.96	1322.28	1330.46	-4.68	Model 1	1346.60	1344.37	1350.09	-2.22	N/A
London North West University	1410.69	1414.71	1425.05	-4.02	Model 2	1430.32	1436.79	1444.68	-6.47	Model 2
Wirral University Teaching Hospital	1378.57	1374.58	1390.07	-3.99	Model 1	1400.66	1399.13	1412.15	-1.54	N/A
Southport and Ormskirk Hospital	1345.86	1341.96	1346.16	-3.90	Model 1	1367.95	1366.51	1368.25	-1.44	N/A
North Tees and Hartlepool	1445.19	1450.01	1448.60	-3.41	N/A	1464.83	1472.10	1468.24	-3.41	N/A
The Royal Wolverhampton	1288.20	1293.03	1291.56	-3.36	N/A	1310.29	1317.58	1313.65	-3.36	N/A
East Kent Hospitals University	1461.01	1465.07	1463.77	-2.77	N/A	1480.64	1487.15	1483.41	-2.77	N/A
Buckinghamshire Healthcare	1277.14	1274.55	1277.14	-2.59	N/A	1296.77	1296.64	1296.77	-0.14	N/A
Royal Devon and Exeter	1125.70	1123.73	1126.63	-1.97	N/A	1143.80	1144.09	1144.73	-0.30	N/A
Countess of Chester Hospital	1353.82	1353.85	1351.85	-1.96	N/A	1373.45	1375.93	1371.49	-1.96	N/A
Airedale	1342.76	1340.31	1342.06	-1.75	N/A	1364.85	1364.85	1364.15	-0.70	N/A
Poole Hospital	919.36	922.61	921.08	-1.72	N/A	938.93	944.36	940.65	-1.72	N/A
Northern Devon Healthcare	1258.89	1261.02	1260.38	-1.49	N/A	1280.66	1285.21	1282.15	-1.49	N/A
York and Scarborough Teaching	1387.53	1388.93	1388.32	-0.78	N/A	1407.17	1411.02	1407.95	-0.78	N/A
Mid Yorkshire Hospitals	1351.91	1352.98	1351.12	-0.78	N/A	1371.16	1374.64	1370.38	-0.78	N/A
Royal Free London	1445.70	1443.86	1444.46	-0.60	N/A	1465.34	1465.95	1464.10	-1.24	N/A
East Suffolk and North Essex	1497.07	1498.17	1496.57	-0.50	N/A	1516.70	1520.26	1516.21	-0.50	N/A
Oxford University Hospitals	1331.52	1326.49	1326.94	-0.45	N/A	1351.15	1348.58	1346.58	-2.01	N/A
Chesterfield Royal Hospital	1444.29	1445.80	1444.59	-0.30	N/A	1463.92	1467.89	1464.23	-0.30	N/A

ARK trial site	Akaike information criterion (AIC)					Bayesian information criterion (BIC)				
	Model 1	Model 2	Model 3	Lowest – next lowest	Best model[1]	Model 1	Model 2	Model 3	Lowest – next lowest	Best model[1]
The Newcastle Upon Tyne Hospitals	1349.12	1349.46	1348.84	-0.28	N/A	1368.76	1371.55	1368.47	-0.28	N/A
North Middlesex University Hospital	1473.36	1475.02	1473.08	-0.28	N/A	1492.99	1497.11	1492.72	-0.28	N/A
Frimley Health	1416.56	1415.94	1415.79	-0.15	N/A	1438.65	1440.48	1437.88	-0.77	N/A
Royal Berkshire	1454.68	1454.65	1454.57	-0.08	N/A	1474.32	1476.74	1474.21	-0.11	N/A
University Hospitals Plymouth	1298.08	1296.88	1296.94	-0.07	N/A	1320.17	1321.42	1319.03	-1.14	N/A
Surrey and Sussex Healthcare	1448.75	1447.99	1448.05	-0.06	N/A	1470.84	1472.53	1470.14	-0.70	N/A
South Tees Hospitals	1390.93	1392.01	1390.94	-0.01	N/A	1410.57	1414.10	1410.58	-0.01	N/A
Nottingham University Hospitals	1397.53	1399.36	1397.52	-0.01	N/A	1417.16	1421.45	1417.15	-0.01	N/A

[1] i.e. A model was considered “best” at $p<0.05$ if its AIC or BIC value was at least 3.84 below the next lowest value.

5.3.5 Prescribing trends and implementation effects derived using ARK trial methods (Model 0)

As reported in the ARK trial(105), after adjusting for a cyclical time effect (February 2016-October 2020) and the impact of COVID (March-June 2020), the intervention was associated with an immediate change in total DDDs/admission of -1.0% (95% CI: -4.0, +2.1) and a sustained year-on-year change in trend post- vs pre-implementation through October 2020 of -4.8% per year (95% CI: -9.1, -0.2) in 38 NHS Trusts. Modifications to this model are presented in **Table 5.7** and **Figure 5.11**.

The first modification involved fitting the same model to DDDs/admission but only adjusting for a cyclical time effect up to February 2020, rather than October 2020, as there was no clear seasonal variation after the onset of COVID in March 2020. This had a negligible impact on the immediate implementation effect (+0.4% per year, 95% CI; -2.6, +3.4) but led to a greater sustained change in year-on-year trend post- vs pre-implementation (-7.9% per year, 95% CI: -12.2, -3.4) (appendix **Figure 5.16**). Restricting the model for DDDs/admission to sites in England (n=29) suggests the English ARK sites achieved even greater sustained reductions in DDDs/admission (-10.3% per year, 95% CI: -15.2, -5.2) than when all ARK sites (including Scotland, Wales and Northern Ireland) were included, again with no evidence of an immediate implementation effect (+1.5%, 95% CI: -1.8, +5.0) (appendix **Figure 5.17**). Fitting this model without normalising for admissions, i.e. to total DDDs reported as used in general medicine by ARK sites within the trial, resulted in a similar change in year-on-year trend post- vs pre-implementation (-10.8% per year, 95% CI: -16.1, -5.2), but resulted in evidence of an immediate increase in DDDs at implementation (+4.2%, 95% CI: +0.8, +7.8) that was not seen in models of DDDs/admission using this ARK trial data (appendix **Figure 5.18**). However, notably without normalising for admissions total DDDs dropped significantly with the onset of the COVID pandemic, whereas DDDs/admission increased significantly.

Analyses of IQVIA DDDs/admission between February 2016-October 2020 (i.e. the timeframe analysed in the original ARK trial) showed ARK implementation was associated with a non-significant immediate change of -4.3% (95% CI: -9.2, +0.7) and a change in year-on-year trend post- vs pre-

implementation of -6.2% per year (95% CI: -12.9, +1.1) (appendix **Figure 5.19**). When this model was fitted without normalising for admissions estimates attenuated slightly, remaining with no evidence of immediate (-3.1%, 95% CI: -7.7, +1.7) or sustained implementation effect (-3.7% per year, 95% CI: -10.8, +4.0) (appendix **Figure 5.20**).

Extending follow-up to include IQVIA DDDs in general medicine from May 2015-June 2022, not normalised to admissions since this data was not available, resulted in:

- An increasing pre-implementation year-on-year trend (IRR: 1.04 per year, 95% CI; 1.01, 1.07) that was consistent with models 1-4 but not observed in any previous model fitted on the shorter timeframe, perhaps due to lower DDD use across most sites in 2015. This in turn contributed to an immediate decline in IQVIA DDDs (-10.3%, 95% CI: -15.0, -5.2) at ARK implementation that was also seen in model 2 (-7.0%, 95% CI: -11.8, -1.8), but not in previous models on the shorter timeframe (**Table 5.7** and **Figure 5.11**).
- An increasing post-implementation trend (IRR: 1.09 per year, 95% CI: 1.05, 1.13) that was similar to the post-COVID trends estimated in models 1-4 (where IRR per year ranged from 1.08-1.11), but differed from other estimates fitted on the shorter timeframe ending in October 2020, which showed either declining trends post-implementation or, in the case of IQVIA DDDs, no evidence of change.
- A year-on-year change in trend post- vs pre-implementation of +4.7% per year (95% CI: -0.1, +9.7) (appendix **Figure 5.21**), which is similar to the +7-8% change in trend after March 2020 vs pre-implementation observed in models 2 and 3. The analysis of IQIVA DDDs across all specialties yielded similar estimates (**Table 5.7** and **Figure 5.11**).

As in models 1-4, the immediate COVID effect estimates remained similar across each modification of these models, showing either an increase in DDDs/admission (IRR range: 1.16-1.22) or a decrease in DDDs (IRR range: 0.65-0.80).

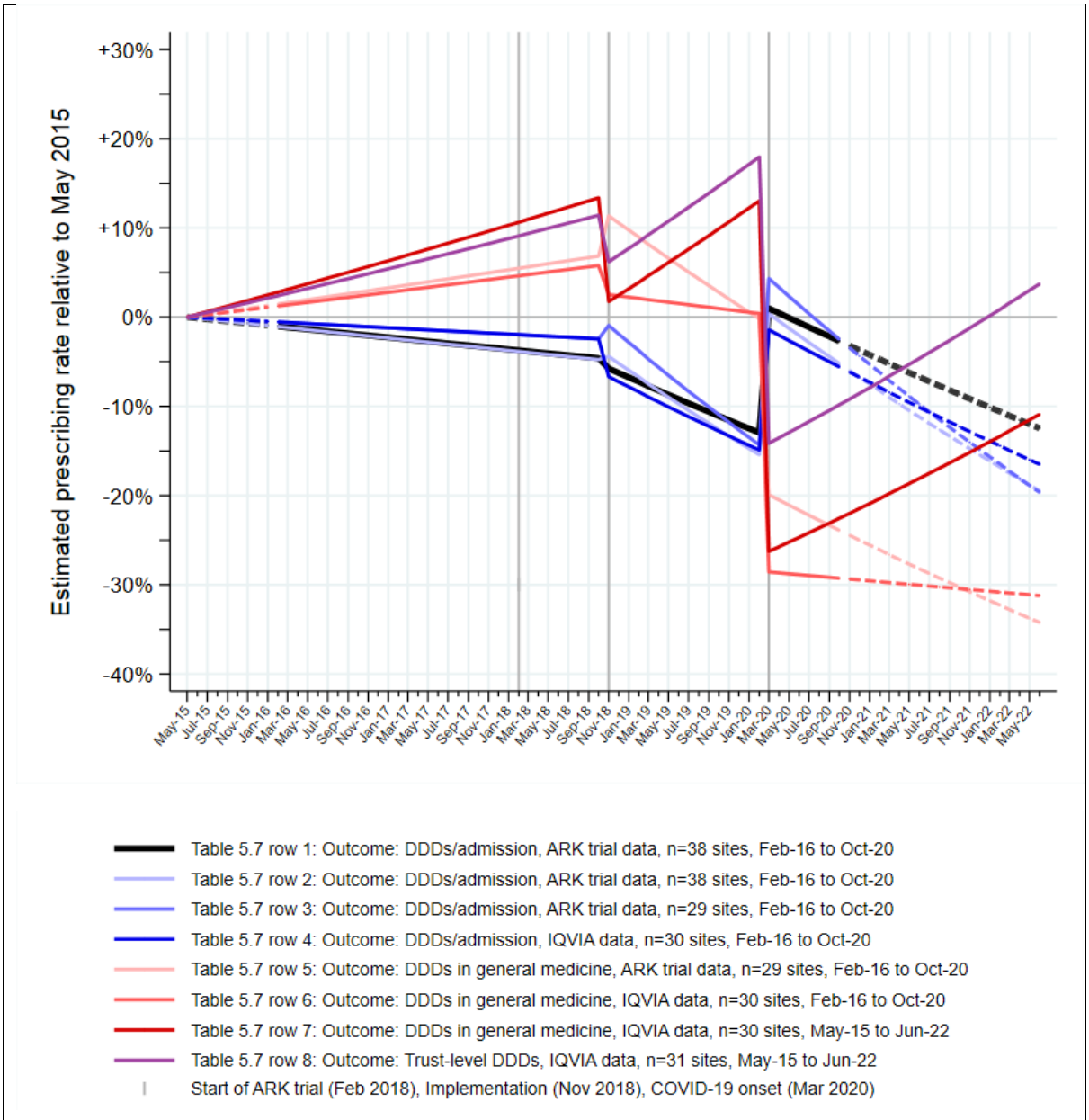
Table 5.7 Impact of the ARK intervention and prescribing trend estimates among ARK trial sites (Model 0)

Outcome[1]	Data source and no. sites	COVID adjust-ment	Cyclical time effect	Time span	Trend before implementation, per year		Trend after implementation, per year		Immediate COVID effect in March 2020		Immediate implementation effect		Change in trend post- vs pre-implementation, per year	
					IRR (95% CI)	p	IRR (95% CI)	p	IRR (95% CI)	p	% Δ IRR (95% CI)	p	% Δ IRR (95% CI)	p
DDDs per admission	ARK n=38[2]	Mar-Jun-20	≤Oct-20	Feb-16, Oct-20	0.99 (0.97,1.01)	0.205	0.94 (0.90,0.97)	**0.001**	1.16 (1.10,1.22)	**<0.001**	-1.0% (-4.0,+2.1)	0.540	-4.8% (-9.1,-0.2)	**0.042**
DDDs per admission	ARK n=38[2]	Mar-Jun-20	≤Feb-20	Feb-16, Oct-20	0.99 (0.97,1.01)	0.187	0.91 (0.87,0.94)	**<0.001**	1.19 (1.13,1.25)	**<0.001**	+0.4% (-2.6,+3.4)	0.812	-7.9% (-12.2,-3.4)	**<0.001**
DDDs per admission	ARK n=29	Mar-Jun-20	≤Feb-20	Feb-16, Oct-20	0.99 (0.97,1.02)	0.596	0.89 (0.85,0.93)	**<0.001**	1.22 (1.15,1.29)	**<0.001**	+1.5% (-1.8,+5.0)	0.362	-10.3% (-15.2,-5.2)	**<0.001**
DDDs per admission	IQVIA n=30	Mar-Jun-20	≤Feb-20	Feb-16, Oct-20	0.99 (0.95,1.03)	0.729	0.93 (0.88,0.98)	**0.007**	1.16 (1.09, 1.24)	**<0.001**	-4.3% (-9.2,+0.7)	0.093	-6.2% (-12.9,+1.1)	0.095
DDDs	ARK n=29	≥Mar-20	≤Feb-20	Feb-16, Oct-20	1.02 (1.00,1.04)	0.073	0.92 (0.87,0.96)	**<0.001**	0.80 (0.76,0.85)	**<0.001**	+4.2% (+0.8,+7.8)	**0.015**	-10.8% (-16.1,-5.2)	**<0.001**
DDDs	IQVIA n=30	≥Mar-20	≤Feb-20	Feb-16, Oct-20	1.02 (0.98,1.05)	0.365	0.98 (0.93,1.04)	0.561	0.71 (0.65,0.78)	**<0.001**	-3.1% (-7.7,+1.7)	0.207	-3.7% (-10.8,+4.0)	0.336
DDDs	IQVIA n=30	≥Mar-20	≤Feb-20	May-15 Jun-22	1.04 (1.01,1.07)	**0.008**	1.09 (1.05,1.13)	**<0.001**	0.65 (0.58,0.73)	**<0.001**	-10.3% (-15.0,-5.2)	**<0.001**	+4.7% (-0.1,+9.7)	0.054
Trust-level DDDs	IQVIA n=31	≥Mar-20	≤Feb-20	May-15 Jun-22	1.03 (1.02,1.05)	**<0.001**	1.09 (1.07,1.10)	**<0.001**	0.73 (0.71,0.75)	**<0.001**	-4.7% (-6.8,-2.5)	**<0.001**	+5.2% (+2.7,+7.7)	**<0.001**

[1] Measured monthly. DDDs refer to general medicine unless otherwise specified.

[2] Included 7 sites outside England and 31 sites in England. Results presented in top row were reported in the ARK trial(105).

Figure 5.11 Estimated antibiotic prescribing rates among ARK trial sites (Model 0)



Note: These predictions are from **Table 5.7**. The implementation date was fixed at Nov 2018 to facilitate model comparisons. The dotted lines have been extrapolated from trend estimates derived using data spanning Feb 2016-Oct 2020.

5.4 Discussion

This chapter assessed whether hospital-level antibiotic prescribing trends differed between ARK and non-ARK sites, as trends in non-ARK sites were unknown (within the trial) and systematic differences had the potential to change inferences drawn from the trial. For example, if secular trends in non-ARK sites showed relatively greater year-on-year increases in antibiotic use than suggested by pre-implementation trends in ARK sites, the impact of the ARK intervention may have been greater than suggested by the modest -4.8% per year decline in DDDs/admission post- vs pre-implementation. This comparison was addressed by fitting model 1, which estimated the year-on-year change in trend for IQVIA DDDs in general medicine after March 2020 (post-COVID) vs before the ARK trial (pre-February 2018). While the estimates differed between ARK (+5.7% per year, 95% CI: +0.1, +11.6) and non-ARK sites (-1.4% per year, 95% CI: -4.6, +1.9), the 95% confidence bounds on these estimates overlapped considerably, suggesting the change in trend is consistent with no difference over the longer-term post-COVID between sites that implemented the ARK intervention and sites that did not. Supporting evidence for the lack of difference between ARK vs non-ARK sites post-COVID is also seen in model 4, which allowed for an additional impact of the ARK intervention pre-COVID. Although model 4 could only be fitted in a subset of ARK sites (n=17/30), it showed DDDs in general medicine changed by +0.3% per year (95% CI: -7.4, +8.6) in ARK sites after March 2020 vs pre-February 2018, again overlapping with estimates from non-ARK sites. Notably, this model suggested a substantial impact of the ARK intervention post-implementation through to COVID vs pre-February 2018 (-7.0% year-on-year decline in DDDs), but this did not reach statistical significance, potentially due to the smaller number of sites that could be included. Without a clear difference in longer term post-COVID trends in DDDs between ARK and non-ARK sites, the sustained impact of ARK implementation on the trial's co-primary outcome (total DDDs/admission) may need to be reconsidered, though caution is warranted in reaching this conclusion due to several limitations outlined below, and results are consistent with the impact of the intervention being temporarily abrogated by the COVID-19 pandemic.

A second aim of this chapter was to assess the stability of prescribing trend and ARK implementation effect estimates, and whether estimates from the trial changed with access to more longitudinal data post-COVID. This aim was first addressed by fitting models 1-4, each of which estimated prescribing trends for IQVIA DDDs in general medicine using slight variations of the interrupted time series analysis used in the ARK trial (model 0), either by modifying the specific months over which trends were estimated, or by excluding the immediate implementation effect. A comparison of AIC and BIC values from these models revealed that among models 1-3 there was no single model that fit the data best, with most models being statistically indistinguishable at the site level. Despite this, trend estimates from models 1-4 were generally stable, with increasing trends before implementation (+3-8% per year), after COVID onset (+8-11% per year), and post-COVID vs pre-implementation (+6-7% per year in models 1-3). The estimated impact of COVID in March 2020 was also consistent across models, associated with an overall immediate decline in IQVIA DDDs between 29-35% in general medicine, or 23-27% across all specialties. Since COVID effects were in opposite directions when normalising for admissions (i.e. DDDs/admission increased 26-22% in model 0), relative declines in admissions in March 2020 were likely greater than the relative declines in DDDs. Annual data reported by ESPAUR corroborates this, showing Trust-level DDDs (measured using the 2019 WHO DDD index(199)) increased 3.1% per year from 2017-2019 and then dropped 18.4% between 2019-2020, while admissions dropped 22.0% between 2019-2020, resulting in a 4.6% increase in Trust-level DDDs/admission(34). The ESPAUR report also noted that from 2020-2021 Trust-level DDDs increased just 1.5% while admissions increased 13.4%, resulting in a decline in DDDs/admission of 10.4%. This modest 2020-2021 increase in Trust-level DDDs is lower than the 11% per year increase I estimated after the onset of COVID in models 1-4, perhaps reflecting the fact that the annual totals reported by ESPAUR included January-February 2020 (i.e. before most sites experienced any impact on prescribing from COVID) and were estimated as a simple proportion (i.e. DDDs in 2021/DDDs in 2020), rather than as a weighted sum of site-specific adjusted trend estimates between March 2020-June 2022.

To understand how ARK implementation estimates changed with access to more recent longitudinal data, I made several iterative changes to model 0. Using data from the ARK trial, these changes showed the impact of ARK implementation on DDDs in general medicine/admission post- vs pre-implementation was even greater among sites in England (-10.3% per year, 95% CI: -15.2, -5.2) than among all 38 ARK sites in the UK (-4.8% per year, 95% CI: -9.1, -0.2), albeit with overlapping confidence bounds. This finding is supported by evidence from Chapter 4, which assessed whether a Trust's location within the UK acted as an effect modifier on the ARK trial's co-primary endpoint (DDDs/admission) by using univariate random-effects meta-regression. Chapter 4 fit UK region as a categorical factor with many sub-regions; had region instead been fit as a binary factor (England vs not England), the analysis would have found moderate evidence of heterogeneity (p=0.025) suggesting English sites had greater year-on-year decreases in DDDs/admission than non-English sites (relative difference: -15.0%, 95% CI: -26.1, -2.1). Given that COVID onset in March 2020 was often associated with a sharp increase in DDDs/admission and adjustment for COVID is likely to be imperfect, the detection of a sustained decline in DDDs/admission may be a reassuring sign that the intervention was effective at reducing total DDDs in general medicine/admission. Alternatively, given other findings it is impossible to exclude sub-optimal model fit also contributing to this finding. However, when model 0 was fit using IQVIA DDDs from February 2016-October 2020 (i.e. overlapping with the trial analysis dates), this sustained decline was no longer observed, and when follow-up was extended from May 2015-June 2022, the estimates differed from those obtained using the trial data, showing an immediate decline in DDDs at implementation (-10.3%) and a non-significant change in trend post- vs pre-implementation of +4.7% per year (95% CI: -0.1, +7.7). This latter effect estimate is consistent with findings from models 1-3, likely because IQVIA DDDs continued to recover towards pre-COVID levels after October 2020, and thus estimates of post-implementation trends through June 2022 shifted towards greater year-on-year increases than analyses ending in October 2020.

There are many limitations affecting analyses of the IQVIA data that complicate the interpretation of these findings. First, I identified likely changes in specialty coding in nearly one-third of sites, and whilst I took steps to minimise the bias introduced by such changes through adjustment or exclusion of the affected data (based on fixed rules), I was cautious in my approach and only flagged the most obvious examples (e.g. **Figure 5.4**), as I did not want to adjust away genuine changes in prescribing over time. Thus, it is plausible some sites with changes in specialty coding have been missed, particularly where the impacts of this change on monthly DDDs are more subtle, and this may introduce a source of residual confounding in the estimated prescribing trends based on IQVIA data. Also, whilst my analysis of Trust-level DDDs avoided the potential bias introduced by changes in specialty coding, this outcome is likely to dilute any implementation effects as DDDs in general medicine (targeted by the ARK intervention) only accounted for a small share of Trust-level DDDs in most sites.

Second, the algorithm used to define general medicine likely differs between the IQVIA data and the ARK trial data, as the absolute magnitude of DDDs in general medicine was substantially lower in the IQVIA data in 23/30 sites. The reason for this difference is not clear, as I applied the same cleaning steps to both datasets (e.g. restricting analyses to the same systemic antibiotics). One possible source of difference is that the IQVIA data did not differentiate between adult and paediatric prescribing in general medicine and may therefore include both, while the ARK data was restricted to prescribing on adult wards where the intervention was implemented. However, this should result in IQVIA DDDs being slightly higher rather than lower, so cannot explain my results. Notably, the relative magnitude of the differences remained fairly constant over time in most ARK sites, suggesting the IQVIA data may reflect a subset of the clinical areas incorporated in the ARK trial definition of general medicine, including where the intervention was genuinely used. Therefore, at least in theory, the IQVIA data could be used assess the longer-term impact of the ARK intervention.

Third, as NHS Digital's IGARD committee was very delayed in processing my application for admissions data, my only source of admission data was from sites in the ARK trial, which meant I could not normalise DDDs for admissions when comparing antibiotic use trends in ARK vs non-ARK sites (model 1). I considered obtaining admissions data from public sources, however, the available datasets all had limitations preventing their use as a denominator (see appendix for details). Since the available admissions data from the trial only spanned February 2016-October 2020, I also could not normalise for admissions when estimating how the ARK implementation effect changed with access to more longitudinal data. As a result, it was not possible to make a direct comparison between the implementation effect presented in the ARK trial (outcome: DDDs/admission) with the estimate obtained using more longitudinal IQVIA data (outcome: DDDs); instead, a number of iterative changes to model 0 were required to facilitate this comparison. Although analyses of DDDs will be biased if monthly changes in DDDs reflect changes in admissions rather than genuine changes in antibiotic prescribing behaviour, it is reassuring that changes in year-on-year trends post- vs pre-implementation were similar between February 2016-October 2020 regardless of whether the analysis considered ARK DDDs/admission (-10.3% per year) vs ARK DDDs (-10.8% per year), or IQVIA DDDs/admission (-6.2% per year) vs IQVIA DDDs (-3.7%), suggesting this source of bias may not change inferences about the sustained implementation effects.

Lastly, estimates from individual Trusts showed substantial heterogeneity (appendix Figures; assessed using I^2 statistics), reflecting genuine substantial differences in implementation effects between sites, as well as potentially undetected specialty coding changes and impact of other changes at the Trust level (e.g. in antimicrobial stewardship, clinical practice, or AMR rates).

Given the challenges noted above, I focused this analysis on the ARK trial's co-primary endpoint (total DDDs/admission), rather than examining the trial's many other secondary antibiotic endpoints (e.g. broad-spectrum or Reserve antibiotic DDDs/admission). As the latest ESPAUR report makes clear, COVID had a profound impact on antibiotic prescribing trends, hospital admissions, and case

mix(34), and though I tried adjusting for COVID, it likely remains a source of residual time dependent confounding. A key question is whether ARK sites continue to implement the intervention effectively post-COVID. Although additional follow-up is needed to assess this, the fundamental challenges with the IQVIA data identified here make repeating these analyses extremely problematic and ultimately I do not consider that the long-term effects of the ARK intervention can be estimated using these routine electronic health datasets.

5.5 Appendix

Analyses of IQVIA DDDs in general medicine from May 2015-June 2022 were not normalised for admissions in general medicine because the required admissions data was not available over this timeframe. Although NHS England publishes monthly admissions by Trust, the data are aggregated across many specialties (not just general medicine) and how the specialties are combined also varies over time. For example, one data source reports admissions for all “specific acute specialties” from April 2018 onwards, but this includes over 100 separate specialties combined (including surgical and paediatric specialties) and is based on the specialty where attending consultants worked (HES field: tretspef) in the last episode of each finished spell(271). A separate, older data source (i.e. Monthly Activity Return data) reports admissions from April 2008-May 2020, but does so for “general and acute” specialties instead, which is defined slightly differently (includes 69 specialties combined) and is based on the specialty where attending consultants were contracted (HES field: mainspef) in the first episode of each spell(272). Although NHS England also reports consultant-led bed-days by Trust and the ARK trial analysed total DDDs/bed-day as a secondary endpoint(34), the bed-days data share similar problems as the admissions data and were therefore not used to normalise IQVIA DDDs. For example, one data source aggregates provider-level monthly bed-days across all “general and acute” beds (defined broadly, as above), but only from March 2020 onwards(273), while another data source provides occupied beds for general medicine specifically but only does so quarterly, rather than monthly(274).

Table 5.8 Rationale for analysing specific sites from a later start date (after May 2015) or early end date (before June 2022) in IQVIA data

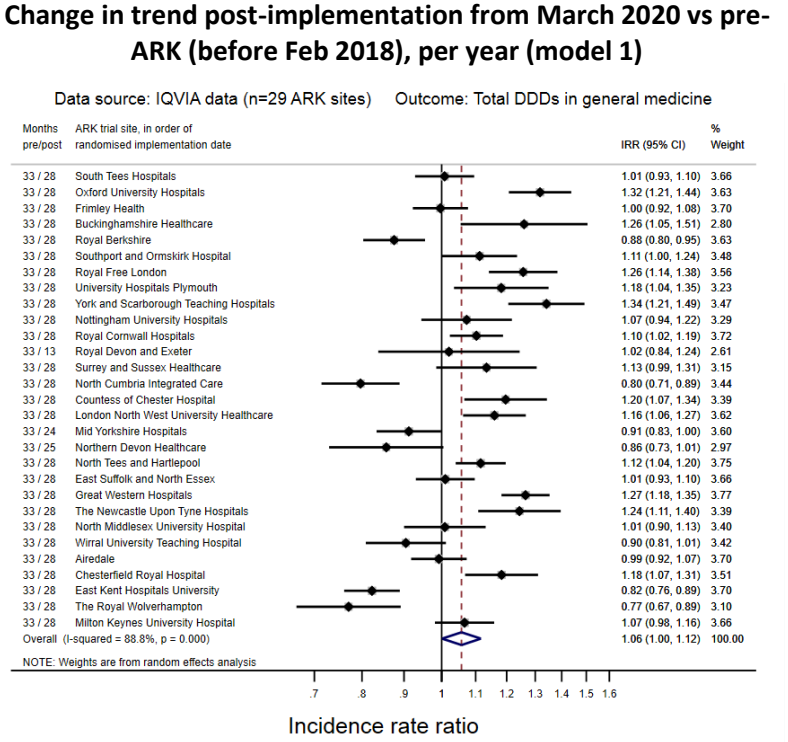
Site	Late start date	Early end date	Rationale for analysis starting after May 2015 or ending before June 2022
James Paget University Hospitals	Sept 2015	N/A	A probable change in specialty coding occurred in September 2015, therefore the start date was pushed forward by 4 months.
Tameside and Glossop Integrated Care	Oct 2015	N/A	A probable change in specialty coding occurred in October 2015, therefore the start date was pushed forward by 5 months.
The Princess Alexandra Hospital	Dec 2015	N/A	A probable change in specialty coding occurred in December 2015, therefore the start date was pushed forward by 7 months.
Maidstone and Tunbridge Wells	Jan 2016	N/A	A probable change in specialty coding occurred in January 2016, therefore the start date was pushed forward by 8 months.
South Warwickshire	Jan 2016	May 2021	A probable change in specialty coding occurred in January 2016, therefore the start date was pushed forward by 8 months. Also, prescribing data was only provided up to June 2021; since this month was missing data for most antibiotics, the end date was pulled back by 1 month.
Torbay and South Devon	Mar 2016	N/A	Between May 2015 and February 2016 the monthly DDDs in general medicine were near zero, therefore the start date was pushed forward by 10 months.
East Sussex Healthcare	Apr 2016	N/A	A probable change in specialty coding occurred in April 2016, therefore the start date was pushed forward by 11 months.
University Hospitals of Bristol	N/A	Sept 2019	Although this site merged with Weston Area Health in April 2020, the University Hospitals of Bristol reported no antibiotic data after September 2019 and the sites were therefore analysed separately.
University College London Hospitals	N/A	Dec 2019	The raw data did not contain antibiotic data after December 2019.
Weston Area Health	N/A	Mar 2020	Merged with University Hospitals of Bristol in April 2020, but since Bristol reported no antibiotic data after Sept 2019, these sites were analysed separately and data from Weston Area Health was included up to the last month before the sites merged.
Poole Hospital; also The Royal	N/A	Sept 2020	To enable assessment of Poole Hospital (an ARK trial site), its merger with The Royal

Site	Late start date	Early end date	Rationale for analysis starting after May 2015 or ending before June 2022
Bournemouth and Christchurch Hospitals			Bournemouth and Christchurch Hospitals in October 2020 was not accounted for and data from these sites was analysed separately, up to September 2020.
Barking, Havering and Redbridge University Hospitals	N/A	Jan 2021	Raw antibiotic data was only provided up to January 2021.
Cambridge University Hospitals[1]	N/A	Oct 2021	A probable change in specialty coding occurred in November 2021, therefore the end date was pulled back by 8 months.
Calderdale and Huddersfield	N/A	Nov 2021	A probable change in specialty coding occurred in December 2021, therefore the end date was pulled back by 7 months.
Mid Yorkshire Hospitals	N/A	Feb 2022	A probable change in specialty coding occurred in March 2022, therefore the end date was pulled back by 4 months.
Northern Devon Healthcare; also Royal Devon and Exeter	N/A	Mar 2022	To enable assessment of these ARK sites, their merger in April 2022 was not accounted for and data from these sites was analysed separately, up to March 2022.
Mid Cheshire Hospitals	N/A	May 2022	June 2022 was missing data for most antibiotics so analyses included data up to May 2022 instead.

[1] In Cambridge University Hospitals, DDDs for most specialties (including general medicine) were not reported from July-December 2016, so DDDs for these months were set to missing in analyses of Trust-level antibiotic use.

Figure 5.12 Model 1: Change in trend from Mar 2020 vs before Feb 2018 per year, IQVIA DDDs (May 2015-Jun 2022), ARK sites (A) and non-ARK sites (B)

A

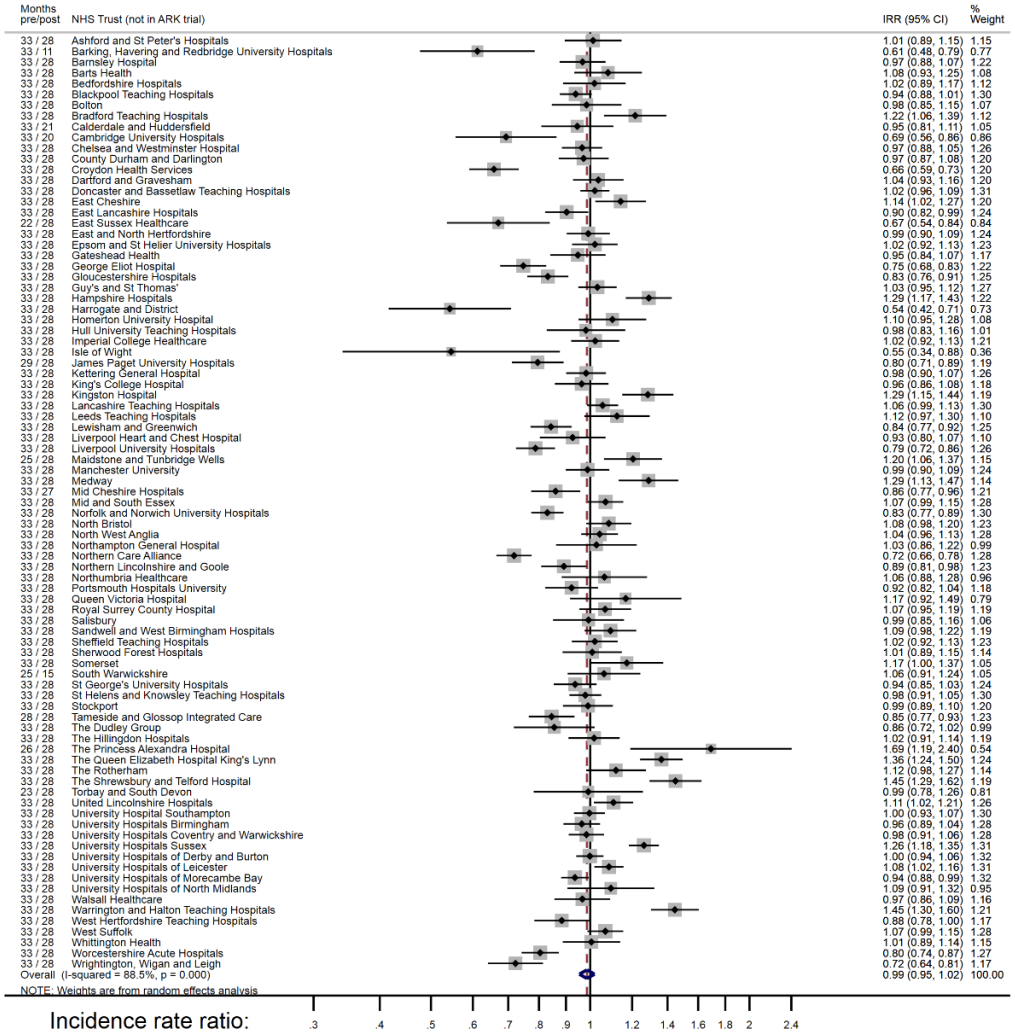


[1] Overall pooled effect p=0.0449.

[2] All estimates adjusted for COVID and a cyclical time effect up to Feb 2020; select sites adjusted for a change in specialty coding or post-COVID step.

[3] Excluded Poole Hospital as the site only had 7 months of post-COVID data avialable (to Sept 2020).

B

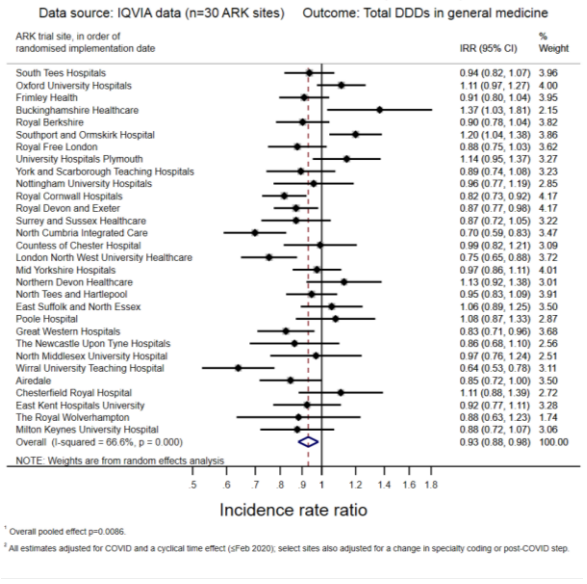


[1] Overall pooled trend p=0.3950. All estimates adjusted for COVID and a cyclical time effect up to Feb 2020; select sites adjusted for a change in specialty coding or a post-COVID step. Excluded 2 sites without post-COVID data and 2 sites lacking sufficient post-COVID data (≤7 months).

Figure 5.13 Model 2: Immediate ARK implementation effect (A), change in trend post-implementation to Feb 2020 vs pre-implementation per year (B), and change in trend post-implementation from March 2020 vs pre-implementation per year (C), IQVIA DDDs (May 2015-June 2022), ARK sites in England

A

Immediate intervention effect (model 2)

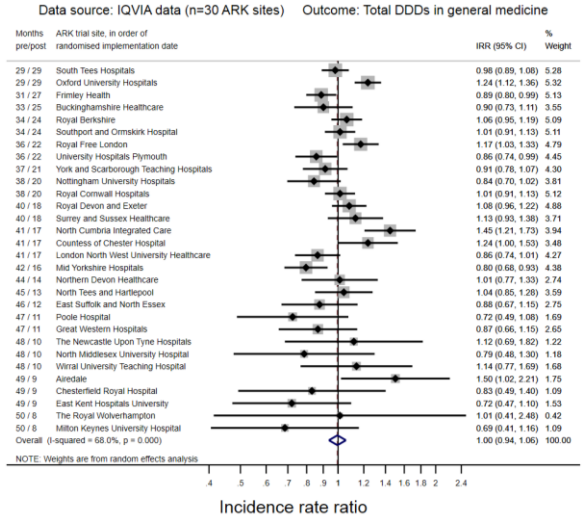


[1] Overall pooled effect p=0.0086.

[2] All estimates adjusted for COVID and a cyclical time effect (sFeb 2020); select sites also adjusted for a change in specialty coding or post-COVID step.

B

Change in trend post-implementation to Feb 2020 vs pre-implementation, per year (model 2)

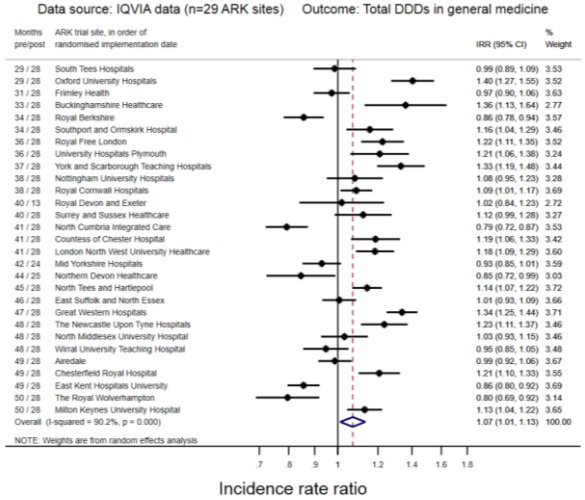


[1] Overall pooled effect p=0.8732.

[2] All estimates adjusted for COVID and a cyclical time effect (sFeb 2020); select sites also adjusted for a change in specialty coding or post-COVID step.

C

Change in trend post-implementation from March 2020 vs pre-implementation, per year (model 2)



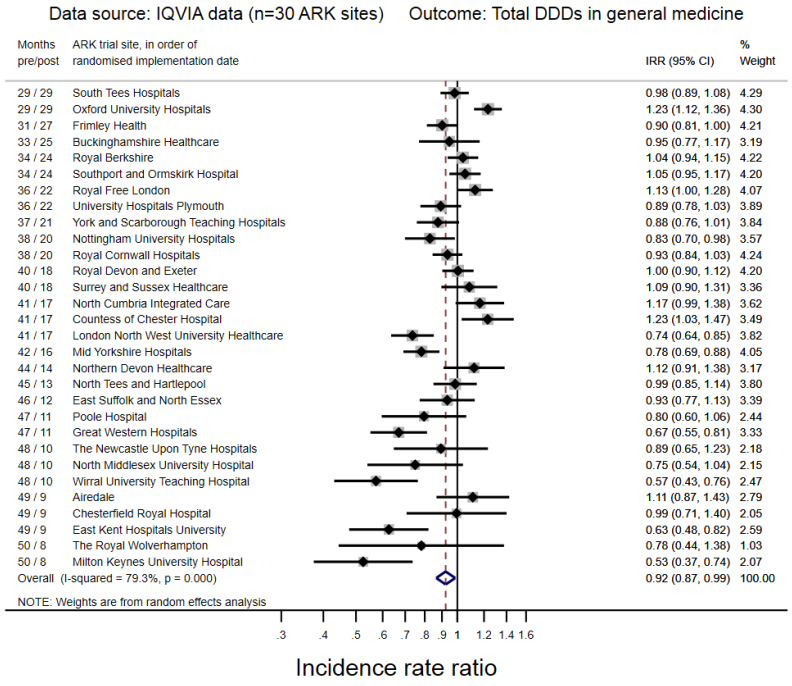
[1] Overall pooled effect p=0.0203.

[2] All estimates adjusted for COVID and a cyclical time effect (sFeb 2020); select sites also adjusted for a change in specialty coding or post-COVID step.

[3] Excluded Poole Hospital as the site only had 7 months of post-COVID data avialable (to Sept 2020).

Figure 5.14 Model 3: Change in trend post-implementation to Feb 2020 vs pre-implementation per year (A) and from March 2020 vs pre-implementation per year (B), IQVIA DDDs (May 2015-June 2022), ARK sites in England

A Change in trend post-implementation to Feb 2020 vs pre-implementation, per year (model 3)

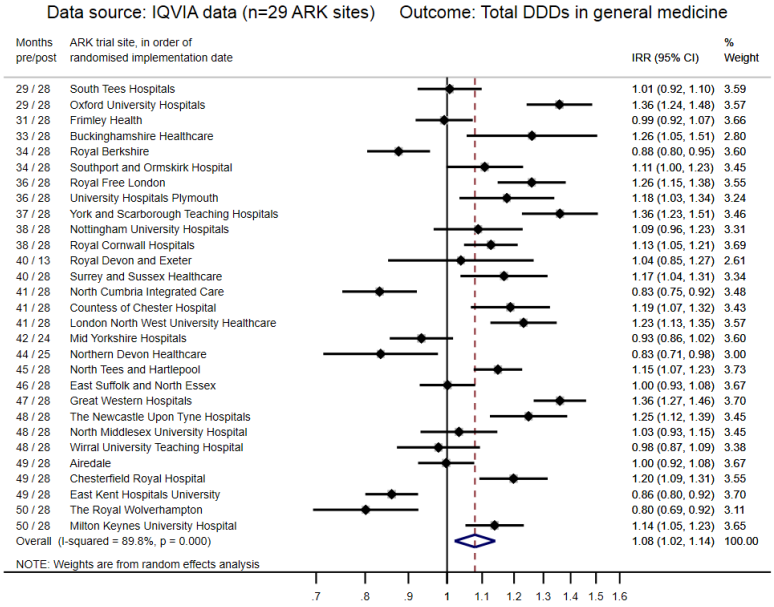


[1] Overall pooled effect p=0.0158.

[2] All estimates adjusted for COVID and a cyclical time effect up (≤Feb 2020); select sites also adjusted for a change in specialty coding or post-COVID step.

[3] Estimates were not adjusted for the immediate implementation effect.

B Change in trend from March 2020 vs pre-implementation, per year (model 3)



[1] Overall pooled effect p=0.0072.

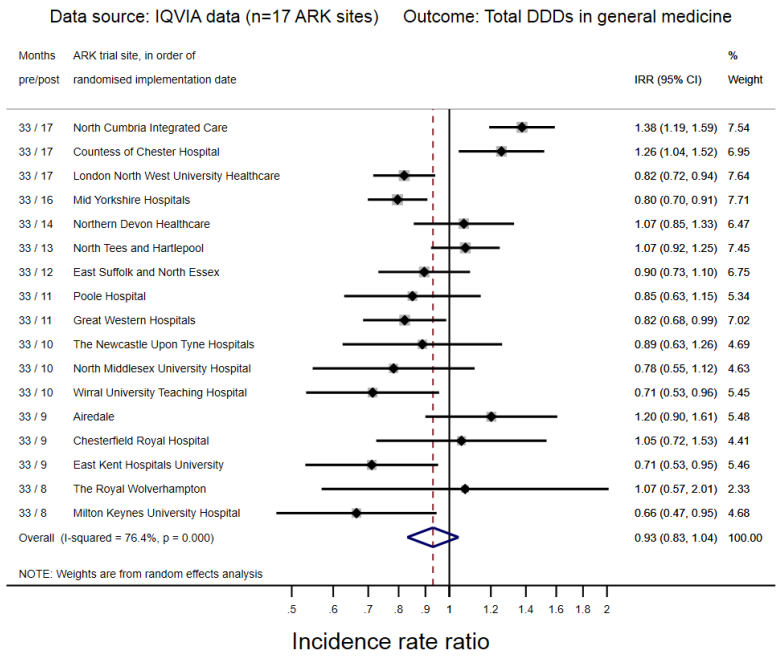
[2] All estimates adjusted for COVID and a cyclical time effect (≤Feb 2020); select sites adjusted for a change in specialty coding or post-COVID step.

[3] Estimates were not adjusted for the immediate implementation effect.

[4] Excluded Poole Hospital as the site only had 7 months of post-COVID data avialable (to Sept 2020).

Figure 5.15 Model 4: Change in trend post-implementation to Feb 2020 vs before Feb 2018 per year (A), and from March 2020 vs before Feb 2018 per year (B), IQVIA DDDs (May 2015-June 2022), ARK sites in England

A Change in trend post-implementation to February 2020 vs pre-ARK (before February 2018), per year (model 4)



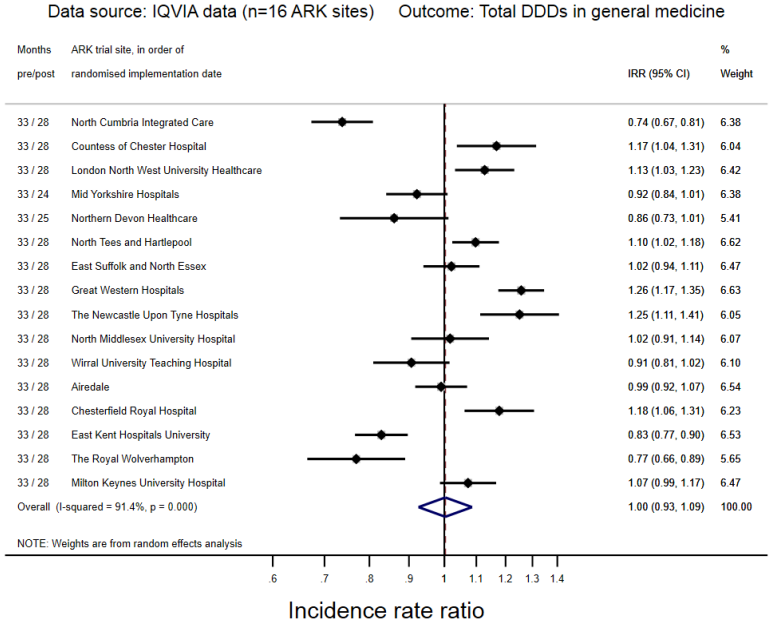
[1] Overall pooled effect p=0.2092.

[2] All estimates adjusted for COVID and a cyclical time effect (≤Feb 2020); select sites adjusted for a change in specialty coding or post-COVID step.

[3] Estimates were not adjusted for the immediate implementation effect.

[4] Excluded 13 sites with <8 months of data available to estimate trends between Feb 2018 and implementation.

B Change in trend from March 2020 vs pre-ARK (before February 2018), per year (model 4)



[1] Overall pooled effect p=0.9445.

[2] All estimates adjusted for COVID and a cyclical time effect (≤Feb 2020); select sites adjusted for a change in specialty coding or post-COVID step.

[3] Estimates were not adjusted for the immediate implementation effect.

[4] Excluded 13 sites with <8 months of data available to estimate trends between Feb 2018 and implementation.

[5] Excluded Poole Hospital as the site only had 7 months of post-COVID data avialable (to Sept 2020).

Figure 5.16 Model 0: Immediate ARK implementation effect (A) and change in trend post- vs pre-implementation per year (B), ARK trial DDDs/admission (Feb 2016-Oct 2020), 38 ARK sites in UK

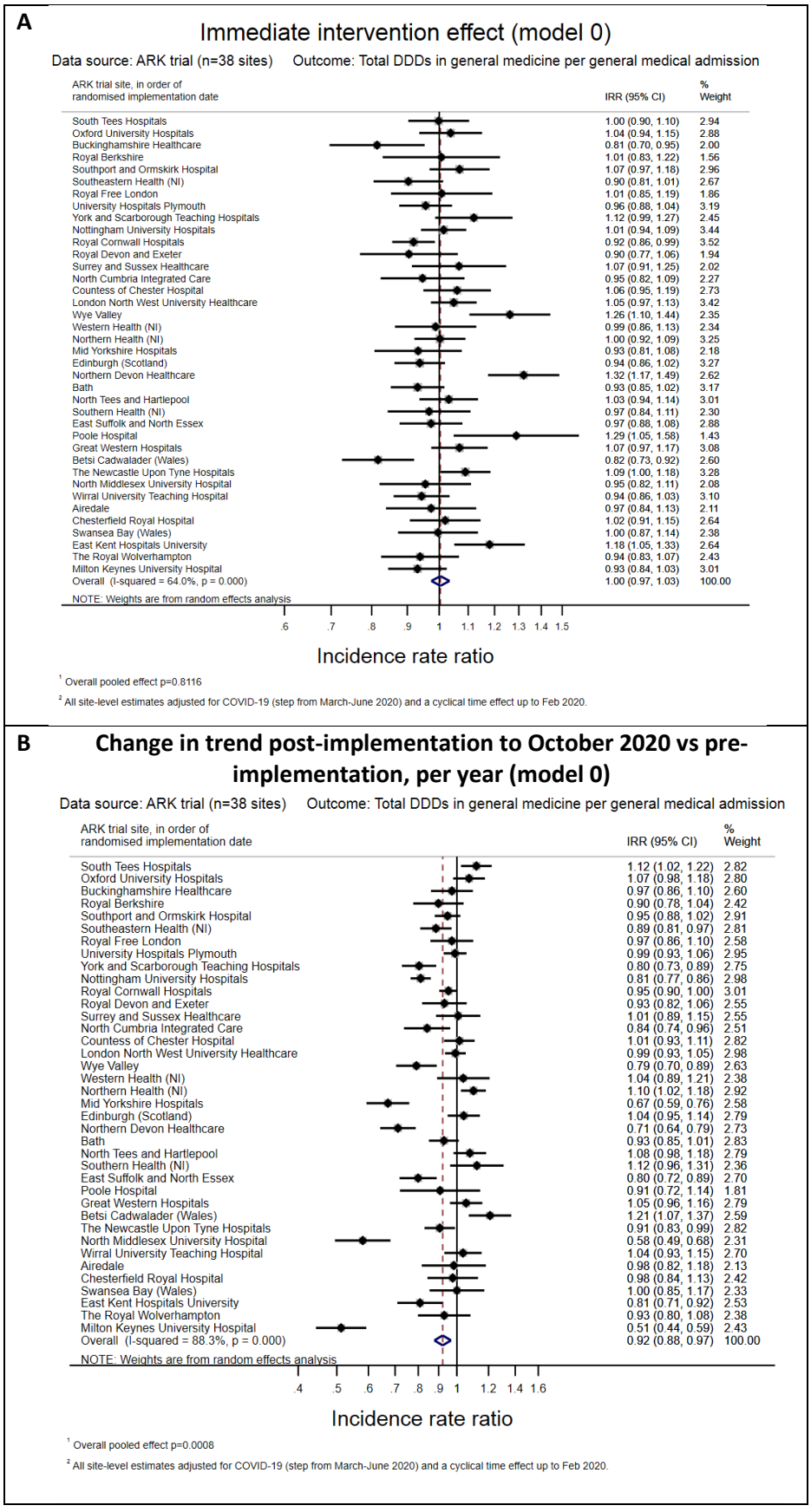


Figure 5.17 Model 0: Immediate ARK implementation effect (A) and change in trend post- vs pre-implementation per year (B), ARK trial DDDs/admission (Feb 2016-Oct 2020), ARK sites in England

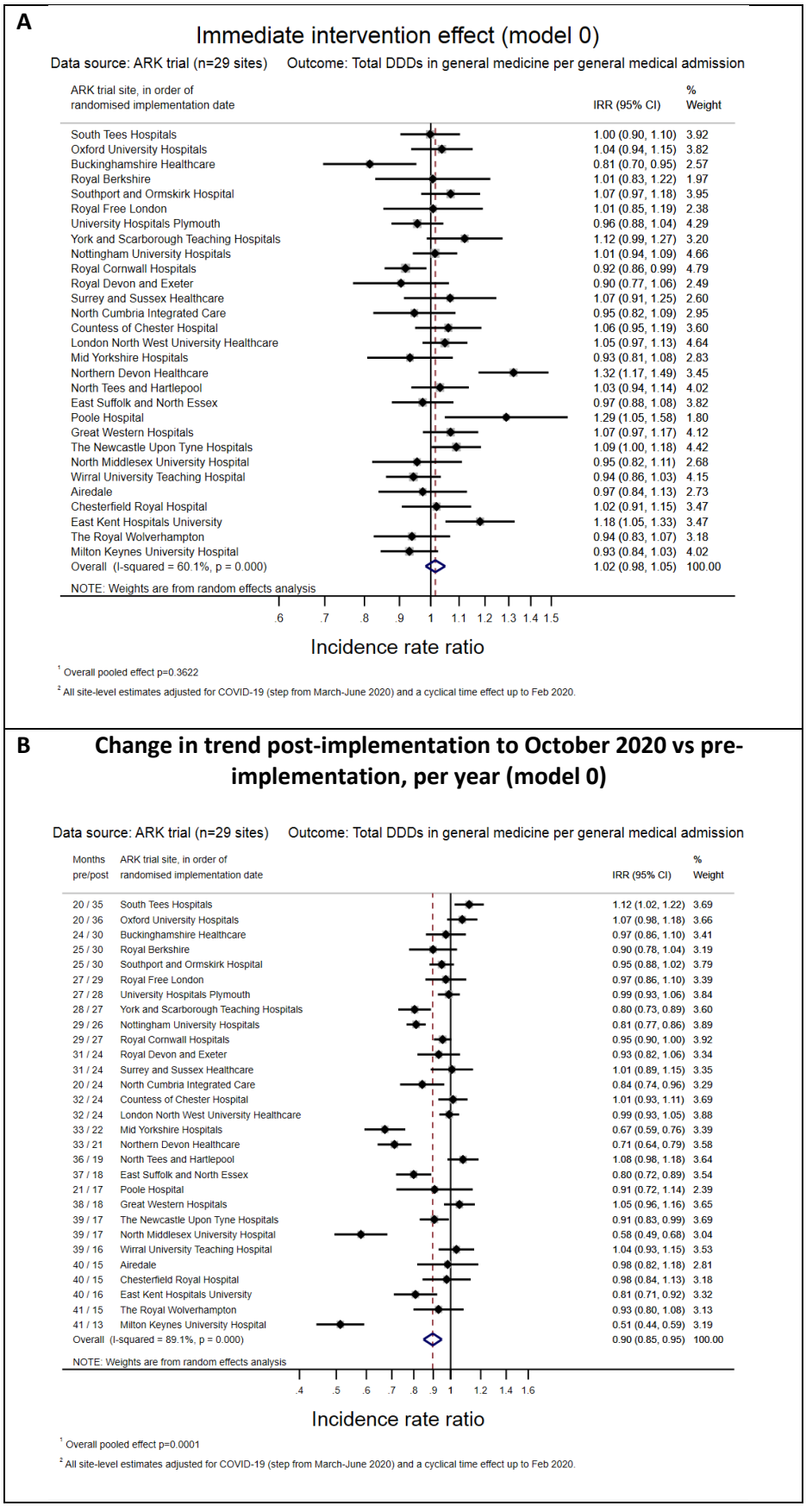


Figure 5.18 Model 0: Immediate ARK implementation effect (A) and change in trend post- vs pre-implementation per year (B), ARK trial DDDs (Feb 2016-Oct 2020), ARK sites in England

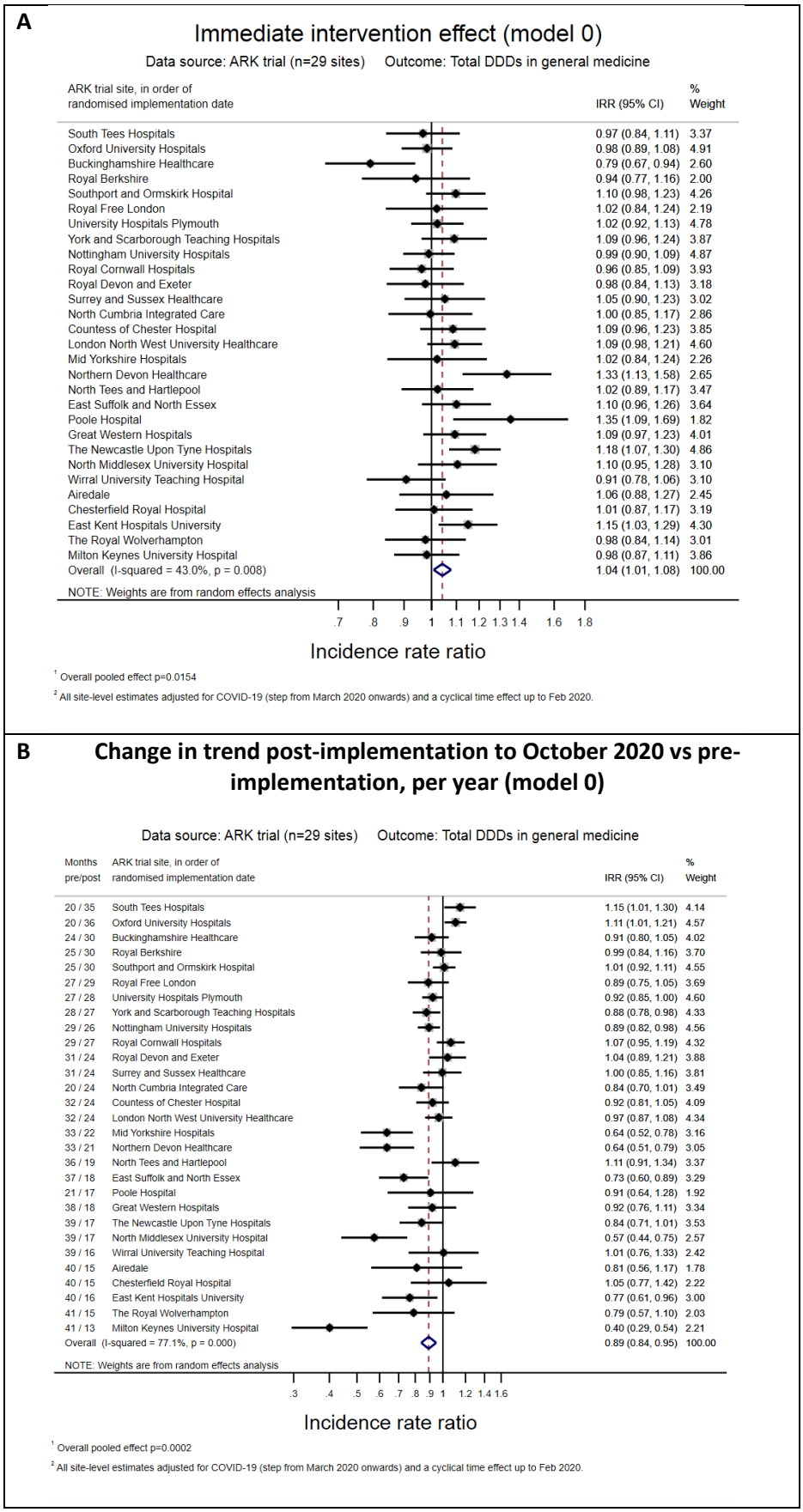


Figure 5.19 Model 0: Immediate ARK implementation effect (A) and change in trend post- vs pre-implementation per year (B), IQVIA DDDs/admission (Feb 2016-Oct 2020), ARK sites in England

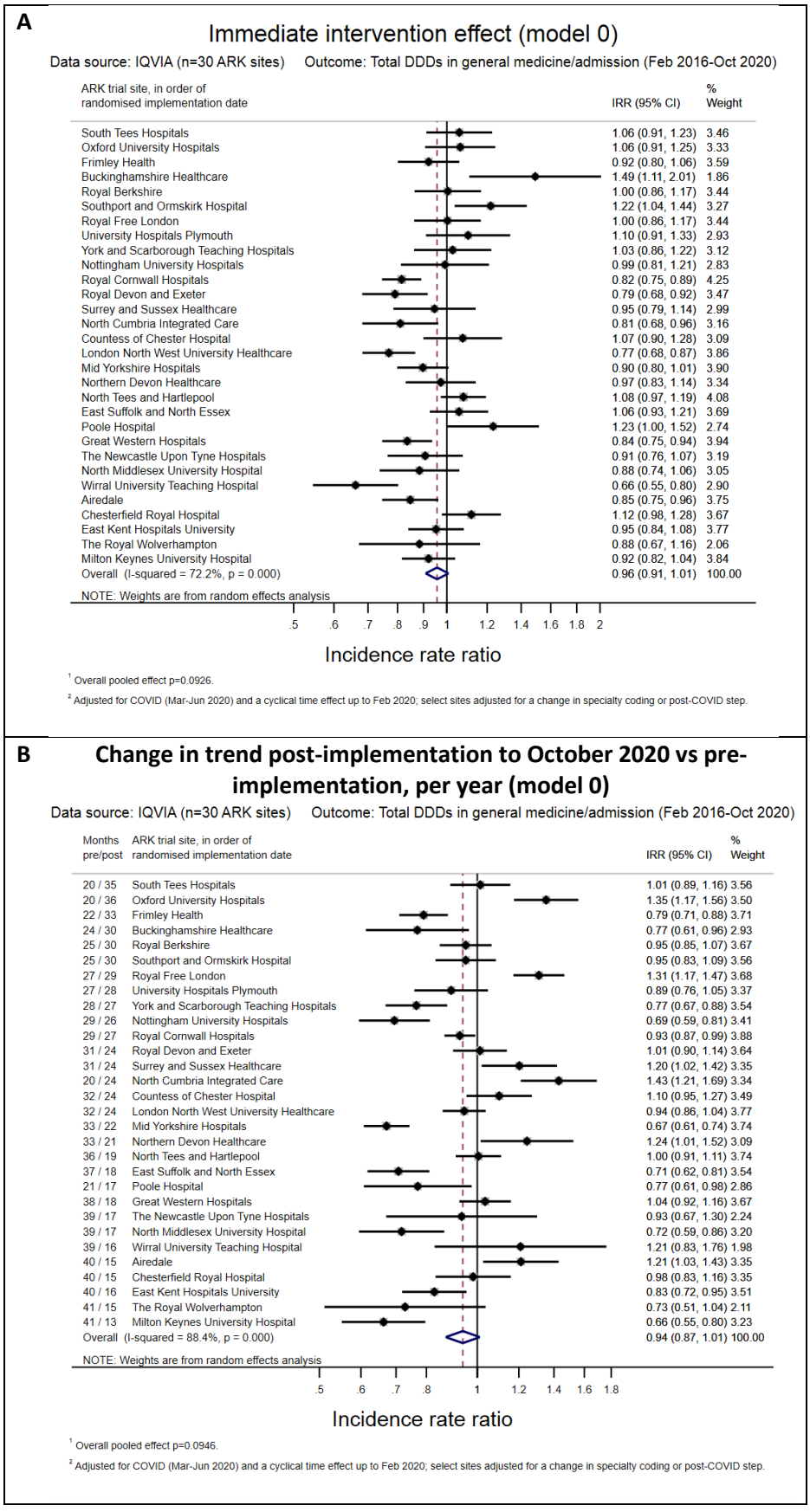


Figure 5.20 Model 0: Immediate ARK implementation effect (A) and change in trend post- vs pre-implementation per year (B), IQVIA DDDs (Feb 2016-Oct 2020), ARK sites in England

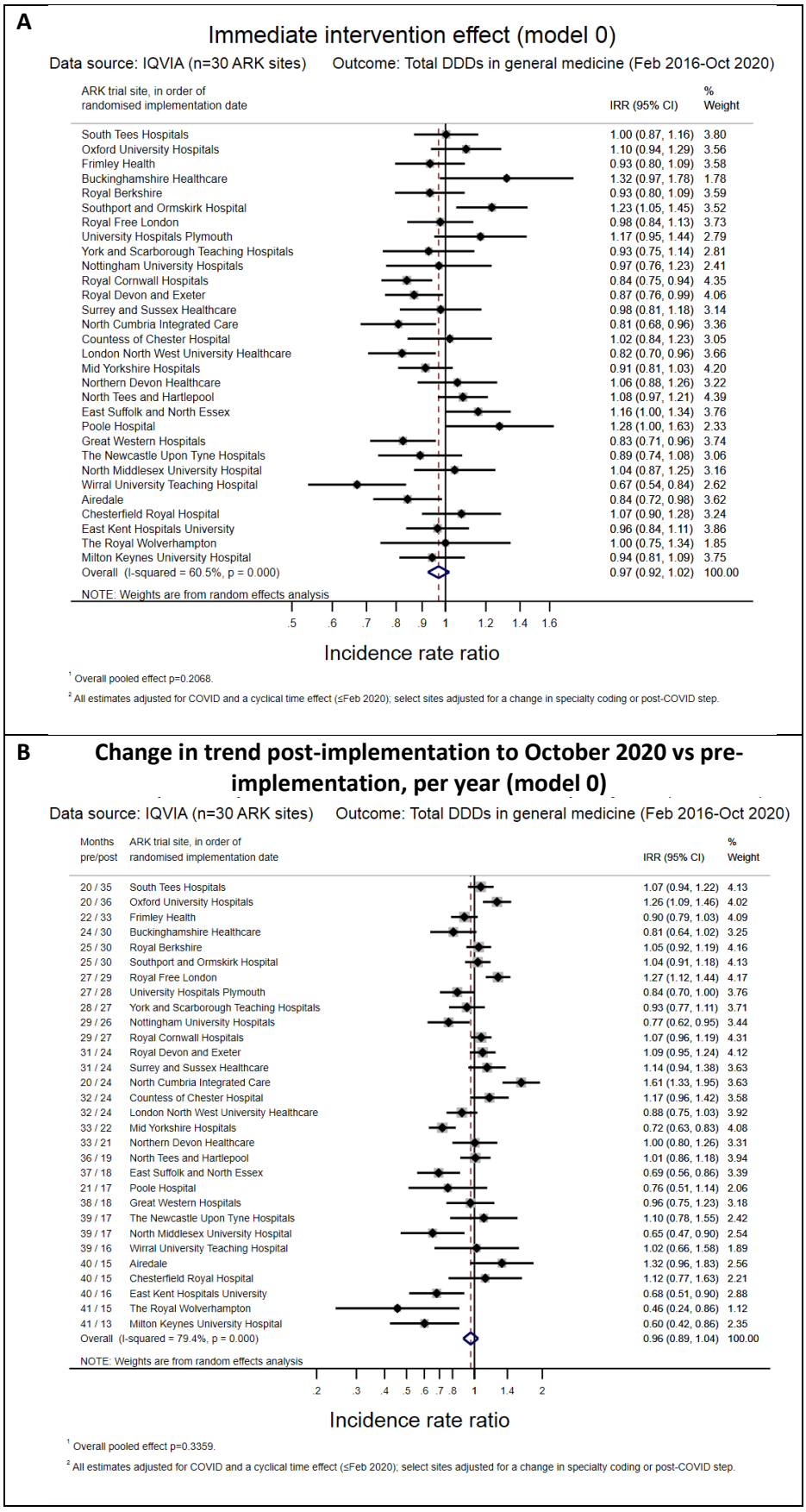
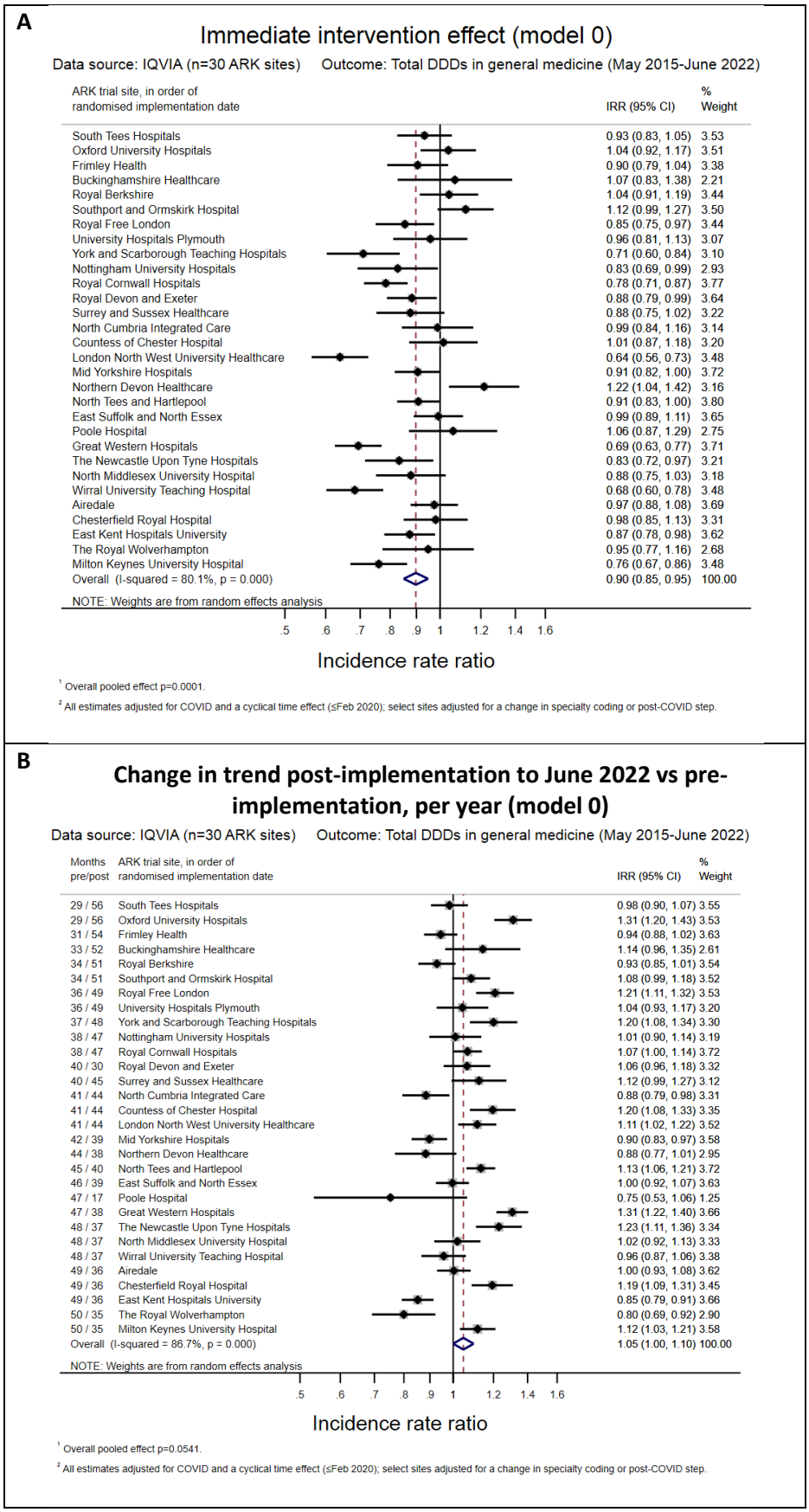


Figure 5.21 Model 0: Immediate ARK implementation effect (A) and change in trend post- vs pre-implementation per year (B), IQVIA DDDs (May 2015-June 2022), ARK sites in England



Chapter 6: Conclusions and future work

6.1 Summary

In this thesis I leveraged routine data from hospital pharmacies and electronic health records to investigate the clinical impact of lower patient- and hospital-level antibiotic use and addressed key questions emerging from the recently completed ARK-Hospital trial. Although my findings have shed light on possible mediators of impact in the ARK trial and provided some reassurance to practitioners that substantial reductions in antibiotic use could be safe, I identified several analytic challenges that stem from the underlying routine data, with broad implications for investigators in the NHS who might rely on this or similar data to measure antimicrobial usage and clinical outcomes to evaluate the impact of antimicrobial stewardship programmes and reduced antibiotic use.

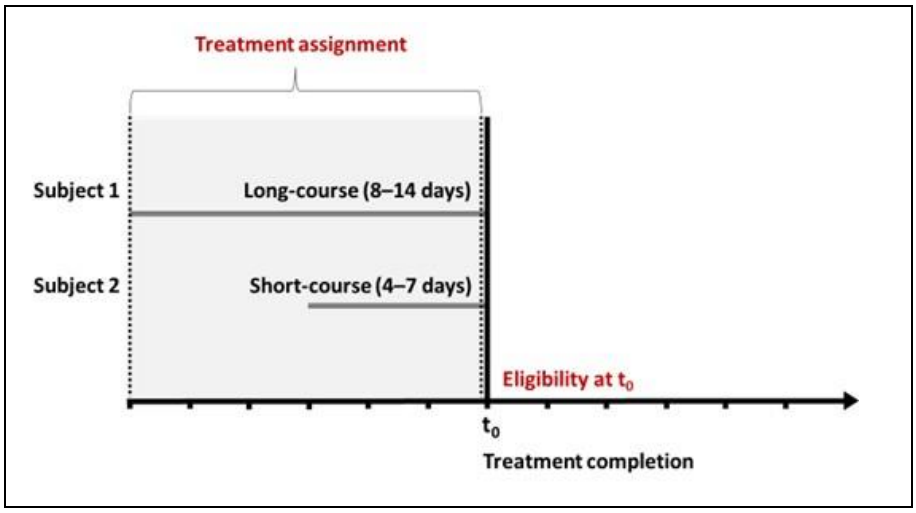
In Chapter 2 I found no evidence that the widely varying antibiotic use among acute hospitals in England was associated with mortality risk in general medical inpatients, and concluded that (i) the scale of antibiotic reductions that may be safely achieved likely dwarfs the reduction targets found in NHS hospital contracts, and that (ii) hospitals (or individual wards) could therefore be benchmarked, following patient case-mix adjustment, against their lower-prescribing peers to drive these reductions. Many risk adjustment procedures are available to benchmark antimicrobial usage for inter-hospital comparisons, and they often rely on the availability of patient-level characteristics that are predictive of antibiotic use, either for regression adjustment(275,276) or standardisation(277–279), including factors such as blood biomarkers and microbiology results, primary and secondary diagnosis codes, and history of outpatient antibiotic use or drug-resistant infections. However, with the exception of diagnosis codes, the HES data used to derive marginal effects in Chapter 2 did not include these other patient-level factors and laboratory test results, which likely contributes to some residual confounding by indication. Risk adjustment procedures could just condition on location or facility-level factors when patient-level factors are not available(280,281), but the inclusion of patient-level data in case-mix adjustments has been shown to consistently improve inpatient

antibiotic use predictions over models that rely on location and facility-level factors alone, resulting in substantial differences in the ranking of hospitals and more accurate observed-to-predicted antibiotic use ratios(275). Though not available in HES, patient-level laboratory tests can, at least in theory, be obtained directly from UK hospitals (e.g. through data sharing agreements, as seen in the data from University Hospitals Birmingham in Chapter 3); however, obtaining this data from large numbers of hospitals for benchmarking purposes would require the collaboration of many informatics and microbiology departments whose capacity and availability to deliver on complex data requests varies widely across the NHS. This variation in capacity was noted in the ARK-Hospital trial, which attempted to obtain routinely collected patient-level blood biomarker and microbiology test results, but was unsuccessful in 90% of participating trial sites (appendix p.5 in (105)), despite the availability of financial reimbursements for time spent on data extraction and repeated requests for the data in the year after the trial ended in October 2020. The COVID pandemic likely contributed to this difficulty through its impact on staffing resources, but attempts to collect laboratory test data at earlier timepoints in the trial (before COVID) also proved difficult, suggesting this challenge is likely to limit data availability and the resulting scope of case-mix adjustments for antimicrobial stewardship leads or policy makers interested in benchmarking antimicrobial use across large numbers of UK hospitals.

In addition to benchmarking antimicrobial use, shorter antibiotic courses offer one means of reducing antibiotic overuse, but prescribing guidelines are only as robust as their evidence base and for many bacterial infections better evidence is needed regarding the impact of shorter versus longer antibiotic prescriptions(61). Although observational comparisons of antibiotic duration offer several practical advantages over RCTs (e.g. affordability, ability to detect harm with large representative population samples)(65), published examples often suffer from avoidable methodological weaknesses that increase risk of bias, such as misclassified or left-censored follow-up (**Figure 1.1**).

As an illustrative case example, a recent observational analysis compared short (4–7 days) vs long-course (8–14 days) antibiotic treatment in hospitalised adults admitted with community-acquired pneumonia(282), motivated by growing evidence that shorter antibiotic courses for mild or moderate community-acquired pneumonia in adults may be noninferior to longer courses across antibiotic classes(168,169). Since treatment was not randomly assigned, confounding by indication was addressed by including inverse probability of treatment weights (IPTW) in logit models, but each treatment strategy was modelled as time-fixed and follow-up began at treatment completion, rather than treatment initiation, meaning outcomes (deaths within 30 days) were only ascertained for those who remained alive and event-free at the end of treatment (as depicted in **Figure 6.1**). Although counting time from the end of treatment avoids immortal time, the inclusion of such patients (called "prevalent users") will still induce a survivor bias since mortality risk likely varies with time and deaths early on treatment were not ascertained, and because confounding factors measured before the start of follow-up may have been affected by treatment(78,79). The study also imposed post-treatment eligibility criteria by restricting the analysis to patients who achieved clinical stability within 3 days of antibiotic therapy, who constitute a subset of otherwise eligible patients alive at the end of treatment, enhancing selection bias(74,100). If there was non-ignorable attrition or differential adherence across exposure groups before the end of treatment, this too could increase bias. Although the authors noted the potential for survivor bias, they did not acknowledge the steps they could have taken to align exposure assignment with eligibility and start of follow-up; for example, they could have cloned subjects to the short and long-course treatment strategies and then censored them when they deviated from the assigned strategy, followed by the use of inverse probability of censoring weighting to avoid selection bias from informative censoring(76).

Figure 6.1 Selection bias from the use of post-treatment eligibility criteria and starting follow-up at the end of treatment (immortal time excluded)



Despite its potential as a technique to avoid misclassified and left-censored follow-up, this 3-step cloning-censoring-weighting procedure has only recently been applied to questions of antibiotic treatment duration. For example, in a paper published in 2022, Evans et al(98) used the procedure to compare the effect of short (<10 days), intermediate (10–18 days) and long-courses (>18 days) of active antibiotic therapy on 28-day in-hospital mortality in 587 patients with *S. aureus* bacteraemia. Also, Zuercher et al(99) very recently (2023) used the cloning-censoring-weighting procedure and a time-dependent Cox model(283) to compare the effect of short (≤10 days) vs long-course (>10 days) antibiotic treatment on time to the earliest of in-hospital death or discharge, in a retrospective cohort of 382 ICU patients with bacteraemia. However, these examples are outliers and it remains a far more common practice in cohort studies comparing treatment durations to model antibiotic use as time-fixed and avoid immortal time by starting follow-up at the end of treatment or excluding deaths on treatment (resulting in selection bias). This can be seen in many studies of treatment for *S. aureus* bacteraemia(284–289) and other bloodstream infections, where the authors failed to take steps to avoid immortal time(290) or selection bias introduced by left-censored follow-up(291) and/or excluded deaths on treatment(292–295). Where the effect of shorter vs longer antibiotic treatment has been treated as time-varying, outcome models often condition only on baseline

confounding factors, rather than accounting for time-varying markers of infection severity such as CRP that might be used to guide treatment duration in patients with bacteraemia(293,296).

Another relatively common weakness in longitudinal studies of antibiotic treatment relates to how treatment assignment is defined, with some studies combining the absence of antibiotic treatment or inactive therapy with an active comparator(290), or describing treatment strategies imprecisely, resulting in contrasts between treatments that may be composed of many different antibiotic regimens, doses, and administration strategies that may have different causal effects, thereby violating a key identifiability condition (known as "consistency") required for causal inference(297,298). Concerningly, contrasts between alternate time-fixed treatment durations are also occasionally limited to unadjusted differences in simple proportions or means/medians, with no acknowledgement of the potential for confounding by indication or the evident immortal follow-up(299).

6.1.1 Avoiding immortal time and prevalent user bias

Given the central importance of avoiding these methodological weaknesses in future observational analyses of antibiotic treatment duration, I next discuss ways to improve these analyses.

Non-random treatment assignment may be the most obvious characteristic distinguishing observational studies from RCTs, but as outlined above, other differences in design and analysis are common and may account for much of the variation between treatment effect estimates in randomised versus observational studies(104). At the study design stage, immortal and prevalent user bias can be avoided by aligning eligibility and cohort entry with exposure assignment, such that all follow-up time is accounted for and correctly classified as exposed or unexposed(74). This can be facilitated by designing a hypothetical ("target") RCT protocol, in which short vs long-course antibiotic treatment is randomly assigned, and then trying to emulate the protocol as closely as possible in the observational analysis by adopting the same eligibility criteria, treatment strategies,

outcomes, and follow-up period(72,73), and complying with relevant reporting guidelines(300). When emulation differences are minimised, similar inferences may be derived from RCTs and observational analyses, but this may be difficult to achieve in practice with routine databases(70).

Although incident (vs prevalent) exposures are often preferrable to avoid selection bias(301), arguments justifying the inclusion of prevalent users may point to scenarios where the only available data are left-truncated, or where right censoring from attrition and missing data means new users lack sufficient follow-up to study long-term outcomes(302). However, this latter justification overlooks the fact that while analyses of prevalent users simply ignore right censoring that occurs before the start of follow-up, new user designs may adjust for right censoring through inverse probability weighting(78). Despite the advantages of new-user designs, a recent review of reporting practices in cohorts of drug-outcome associations found only 40% (36/89) of studies implemented a new-user design, of which only half (19/36) reported full alignment between start of follow-up, eligibility, and cohort entry(81). A second review also found a quarter (19/77) of studies were at high risk of bias from misclassified or left-censored follow-up, most of which failed to acknowledge the problem when discussing limitations(80).

At the analysis stage, the steps required to avoid immortal time bias will depend on whether the causal contrast of interest is the observational analogue of an intention to treat (ITT) effect (that is, the effect of being assigned to a treatment strategy, regardless of what happens subsequently), or a per-protocol effect (that is, the effect of strictly adhering to an assigned treatment strategy throughout follow-up). However, studies of drug-outcome associations rarely report these target estimands explicitly(81). For example, if data on treatment adherence were unavailable, a study could estimate the average causal effect of short- versus long-course prescription assignment (an ITT analogue), as might occur if relying on outpatient prescribing records(303), such as those captured in the UK's Clinical Practice Research Datalink (CPRD) database(151,304,305). In this scenario, the assigned antibiotic prescription duration would be known at the start of follow-up and exposure

status could be modelled as time-fixed without misclassifying follow-up. Confounding adjustment could then be limited to baseline factors, controlled through standard outcome regression methods or by estimating and conditioning on each subject's probability of exposure assignment given the measured baseline covariates (i.e. propensity score (PS)), either through PS matching, PS stratification, or PS regression adjustment using IPTWs(306,307). However, if loss to follow-up or treatment adherence differed across the exposure groups (e.g. due to treatment-associated toxicity), prescription assignment would provide a biased reflection of treatment efficacy, which may be of greater clinical interest to researchers assessing the safety of alternate regimens(308). Alternatively, if data on treatment adherence were available, investigators might emulate an ITT comparison of the effect of stopping (vs continuing) antibiotic treatment at different timepoints, regardless of subsequent management(104). As illustrated in Chapter 3, this can be modelled by emulating a sequence of non-randomised nested "trials" at each relevant timepoint, adjusted for baseline confounding up to each timepoint. This nested "trials" approach has recently been used to investigate the effect of initiating (vs not initiating) antibiotic therapy at different timepoints(309), but to my knowledge Chapter 3 provides the first application of the technique to understanding the consequences of stopping therapy.

Of note, the analysis in Chapter 3 is an example of a scenario where including prevalent users is necessary and unproblematic, as it avoids misclassified follow-up and selection bias by design. Using this approach, I found no evidence of elevated mortality risk from early discontinuation of inpatient antibiotics (i.e. when discontinuation was lagged to minimise bias in patients who stop active treatment due to imminent death), suggesting clinical harm is not an inevitable consequence of shorter antibiotic courses. However, this analysis faced multiple challenges stemming from the available routine EHR data. Some of these limitations, such as the absence of antibiotic consumption data prior to admission, could, in theory, be resolved through the collection of additional routine data, allowing adjustment for additional sources of unmeasured confounding. Other limitations are more intractable however, such as the fact that primary and secondary diagnosis codes are recorded

retrospectively based on discharge summaries and may be later ruled out or unrelated to the clinical indication for antibiotic prescribing. Not adjusting for these diagnoses could result in substantial residual confounding as many will be prognostic for treatment outcomes, but adjusting for them (e.g. as CCS groups or comorbidity scores) may use the future to predict the past in some patients, and it is not clear this problem can be resolved using current EHR systems.

Where the causal estimand is adherence to an assigned regimen (analogous to a per-protocol effect), treatment durations may instead be contrasted using the 3-step cloning-censoring-weighting procedure noted above(76) to inform treatment duration. Alternatively, antibiotic use could be modelled as a continuous time-varying exposure(83) with adjustments for time-varying markers of infection severity (e.g. CRP), however, standard regression methods (e.g. time-dependent Cox models, or generalised estimating equations) will generate biased effect estimators in the presence of time-varying confounders that are affected by previous treatment (even in the absence of model misspecification or unmeasured confounding, as illustrated in **Figure 1.2**). Thus alternative structural models are needed to obtain consistent estimates of causal effects. Methods developed for this purpose include using IPTWs to estimate the parameters of a marginal structural model (MSM)(91) or dynamic MSM(196,310,311), g-computation (parametric g-formula)(312), g-estimation of structural nested models(313), and hybrid approaches such as doubly robust methods(314). A detailed description of each of these "g methods" and their relative advantages and disadvantages is beyond the scope of this discussion, but has been covered elsewhere(92–94). Use of more than one of these methods may be advisable, as inconsistent results could flag problems with the analyses, which would be helpful as each approach requires certain structural and parametric modelling assumptions to hold, but violation of these assumptions can be difficult to ascertain in practice with routine EHR data(315). For example, each approach requires causal diagrams to be correct and time-varying treatment and confounders to be correctly measured throughout follow-up, but as highlighted above, prognostic diagnosis codes are captured retrospectively and may be misclassified in routine data, and the construction of treatment episodes can be challenging. Ideally, electronic

dispensing records would accurately capture the dates, dose, and administration route for each antibiotic, but in practice analyses may have to rely on paper-based records, patient-reports, or prescribing registers that do not accurately reflect dispensing(316), with temporal gaps and overlaps of varying duration between consecutive prescriptions(317) or non-compliance (e.g. tablet splitting, missed doses) leading to justifiable concerns about treatment misclassification errors and the plausibility of the "consistency" assumption(297). The use of IPTWs also assumes "positivity", meaning there must be a non-zero (positive) probability a subject will be assigned treatment at each timepoint of follow-up, given their covariates(318), which if violated can lead to extreme weights that introduce bias. While it may be reasonable to truncate extreme weights, truncation can also introduce bias and it would therefore be sensible to assess the sensitivity of estimands to weight model specifications(319).

Systematic reviews of the use of causal methods (up to 2016) did not identify any studies examining antibiotic treatment duration(95,96), and while I found a couple recently published examples applying the cloning-censoring-weighting procedure(98,99), these appear to be outliers. Thus, despite the challenges noted above, where the causal estimand is analogous to a per-protocol effect, more widespread use of these causal methods could be helpful, as it would help to avoid many of the preventable and relatively common methodological weaknesses that remain in observational comparisons of antibiotic treatment durations.

6.1.2 Addressing questions that emerged from the ARK-Hospital trial

Evaluations of AMS interventions are often marred by study design weaknesses that limit the translation of evidence into clinical practice(47,220–223). The ARK intervention design process and trial evaluation overcame many of these weaknesses, providing good quality evidence the intervention was effective in achieving safe and sustained reductions in organisation-level antibiotic use in general medicine(105). However, given the multifaceted nature of the intervention and the heterogeneity of the participating sites (e.g. in terms of geographic distribution, size, fidelity of

implementation), it may not be surprising that site-level effect estimates varied quite widely. In Chapter 4, I examined the role of 26 contextual factors as potential mediators of impact, to better understand the sources of this variation between sites. This exploratory analysis revealed the drivers of short and long-term intervention effects may differ, with the strongest evidence for effect modification seen in sites that increased prescription stop rates after baseline (vs did not), which was associated with greater immediate declines in parenteral, carbapenem and Reserve antibiotics, followed by sites that introduced processes for ongoing audit and feedback by the implementation date (vs did not), which reflected better implementation preparation and was associated with greater immediate declines in total, oral, and narrow-spectrum antibiotics. However, due to limited power, the evidence for observed effect modifications was relatively weak, and, in particular, the limited evidence for effect modification on sustained implementation effects was consistent with what might be expected under the global null, given the large number of tests performed. Ultimately, the interaction between these contextual factors and the different intervention components may be better explored through qualitative and mixed-methods approaches, to better understand how the intervention interacts with the specific implementation context and contributes to system change (e.g. in sites with electronic vs paper-based prescribing systems). This qualitative work is being pursed separately in ongoing work by ARK collaborators, and may provide further complimentary insights into the barriers and facilitators of impact and challenges of sustaining ARK over time and across different acute care settings.

In Chapter 5, I leveraged data from IQVIA to examine the longer-term impact of the ARK intervention and estimated prescribing trends for all acute Trusts in England, using non-ARK sites as a counterfactual to understand whether the modest impact of the ARK intervention may have been even greater (or less) than suggested by the trial-site specific analysis alone. Although the impact of COVID on prescribing trends greatly complicated assessments of the longer-term impact of ARK, the findings revealed no clear differences in prescribing trends, given the considerable overlap in confidence bounds on trend estimates, between ARK and non-ARK sites before the trial began,

during the trial, or after COVID onset in March 2020 (post-intervention in ARK sites). That differences in trends in ARK vs non-ARK sites were not observed prior to the trial (before February 2018) suggests ARK sites are broadly representative of acute Trusts in England, and the trial findings may therefore be generalizable in this setting. However, the analysis highlighted a number of challenges that stem from the underlying routine data, with broad implications for other researchers who might rely on similar organisation-level routine prescribing data to estimate trends and evaluate the impact of AMS interventions. For example, the algorithm used to define DDDs in general medicine appeared to differ between the trial data (obtained directly from sites through data sharing agreements) and the data obtained from IQVIA, resulting in monthly DDDs that varied greatly in absolute (and occasionally relative) magnitude, though the reasons for this difference are unclear, as both nominally reflect ward-level prescribing of systemic antibiotic DDDs in general medicine. Additionally, analysis of the IQVIA data revealed possible administrative changes in reported specialty coding for antibiotic use in nearly one-third of acute Trusts in England between May 2015-June 2022. Although detecting these changes was inevitably somewhat subjective, many of these changes have the potential to introduce substantial bias in trend estimates, compromising any interrupted time series analysis relying on these data. Finally, although they enable standardised comparisons of antibiotic use without the need for patient-level data, DDDs have their own inherent limitations as a metric of antibiotic use and may over- or under-estimate days of therapy (as discussed in the Chapter 2 Discussion)(157), and measuring ward-level antibiotic use as DDDs/admission necessarily relies on two separate routine data sources (i.e. one for the numerator, and another for the denominator). Ideally, patients reflected in the denominator would account for all of the DDDs in the numerator (and vice-versa), however, this is unlikely in practice given the challenges with specialty coding identified in the IQVIA data, and the difficulty of identifying general medical inpatients from HES consultant specialty codes (see **Figure 2.1**). Given these varied challenges, the long-term effects of the ARK intervention (post-COVID) likely cannot be estimated unbiasedly using these routine electronic health datasets, precluding any future assessments.

6.1.3 Directions for future work

Although more than one-third (14/39) of the NHS Trusts participating in the ARK-Hospital trial had electronic prescribing systems, only 4 sites had the informatics resources available to provide patient-level antibiotic prescribing data when requested. With more time available, this patient-level data could potentially be used to explore the impact of the ARK intervention on patient-level antibiotic outcomes and the role of contextual factors as mediators of impact on these patient-level outcomes. Also, as analyses in Chapter 3 relied on relatively old data (limiting generalisability), this analysis could be updated using the more recent patient-level antibiotic data, and the effect of treatment discontinuation at different timepoints could be explored using the cloning-censoring-weighting procedure to emulate the per-protocol effect of a target RCT with full adherence. However, the limitations inherent in these routine data would remain, notably the retrospective recording of diagnosis codes and challenges of reconstructing treatment episodes from e-prescribing records. In Chapter 2, future observational work using an England-wide dataset could also expand on the outcome measures already considered to investigate other markers of harm (e.g. length of hospital stay, ICU admission) or benefit (e.g. AMR and *C. difficile* infection rates) that might be associated with lower hospital-level antibiotic use. Similarly, investigations could explore how, after case-mix adjustments, some hospitals achieve low-levels of antibiotic use, as this information could inform ongoing AMS activity and evaluation work.

Over the last 5 years, e-prescribing systems have become more widely available and are likely to become ubiquitous in future. However, as described above, the detail included in e-prescribing systems means that data manipulation into meaningful antibiotic treatment episodes is complex, compared with traditionally recorded start and stop dates on paper drug charts. Also, with the adoption of common data models across distributed data networks in the European Union(320), Canada(321), and the United States(322), multi-database studies could be used more widely to compare antibiotic treatment strategies, allowing investigators to remotely replicate the same

protocol-driven treatment comparisons in datasets that are stored and maintained separately by network partners using standardised formats and content(323,324). This approach to comparative effectiveness research has been used to overcome the logistical and regulatory barriers associated with sharing and pooling confidential or proprietary health record data(325–327), generating more precise and generalisable database-specific effect estimates that can then be compared and summarised using meta-analytic techniques(328).

6.1.4 Concluding remarks

This thesis has shown that whilst there is widespread interest in the potential of routine electronic databases for use in clinical trials(258,259) and AMS research(260,261), there are many analytic challenges stemming from the use of these data that might temper expectations, and I have shown there is considerable scope for improvement in the methods and reporting practices of observational studies comparing antibiotic treatment durations. Inherent limitations in EHR systems and specific data accessibility and quality issues have been highlighted, with broad implications for investigators and policy makers relying on this data to benchmark antimicrobial use across UK hospitals, evaluate AMS initiatives, or compare the effects of shorter antibiotic therapies. Overall, the thesis demonstrates the need for high-quality health-record-linked e-prescribing systems to support robust assessments of the impact of antibiotic reduction strategies.

References

1. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect Dis. 2019 Jan 1;19(1):56–66.

2. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022 Feb;399(10325):629–55.

3. Naylor NR, Atun R, Zhu N, Kulasabanathan K, Silva S, Chatterjee A, et al. Estimating the burden of antimicrobial resistance: a systematic literature review. Antimicrob Resist Infect Control. 2018 Apr 25;7:58.

4. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. London: Review on Antimicrobial Resistance. 2016. Available from: https://amr-review.org/sites/default/files/160518_Final paper_with cover.pdf

5. de Kraker MEA, Davey PG, Grundmann H, BURDEN study group. Mortality and hospital stay associated with resistant Staphylococcus aureus and Escherichia coli bacteremia: estimating the burden of antibiotic resistance in Europe. Opal SM, editor. PLoS Med. 2011 Oct 11;8(10):e1001104.

6. Holmes AH, Moore LSP, Sundsfjord A, Steinbakk M, Regmi S, Karkey A, et al. Understanding the mechanisms and drivers of antimicrobial resistance. The Lancet. 2016 Jan 9;387(10014):176–87.

7. Allel K, Day L, Hamilton A, Lin L, Furuya-Kanamori L, Moore CE, et al. Global antimicrobial-resistance drivers: an ecological country-level study at the human-animal interface. Lancet Planet Health. 2023 Apr 1;7(4):e291–303.

8. Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. A systematic review and meta-analysis of the effects of antibiotic consumption on antibiotic resistance. BMC Infect Dis. 2014 Dec 9;14(1):13.

9. Chatterjee A, Modarai M, Naylor NR, Boyd SE, Atun R, Barlow J, et al. Quantifying drivers of antibiotic resistance in humans: a systematic review. Lancet Infect Dis. 2018 Dec;18(12):e368–78.

10. Lee K, Raguideau S, Sirén K, Asnicar F, Cumbo F, Hildebrand F, et al. Population-level impacts of antibiotic usage on the human gut microbiome. Nat Commun. 2023 Dec 1;14(1).

11. Goossens H, Ferech M, Vander Stichele R, Elseviers M, ESAC Project Group. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. The Lancet. 2005 Feb;365(9459):579–87.

12. McDonnell L, Armstrong D, Ashworth M, Dregan A, Malik U, White P. National disparities in the relationship between antimicrobial resistance and antimicrobial consumption in Europe: an observational study in 29 countries. J Antimicrob Chemother. 2017 Nov 1;72(11):3199–204.

13. Costelloe C, Metcalfe C, Lovering A, Mant D, Hay AD. Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis. BMJ. 2010 May 18;340:c2096.

14. Bryce A, Hay AD, Lane IF, Thornton H V., Wootton M, Costelloe C. Global prevalence of antibiotic resistance in paediatric urinary tract infections caused by Escherichia coli and association with routine use of antibiotics in primary care: systematic review and meta-analysis. BMJ. 2016 Mar 15;352:i939.

15. Rowlinson E, Soge OO, Hughes JP, Berzkalns A, Thibault C, Kerani RP, et al. Prior Exposure to Azithromycin and Azithromycin Resistance Among Persons Diagnosed With Neisseria gonorrhoeae Infection at a Sexual Health Clinic: 2012-2019. Clin Infect Dis. 2023 Feb 1;76(3):e1270–6.

16. Fleming A. Penicillin. Nobel Lecture. 1945. Available from: https://www.nobelprize.org/prizes/medicine/1945/fleming/lecture/

17. Abraham E, Chain E, Fletcher C, Gardner A, Heatley N, Jennings M, et al. Further Observations on Penicillin. The Lancet. 1941;238(6155):177–89.

18. Medical Research Council. Streptomycin Treatment of Pulmonary Tuberculosis. Br Med J. 1948 Oct 10;2(4582):769–82.

19. Swedish National Association Against Tuberculosis. Therapeutic Trials Committee. Para-aminosalicylic acid treatment in pulmonary tuberculosis. Am Rev Tuberc. 1950;61(5):597–612.

20. Daniels M, Hill AB. Chemotherapy of Pulmonary Tuberculosis in Young Adults. Br Med J. 1952 May 5;1(4769):1162.

21. Llewelyn MJ, Fitzpatrick JM, Darwin E, SarahTonkin-Crine, Gorton C, Paul J, et al. The antibiotic course has had its day. BMJ. 2017 Jul 26;358:j3418.

22. Carattoli A. Plasmids and the spread of resistance. Int J Med Microbiol. 2013 Aug;303(6–7):298–304.

23. Huddleston JR. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. Infect Drug Resist. 2014 Jun 20;7:167–76.

24. Bhalodi AA, van Engelen TSR, Virk HS, Wiersinga WJ. Impact of antimicrobial therapy on the gut microbiome. J Antimicrob Chemother. 2019 Jan 1;74(Suppl 1):i6–15.

25. Baker S, Thomson N, Weill FX, Holt KE. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. Science. 2018 May 18;360(6390):733–8.

26. Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term impacts of antibiotic exposure on the human intestinal microbiota. Microbiology (N Y). 2010 Nov;156(Pt 11):3216–23.

27. Chng KR, Li C, Bertrand D, Ng AHQ, Kwah JS, Low HM, et al. Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. Nat Med. 2020 Jun 1;26(6):941–51.

28. Hess AS, Kleinberg M, Sorkin JD, Netzer G, Johnson JK, Shardell M, et al. Prior colonization is associated with increased risk of antibiotic-resistant Gram-negative bacteremia in cancer patients. Diagn Microbiol Infect Dis. 2014;79(1):73.

29. Barcudi D, Sosa EJ, Lamberghini R, Garnero A, Tosoroni D, Decca L, et al. MRSA dynamic circulation between the community and the hospital setting: New insights from a cohort study. J Infect. 2020 Jan 1;80(1):24–37.

30. Jernigan JA, Hatfield KM, Wolford H, Nelson RE, Olubajo B, Reddy SC, et al. Multidrug-Resistant Bacterial Infections in U.S. Hospitalized Patients, 2012-2017. N Engl J Med. 2020 Apr 2;382(14):1309–19.

31. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, et al. Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC. Antimicrob Agents Chemother. 2016 Jun 1;60(6):3767–78.

32. Decraene V, Phan HTT, George R, Wyllie DH, Akinremi O, Aiken Z, et al. A Large, Refractory Nosocomial Outbreak of Klebsiella pneumoniae Carbapenemase-Producing Escherichia coli Demonstrates Carbapenemase Gene Outbreaks Involving Sink Sites Require Novel Approaches to Infection Control. Antimicrob Agents Chemother. 2018 Dec 1;62(12).

33. Department of Health ARHAI MRSA Screening Implementation Group. Implementation of modified admission MRSA screening guidance for NHS. 2014. Available from: https://www.gov.uk/government/publications/how-to-approach-mrsa-screening

34. UK Health Security Agency. English surveillance programme for antimicrobial utilisation and resistance (ESPAUR) report 2021 to 2022. 2022. Available from: https://www.gov.uk/government/publications/english-surveillance-programme-antimicrobial-utilisation-and-resistance-espaur-report

35. Enne VI, Livermore DM, Stephens P, Hall LMC. Persistence of sulphonamide resistance in Escherichia coli in the UK despite national prescribing restriction. Lancet. 2001 Apr 28;357(9265):1325–8.

36. Sundqvist M, Geli P, Andersson DI, Sjölund-Karlsson M, Runehagen A, Cars H, et al. Little evidence for reversibility of trimethoprim resistance after a drastic reduction in trimethoprim use. J Antimicrob Chemother. 2010 Nov 8;65(2):350–60.

37. Baker S, Duy PT, Nga TVT, Dung TTN, Phat VV, Chau TT, et al. Fitness benefits in fluoroquinolone-resistant Salmonella Typhi in the absence of antimicrobial pressure. Elife. 2013 Dec 10;2:e01229.

38. Aliabadi S, Anyanwu P, Beech E, Jauneikaite E, Wilson P, Hope R, et al. Effect of antibiotic stewardship interventions in primary care on antimicrobial resistance of Escherichia coli bacteraemia in England (2013-18): a quasi-experimental, ecological, data linkage study. Lancet Infect Dis. 2021 Dec 1;21(12):1689–700.

39. D'Atri F, Arthur J, Blix HS, Hicks LA, Plachouras D, Monnet DL. Targets for the reduction of antibiotic use in humans in the Transatlantic Taskforce on Antimicrobial Resistance (TATFAR) partner countries. Euro Surveill. 2019 Jul 1;24(28):1800339.

40. HM Government. Tackling antimicrobial resistance 2019–2024: the UK's five-year national action plan. 2019. Available from: https://www.gov.uk/government/publications/uk-5-year-action-plan-for-antimicrobial-resistance-2019-to-2024

41. NHS England. Commissioning for Quality and Innovation. Available from: https://www.england.nhs.uk/nhs-standard-contract/cquin/

42. Budd E, Cramp E, Sharland M, Hand K, Howard P, Wilson P, et al. Adaptation of the WHO Essential Medicines List for national antibiotic stewardship policy in England: being AWaRe. J Antimicrob Chemother. 2019;74(11):3384–9.

43. Public Health England. AMR Local Indicators. Available from: https://fingertips.phe.org.uk/profile/amr-local-indicators

44. NHS England. NHS Standard Contract 2023/24 Service Conditions. 2023. Available from: https://www.england.nhs.uk/nhs-standard-contract/23-24/

45. NHS England. NHS Standard Contracts. Available from: https://www.england.nhs.uk/nhs-standard-contract/previous-nhs-standard-contracts/

46. Islam J, Ashiru-Oredope D, Budd E, Howard P, Walker AS, Hopkins S, et al. A national quality incentive scheme to reduce antibiotic overuse in hospitals: Evaluation of perceptions and impact. J Antimicrob Chemother. 2018 Jun 1;73(6):1708–13.

47. Schweitzer VA, van Heijl I, van Werkhoven CH, Islam J, Hendriks-Spoor KD, Bielicki J, et al. The quality of studies evaluating antimicrobial stewardship interventions: a systematic review. Clinical Microbiology and Infection. 2019 May;25(5):555–61.

48. Department of Health and Social Care. Health and Social Care Act 2008: code of practice on the prevention and control of infections. 2022. Available from: https://www.gov.uk/government/publications/the-health-and-social-care-act-2008-code-of-practice-on-the-prevention-and-control-of-infections-and-related-guidance

49. Ashiru-Oredope D, Sharland M, Charani E, McNulty C, Cooke J, ARHAI Antimicrobial Stewardship Group. Improving the quality of antibiotic prescribing in the NHS by developing a new Antimicrobial Stewardship Programme: Start Smart--Then Focus. J Antimicrob Chemother. 2012 Jul 1;67 Suppl 1(suppl 1):i51-63.

50. Janssen RME, Oerlemans AJM, Van Der Hoeven JG, Ten Oever J, Schouten JA, Hulscher MEJL. Why we prescribe antibiotics for too long in the hospital setting: a systematic scoping review. J Antimicrob Chemother. 2022 Aug 1;77(8):2105.

51. Charani E, Ahmad R, Rawson TM, Castro-Sanchèz E, Tarrant C, Holmes AH. The Differences in Antibiotic Decision-making Between Acute Surgical and Acute Medical Teams: An Ethnographic Study of Culture and Team Dynamics. Clin Infect Dis. 2019 Jun 18;69(1):12–20.

52. Roope LSJ, Buchanan J, Morrell L, Pouwels KB, Sivyer K, Mowbray F, et al. Why do hospital prescribers continue antibiotics when it is safe to stop? Results of a choice experiment survey. BMC Med. 2020 Jul 30;18(1):196.

53. Warreman EB, Lambregts MMC, Wouters RHP, Visser LG, Staats H, van Dijk E, et al. Determinants of in-hospital antibiotic prescription behaviour: a systematic review and formation of a comprehensive framework. Clin Microbiol Infect. 2019. May;25(5):538-45.

54. Charani E, Castro-Sanchez E, Sevdalis N, Kyratsis Y, Drumright L, Shah N, et al. Understanding the determinants of antimicrobial prescribing within hospitals: The role of "prescribing etiquette." Clinical Infectious Diseases. 2013 Jul 15;57(2):188–96.

55. Ashiru-Oredope D, Budd EL, Doble A, Cramp E, Hendrick A, Hopkins S. Development and implementation of a national antimicrobial stewardship surveillance system, with open access data sharing. J Hosp Infect. 2021 Jan 1;107:16–22.

56. Fawcett NJ, Jones N, Quan TP, Mistry V, Crook D, Peto T, et al. Antibiotic use and clinical outcomes in the acute setting under management by an infectious diseases acute physician versus other clinical teams: a cohort study. BMJ Open. 2016 Aug 23;6(8):e010969.

57. Powell N, Stephens J, Rule R, Phillips R, Morphew M, Garry E, et al. Potential to reduce antibiotic use in secondary care: Single-centre process audit of prescription duration using NICE guidance for common infections. Clinical Medicine. 2021 Jan 1;21(1):E39–44.

58. Bitterman R, Hussein K, Leibovici L, Carmeli Y, Paul M. Systematic review of antibiotic consumption in acute care hospitals. Clinical Microbiology and Infection. 2016 Jun 1;22(6):561.e7-561.e19.

59. Zanichelli V, Monnier A, Gyssens I, Adriaenssens N, Versporten A, Pulcini C, et al. Variation in Antibiotic Use Among and Within Different Settings: A Systematic Review. J Antimicrob Chemother. 2018;73(6):vi17–29.

60. Llewelyn MJ, Hand K, Hopkins S, Sarah Walker A. Antibiotic policies in acute English NHS trusts: Implementation of "Start Smart-Then Focus" and relationship with Clostridium difficile infection rates. J Antimicrob Chemother. 2014 Sep 16;70(4):1230–5.

61. Onakpoya I, Walker A, Tan P, Spencer E, Gbinigie O, Cook J, et al. Overview of systematic reviews assessing the evidence for shorter versus longer duration antibiotic treatment for bacterial infections in secondary care. PLoS One. 2018;13(3):e0194858.

62. Evans SR, Follmann D, Liu Y, Holland T, Doernberg SB, Rouphael N, et al. Sequential, Multiple-Assignment, Randomized Trials for COMparing Personalized Antibiotic StrategieS (SMART-COMPASS). Clin Infect Dis. 2019 May 17;68(11):1961–7.

63. Pouwels KB, Yin M, Butler CC, Cooper BS, Wordsworth S, Walker AS, et al. Optimising trial designs to identify appropriate antibiotic treatment durations. BMC Med. 2019 Dec 21;17(1):115.

64. Quartagno M, Walker AS, Carpenter JR, Phillips PPJ, Parmar MKB. Rethinking non-inferiority: a practical trial design for optimising treatment duration. Clinical trials. 2018 Oct 1;15(5):477–88.

65. McDonald EG, Prosty C, Hanula R, Bortolussi-Courval É, Albuquerque AM, Tong SYC, et al. Observational versus randomized controlled trials to inform antibiotic treatment durations: a narrative review. Clin Microbiol Infect. 2023 Feb 1;29(2):165–70.

66. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I. Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research. BMJ. 2013 Nov 11;347:f6409.

67. Golder S, Loke Y, Bland M. Meta-analyses of Adverse Effects Data Derived from Randomised Controlled Trials as Compared to Observational Studies: Methodological Overview. PLoS Med. 2011;8(5):e1001026.

68. Mahar RK, McGuinness MB, Chakraborty B, Carlin JB, IJzerman MJ, Simpson JA. A scoping review of studies using observational data to optimise dynamic treatment regimens. BMC Med Res Methodol. 2021 Feb 22;21(1):39.

69. Lim C, Mo Y, Teparrukkul P, Hongsuwan M, Day NPJ, Limmathurotsakul D, et al. Effect of Delays in Concordant Antibiotic Treatment on Mortality in Patients With Hospital-Acquired Acinetobacter Species Bacteremia: Emulating a Target Randomized Trial With a 13-Year Retrospective Cohort. Am J Epidemiol. 2021 Nov 2;190(11):2395–404.

70. Wang S V, Schneeweiss S, RCT-DUPLICATE Initiative, Franklin JM, Desai RJ, Feldman W, et al. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials. JAMA. 2023 Apr 25;329(16):1376–85.

71. Danaei G, Rodríguez LAG, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. J Clin Epidemiol. 2018 Apr;96:12–22.

72. Hernán MA. Methods of Public Health Research - Strengthening Causal Inference from Observational Data. N Engl J Med. 2021 Oct 7;385(15):1345–8.

73. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758–64.

74. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016 Nov 1;79:70–5.

75. Suissa S. Immortal time bias in pharmacoepidemiology. Am J Epidemiol. 2008 Feb;167(4):492–9.

76. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. BMJ. 2018 Feb 1;360:k182.

77. Airaksinen J, Pentti J, Suominen S, Vahtera J, Kivimäki M. An example of how immortal time bias can reverse the results of an observational study. Epidemiology. 2020 Mar;31(2):e19-e20.

78. Hernán MA. Counterpoint: Epidemiology to Guide Decision-Making: Moving Away From Practice-Free Research. Am J Epidemiol. 2015 Oct 12;182(10):834.

79. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol. 2003 Nov 1;158(9):915–20.

80. Nguyen VT, Engleton M, Davison M, Ravaud P, Porcher R, Boutron I. Risk of bias in observational studies using routinely collected data of comparative effectiveness research: a meta-research study. BMC Med. 2021 Nov 23;19(1):279.

81. Luijken K, Spekreijse JJ, van Smeden M, Gardarsdottir H, Groenwold RHH. New-user and prevalent-user designs and the definition of study time origin in pharmacoepidemiology: A review of reporting practices. Pharmacoepidemiol Drug Saf. 2021 Jul 1;30(7):960.

82. Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 2007 Mar;16(3):241–9.

83. Munoz-Price LS, Frencken JF, Tarima S, Bonten M. Handling Time-dependent Variables: Antibiotics and Antibiotic Resistance. Clinical Infectious Diseases. 2016 Jun 15;62(12):1558–63.

84. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol. 1986 Sep;15(3):413–9.

85. Greenland S, Robins J, Pearl J. Confounding and Collapsibility in Causal Inference. Statistical Science. 1999;14(1):29–46.

86. Walker AS, Mason A, Quan TP, Fawcett NJ, Watkinson P, Llewelyn M, et al. Mortality risks associated with emergency admissions during weekends and public holidays: an analysis of electronic health records. Lancet. 2017 Jul 1;390(10089):62–72.

87. Kyriazopoulou E, Giamarellos-Bourboulis EJ. Antimicrobial Stewardship Using Biomarkers: Accumulating Evidence for the Critically Ill. Antibiotics. 2022 Mar 9;11(3):367.

88. Petel D, Winters N, Gore GC, Papenburg J, Beltempo M, Lacroix J, et al. Use of C-reactive protein to tailor antibiotic use: a systematic review and meta-analysis. BMJ Open. 2018 Dec 1;8(12).

89. Schuetz P, Wirz Y, Sager R, Christ-Crain M, Stolz D, Tamm M, et al. Effect of procalcitonin-guided antibiotic treatment on mortality in acute respiratory infections: a patient level meta-analysis. Lancet Infect Dis. 2018;18(1):95–107.

90. Meier MA, Branche A, Neeser OL, Wirz Y, Haubitz S, Bouadma L, et al. Procalcitonin-guided Antibiotic Treatment in Patients With Positive Blood Cultures: A Patient-level Meta-analysis of Randomized Trials. Clinical Infectious Diseases. 2019 Jul 18;69(3):388–96.

91. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000 Sep;11(5):550–60.

92. Daniel RM, Cousens SN, De Stavola BL, Kenward MG, Sterne JAC. Methods for dealing with time-dependent confounding. Stat Med. 2013 Apr 30;32(9):1584–618.

93. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. BMJ. 2017;359:j4587.

94. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. Int J Epidemiol. 2017;46(2):756–62.

95. Clare PJ, Dobbins TA, Mattick RP. Causal models adjusting for time-varying confounding-a systematic review of the literature. Int J Epidemiol. 2019 Feb 1;48(1):254–65.

96. Farmer RE, Kounali D, Walker AS, Savović J, Richards A, May MT, et al. Application of causal inference methods in the analyses of randomised controlled trials: a systematic review. Trials. 2018 Dec 10;19(1):23.

97. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ. 2010 Apr 24;340(7752):907–11.

98. Evans RN, Harris J, Rogers CA, Macgowan AP. The effect of duration of therapy for treatment of Staphylococcus aureus blood stream infection: an application of cloning to deal with immortal-time bias in an analysis of data from a cohort study (BSI-FOO). J Antimicrob Chemother. 2022 Jan 5;78(1):196–204.

99. Zuercher P, Moser A, Frey MC, Pagani JL, Buetti N, Eggimann P, et al. The effect of duration of antimicrobial treatment for bacteremia in critically ill patients on in-hospital mortality - Retrospective double center analysis. J Crit Care. 2023 Apr 1;74.

100. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004 Sep;15(5):615–25.

101. Greenland S, Pearl J, Robins J. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1):37–48.

102. Schisterman EF, Cole SR, Platf RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology. 2009 Jul;20(4):488–95.

103. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003;14(3):300–6.

104. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008 Nov;19(6):766–79.

105. Llewelyn M, Budgell E, Laskawiec-Szkonter M, Cross E, Alexander R, Bond S, et al. Antibiotic review kit for hospitals (ARK-Hospital): a stepped-wedge cluster-randomised controlled trial. Lancet Infect Dis. 2023 Feb;23(2):207–21.

106. Budgell EP, Davies TJ, Donker T, Hopkins S, Wyllie DH, Peto TEA, et al. Impact of antibiotic use on patient-level risk of death in 36 million hospital admissions in England. J Infect. 2022 Mar 1;84(3):311–20.

107. Tamma PD, Avdic E, Li DX, Dzintars K, Cosgrove SE. Association of Adverse Events With Antibiotic Use in Hospitalized Patients. JAMA Intern Med. 2017 Sep 1;177(9):1308.

108. Curran J, Lo J, Leung V, Brown K, Schwartz KL, Daneman N, et al. Estimating daily antibiotic harms: an umbrella review with individual study meta-analysis. Clin Microbiol Infect. 2022 Apr 1;28(4):479–90.

109. Tennyson L, Averch T. An update on fluoroquinolones: the emergence of a multisystem toxicity syndrome. Urol Pract. 2017;4(5):386–7.

110. Gao Z, Chen Y, Guan MX. Mitochondrial DNA mutations associated with aminoglycoside induced ototoxicity. J Otol. 2017 Mar;12(1):1–8.

111. Kalghatgi S, Spina CS, Costello JC, Liesa M, Morones-Ramirez JR, Slomovic S, et al. Bactericidal Antibiotics Induce Mitochondrial Dysfunction and Oxidative Damage in Mammalian Cells. Sci Transl Med. 2013 Jul 3;5(192):192ra85-192ra85.

112. Malik U, Armstrong D, Ashworth M, Dregan A, L'Esperance V, McDonnell L, et al. Association between prior antibiotic therapy and subsequent risk of community-acquired infections: a systematic review. J Antimicrob Chemother. 2018 Feb 1;73(2):287–96.

113. Brown KA, Khanafer N, Daneman N, Fisman DN. Meta-Analysis of Antibiotics and the Risk of Community-Associated Clostridium difficile Infection. Antimicrob Agents Chemother. 2013 May;57(5):2326–32.

114. Deshpande A, Pasupuleti V, Thota P, Pant C, Rolston DDK, Sferra TJ, et al. Community-associated Clostridium difficile infection and antibiotics: a meta-analysis. J Antimicrob Chemother. 2013 Sep;68(9):1951–61.

115. Slimings C, Riley T V. Antibiotics and hospital-acquired Clostridium difficile infection: update of systematic review and meta-analysis. J Antimicrob Chemother. 2014 Apr;69(4):881–91.

116. Rhee C, Kadri SS, Dekker JP, Danner RL, Chen HC, Fram D, et al. Prevalence of Antibiotic-Resistant Pathogens in Culture-Proven Sepsis and Outcomes Associated With Inadequate and Broad-Spectrum Empiric Antibiotic Use. JAMA Netw Open. 2020 Apr 1;3(4):e202899.

117. Levy MM, Dellinger RP, Townsend SR, Linde-Zwirble WT, Marshall JC, Bion J, et al. The Surviving Sepsis Campaign: Results of an international guideline-based performance improvement program targeting severe sepsis. Crit Care Med. 2010 Feb;38(2):367–74.

118. Ashiru-Oredope D, Sharland M, Charani E, McNulty C, Cooke J, ARHAI Antimicrobial Stewardship Group. Improving the quality of antibiotic prescribing in the NHS by developing a new Antimicrobial Stewardship Programme: Start Smart--Then Focus. J Antimicrob Chemother. 2012 Jul 1;67(suppl 1):i51–63.

119. Centers for Disease Control and Prevention. Core Elements of Hospital Antibiotic Stewardship Programs. Atlanta, GA: US Department of Health and Human Services, CDC. 2019. Available from: https://www.cdc.gov/antibiotic-use/core-elements/hospital.html#_ENREF_4

120. Versporten A, Zarb P, Caniaux I, Gros M, Drapier N, Miller M, et al. Antimicrobial consumption and resistance in adult hospital inpatients in 53 countries: results of an internet-based global point prevalence survey. Lancet Glob Health. 2018;6(6):e619–29.

121. Cross ELA, Sivyer K, Islam J, Santillo M, Mowbray F, Peto TEA, et al. Adaptation and implementation of the ARK (Antibiotic Review Kit) intervention to safely and substantially reduce antibiotic use in hospitals: a feasibility study. J Hosp Infect. 2019 Nov;103(3):268–75.

122. Powell N, McGraw-Allen K, Menzies A, Peet B, Simmonds C, Wild A. Identifying antibiotic stewardship interventions to meet the NHS England CQUIN: an evaluation of antibiotic - prescribing against published evidence-based antibiotic audit tools. Clinical Medicine. 2018 Aug 2;18(4):276–81.

123. Braykov NP, Morgan DJ, Schweizer ML, Uslan DZ, Kelesidis T, Weisenberg SA, et al. Assessment of empirical antibiotic therapy optimisation in six hospitals: An observational cohort study. Lancet Infect Dis. 2014;14(12):1220–7.

124. Hood G, Hand KS, Cramp E, Howard P, Hopkins S, Ashiru-Oredope D. Measuring appropriate antibiotic prescribing in acute hospitals: Development of a national audit tool through a Delphi consensus. Antibiotics. 2019 Jun 1;8(2).

125. Fridkin SK, Lawton R, Edwards JR, Tenover FC, McGowan JE, Gaynes RP. Monitoring antimicrobial use and resistance: Comparison with a national benchmark on reducing vancomycin use and vancomycin-resistant enterococci. Emerg Infect Dis. 2002;8(7):702–7.

126. Ibrahim OM, Polk RE. Benchmarking antimicrobial drug use in hospitals. Expert Rev Anti Infect Ther. 2012 Apr;10(4):445-57.

127. Kuster SP, Ruef C, Bollinger AK, Ledergerber B, Hintermann A, Deplazes C, et al. Correlation between case mix index and antibiotic use in hospitals. J Antimicrob Chemother. 2008;62(4):837–42.

128. NHS England. NHS Standard Contract 2021/22 Service Conditions. 2021. Available from: https://www.england.nhs.uk/nhs-standard-contract/previous-nhs-standard-contracts/21-22/

129. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). Int J Epidemiol. 2017;46(4):1093–1093i.

130. Imison C, Vaughan L. Acute medical care in England: Findings from a survey of smaller acute hospitals. Slide-set resource. 2018. Available from: www.nuffieldtrust.org.uk/research/acute-medical-care-in-england-findings-from-a-survey-of-smaller-acute-hospitals

131. NHS Digital. Summary Hospital-level Mortality Indicator (SHMI) Provider Spells Methodology.. 2022. Available from: https://digital.nhs.uk/data-and-information/publications/ci-hub/summary-hospital-level-mortality-indicator-shmi

132. IQVIA. Welcome to IQVIA. Available from: https://www.iqvia.com

133. WHO Collaborating Centre for Drug Statistics Methodology. International language for drug utilization research. 2019. Available from: https://www.whocc.no

134. UK Health Security Agency. MRSA bacteraemia: annual data. Annual counts and rates of meticillin resistant Staphylococcus aureus (MRSA) bacteraemia by acute trust and clinical commissioning group (CCG). 2018. Available from: https://www.gov.uk/government/statistics/mrsa-bacteraemia-annual-data

135. NHS England. Bed Availability and Occupancy Quarterly Return Definitions (KH03). 2023. Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/

136. NHS Digital. Hospital admitted patient care activity: provider level analysis. 2017. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity

137. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Med. 2015;12(10):e1001885.

138. NHS Digital. HES Data Dictionary Admitted Patient Care. 2018. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary

139. Thompson S, Higgins J. How should meta-regression analyses be undertaken and interpreted? Stat Med. 2002;21(11):1559–73.

140. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. 2004 Apr 1;159(7):702–6.

141. Freemantle N, Richardson M, Wood J, Ray D, Khosla S, Shahian D, et al. Weekend hospitalization and additional risk of death: an analysis of inpatient data. J R Soc Med. 2012 Feb;105(2):74–84.

142. Onukwugha E, Bergtold J, Jain R. A Primer on Marginal Effects—Part I: Theory and Formulae. Pharmacoeconomics. 2015 Jan 5;33(1):25–30.

143. Norton EC, Dowd BE, Maciejewski ML. Marginal Effects—Quantifying the Effect of Changes in Risk Factors in Logistic Regression Models. JAMA. 2019 Apr 2;321(13):1304.

144. Harbord R, Higgins J. Meta-regression in Stata. Stata J. 2008;8(4):493–519.

145. Greenland S, O'Rourke K. Meta-Analysis. In: Rothman K, Greenland S, Lash T, editors. Modern Epidemiology. 3rd ed. Philadelphia, USA: Lippincott Williams & Wilkins; 2008. p. 652–82.

146. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random-effects regression model for meta-analysis. Stat Med. 1995 Feb 28;14(4):395–411.

147. Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. Stat Med. 2009 Apr 30;29(12):1282–97.

148. Royer S, Demerle KM, Dickson RP, Prescott HC. Shorter Versus Longer Courses of Antibiotics for Infection in Hospitalized Patients: A Systematic Review and Meta-Analysis. J Hosp Med. 2018 May 1;13(5):336–42.

149. Sawyer RG, Claridge JA, Nathens AB, Rotstein OD, Duane TM, Evans HL, et al. Trial of Short-Course Antimicrobial Therapy for Intraabdominal Infection. New England Journal of Medicine. 2015 May 21;372(21):1996–2005.

150. Uranga A, España PP, Bilbao A, Quintana JM, Arriaga I, Intxausti M, et al. Duration of Antibiotic Treatment in Community-Acquired Pneumonia. JAMA Intern Med. 2016 Sep 1;176(9):1257.

151. Palin V, Welfare W, Ashcroft DM, Van Staa TP. Shorter and Longer Courses of Antibiotics for Common Infections and the Association With Reductions of Infection-Related Complications Including Hospital Admissions. Clinical Infectious Diseases. 2021 Nov 15;73(10):1805–12.

152. Spellberg B, Srinivasan A, Chambers HF. New Societal Approaches to Empowering Antibiotic Stewardship. JAMA. 2016;315(12):1229–30.

153. Dawson-Hahn EE, Mickan S, Onakpoya I, Roberts N, Kronman M, Butler CC, et al. Short-course versus long-course oral antibiotic treatment for infections treated in outpatient settings: a review of systematic reviews. Fam Pract. 2017 Oct 1;34(5):511–9.

154. NHS England. NHS Standard Contract 2022/23 Service Conditions. 2022. Available from: https://www.england.nhs.uk/nhs-standard-contract/22-23/

155. Safi W, Elnegouly M, Schellnegger R, Umgelter K, Geisler F, Reindl W, et al. Infection and Predictors of Outcome of Cirrhotic Patients after Emergency Care Hospital Admission. Ann Hepatol. 2018 Oct;17(6):948–58.

156. Health Protection Agency. English national point prevalence survey on healthcare-associated infections and antimicrobial use, 2011. 2012. Available from: https://webarchive.nationalarchives.gov.uk/20140714085429/http://www.hpa.org.uk/Publications/InfectiousDiseases/AntimicrobialAndHealthcareAssociatedInfections/1205HCAIEnglishPPSforhcaiandamu2011prelim/

157. Ibrahim OM, Polk RE. Antimicrobial use metrics and benchmarking to improve stewardship outcomes: methodology, opportunities, and challenges. Infect Dis Clin North Am. 2014 Jun;28(2):195–214. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24857388

158. Health and Social Care Information Centre. Methodology for Creation of the HES Patient ID (HESID). 2014. Available from: http://content.digital.nhs.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HES-patient-ID/pdf/HESID_Methodology.pdf

159. NHS Digital. Summary Hospital-level Mortality Indicator (SHMI) Indicator specification. 2023. Available from: https://digital.nhs.uk/data-and-information/publications/ci-hub/summary-hospital-level-mortality-indicator-shmi

160. Binder H, Sauerbrei W, Royston P. Comparison between splines and fractional polynomials for multivariable model building with continuous covariates: a simulation study with continuous response. Stat Med. 2013 Jun 15;32(13):2262–77.

161. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015.

162. Hanretty AM, Gallagher JC. Shortened Courses of Antibiotics for Bacterial Infections: A Systematic Review of Randomized Controlled Trials. Pharmacotherapy. 2018 Jun 1;38(6):674–87.

163. Wald-Dickler N, Spellberg B. Short-course Antibiotic Therapy—Replacing Constantine Units With “Shorter Is Better.” Clin Infect Dis. 2019 Nov 11;69(9):1476.

164. Lee R, Centor R, Humphrey L, Jokela J, Andrews R, Qaseem A, et al. Appropriate Use of Short-Course Antibiotics in Common Infections: Best Practice Advice From the American College of Physicians. Ann Intern Med. 2021 Jun 1;174(6):822–7

165. Lee RA, Stripling JT, Spellberg B, Centor RM. Short-course antibiotics for common infections: what do we know and where do we go from here? Clin Microbiol Infect. 2023 Feb 1;29(2):150–9.

166. Li Q, Zhou Q, Florez ID, Mathew JL, Shang L, Zhang G, et al. Short-Course vs Long-Course Antibiotic Therapy for Children With Nonsevere Community-Acquired Pneumonia: A Systematic Review and Meta-analysis. JAMA Pediatr. 2022 Dec 5;176(12):1199–207.

167. Dinh A, Ropers J, Duran C, Davido B, Deconinck L, Matt M, et al. Discontinuing β-lactam treatment after 3 days for patients with community-acquired pneumonia in non-critical care

wards (PTC): a double-blind, randomised, placebo-controlled, non-inferiority trial. Lancet. 2021 Mar 27;397(10280):1195–203.

168. Furukawa Y, Luo Y, Funada S, Onishi A, Ostinelli E, Hamza T, et al. Optimal duration of antibiotic treatment for community-acquired pneumonia in adults: a systematic review and duration-effect meta-analysis. BMJ Open. 2023;13(3):e061023.

169. Tansarli GS, Mylonakis E. Systematic Review and Meta-analysis of the Efficacy of Short-Course Antibiotic Treatments for Community-Acquired Pneumonia in Adults. Antimicrob Agents Chemother. 2018 Sep 1;62(9)e00635-18.

170. Pugh R, Grant C, Cooke RPD, Dempsey G. Short-course versus prolonged-course antibiotic therapy for hospital-acquired pneumonia in critically ill adults. Cochrane Database Syst Rev. 2015 Aug 24;2015(8)CD007577.

171. Hassinger TE, Guidry CA, Rotstein OD, Duane TM, Evans HL, Cook CH, et al. Longer-Duration Antimicrobial Therapy Does Not Prevent Treatment Failure in High-Risk Patients with Complicated Intra-Abdominal Infections. Surg Infect. 2017 Aug 1;18(6):659–63.

172. Montravers P, Tubach F, Lescot T, Veber B, Esposito-Farèse M, Seguin P, et al. Short-course antibiotic therapy for critically ill patients treated for postoperative intra-abdominal infection: the DURAPOP randomised clinical trial. Intensive Care Med. 2018 Mar 26;44(3):300–10.

173. Hepburn MJ, Dooley DP, Skidmore PJ, Ellis MW, Starnes WF, Hasewinkle WC. Comparison of short-course (5 days) and standard (10 days) treatment for uncomplicated cellulitis. Arch Intern Med. 2004 Aug 9;164(15):1669–74.

174. Huang CY, Hsieh RW, Yen HT, Hsu TC, Chen CY, Chen YC, et al. Short- versus long-course antibiotics in osteomyelitis: A systematic review and meta-analysis. Int J Antimicrob Agents. 2019 Mar;53(3):246–60.

175. Gariani K, Pham TT, Kressmann B, Jornayvaz FR, Gastaldi G, Stafylakis D, et al. Three Weeks Versus Six Weeks of Antibiotic Therapy for Diabetic Foot Osteomyelitis: A Prospective, Randomized, Noninferiority Pilot Trial. Clin Infect Dis. 2021 Oct 1;73(7):E1539–45.

176. Tone A, Nguyen S, Devemy F, Topolinski H, Valette M, Cazaubiel M, et al. Six-week versus twelve-week antibiotic therapy for nonsurgically treated diabetic foot osteomyelitis: a multicenter open-label controlled randomized study. Diabetes Care. 2015 Feb 1;38(2):302–7.

177. Bernard L, Dinh A, Ghout I, Simo D, Zeller V, Issartel B, et al. Antibiotic treatment for 6 weeks versus 12 weeks in patients with pyogenic vertebral osteomyelitis: an open-label, non-inferiority, randomised, controlled trial. Lancet. 2015 Mar 7;385(9971):875–82.

178. McAteer J, Lee JH, Cosgrove SE, Dzintars K, Fiawoo S, Heil EL, et al. Defining the Optimal Duration of Therapy for Hospitalized Patients With Complicated Urinary Tract Infections and Associated Bacteremia. Clin Infect Dis. 2023 May 3;76(9).

179. Drekonja DM, Trautner B, Amundson C, Kuskowski M, Johnson JR. Effect of 7 vs 14 Days of Antibiotic Therapy on Resolution of Symptoms Among Afebrile Men With Urinary Tract Infection: A Randomized Clinical Trial. JAMA. 2021 Jul 27;326(4):324–31.

180. Eliakim-Raz N, Yahav D, Paul M, Leibovici L. Duration of antibiotic treatment for acute pyelonephritis and septic urinary tract infection—7 days or less versus longer treatment:

systematic review and meta-analysis of randomized controlled trials. J Antimicrob Chemother. 2013 Oct;68(10):2183–91.

181. Molina J, Montero-Mateos E, Praena-Segovia J, León-Jiménez E, Natera C, López-Cortés LE, et al. Seven-versus 14-day course of antibiotics for the treatment of bloodstream infections by Enterobacterales: a randomized, controlled trial. Clin Microbiol Infect. 2022 Apr 1;28(4):550–7.

182. Von Dach E, Albrich WC, Brunel AS, Prendki V, Cuvelier C, Flury D, et al. Effect of C-Reactive Protein-Guided Antibiotic Treatment Duration, 7-Day Treatment, or 14-Day Treatment on 30-Day Clinical Failure Rate in Patients With Uncomplicated Gram-Negative Bacteremia: A Randomized Clinical Trial. JAMA. 2020 Jun 2;323(21):2160–9.

183. Yahav D, Franceschini E, Koppel F, Turjeman A, Babich T, Bitterman R, et al. Seven Versus 14 Days of Antibiotic Therapy for Uncomplicated Gram-negative Bacteremia: A Noninferiority Randomized Controlled Trial. Clin Infect Dis. 2019 Sep 13;69(7):1091–8.

184. Falagas ME, Karageorgopoulos DE, Grammatikos AP, Matthaiou DK. Effectiveness and safety of short vs. long duration of antibiotic therapy for acute bacterial sinusitis: a meta-analysis of randomized trials. Br J Clin Pharmacol. 2009 Feb;67(2):161–71.

185. Pouwels KB, Hopkins S, Llewelyn MJ, Walker AS, McNulty CA, Robotham J V. Duration of antibiotic treatment for common infections in English primary care: cross sectional analysis and comparison with guidelines. BMJ. 2019;364.

186. King LM, Hersh AL, Hicks LA, Fleming-Dutra KE. Duration of Outpatient Antibiotic Therapy for Common Outpatient Infections, 2017. Clinical Infectious Diseases. 2021 May 18;72(10):e663–6.

187. Magill SS, O'Leary E, Ray SM, Kainer MA, Evans C, Bamberg WM, et al. Assessment of the Appropriateness of Antimicrobial Use in US Hospitals. JAMA Netw Open. 2021 Mar 18;4(3).

188. Vaughn VM, Flanders SA, Snyder A, Conlon A, Rogers MA, Malani AN, et al. Excess Antibiotic Treatment Duration and Adverse Events in Patients Hospitalized With Pneumonia. Ann Intern Med. 2019 Aug 6;171(3):153.

189. NHS Digital. Summary Hospital-level Mortality Indicator (SHMI). ICD-10 to SHMI diagnosis group lookup table. 2020. Available from: https://digital.nhs.uk/data-and-information/publications/ci-hub/summary-hospital-level-mortality-indicator-shmi

190. Ammerlaan HSM, Harbarth S, Buiting AGM, Crook DW, Fitzpatrick F, Hanberger H, et al. Secular trends in nosocomial bloodstream infections: antibiotic-resistant bacteria increase the total burden of infection. Clin Infect Dis. 2013 Mar 15;56(6):798–805.

191. Rubin D. Inference and missing data. Biometrika. 1976;63(3):581–90.

192. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009 Sep 1;338(jun29 1):b2393–b2393.

193. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):237–42.

194. De Bus L, Depuydt P, Steen J, Dhaese S, De Smet K, Tabah A, et al. Antimicrobial de-escalation in the critically ill patient and assessment of clinical cure: the DIANA study. Intensive Care Med. 2020 Jul 1;46(7):1404.

195. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ. Selection Bias Due to Loss to Follow Up in Cohort Studies. Epidemiology. 2016 Jan 1;27(1):91–7.

196. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. Int J Biostat. 2010;6(2):18.

197. Lash L, Fox M, Fink A. Applying quantitative bias analysis to epidemiologic data. New York: Springer; 2009.

198. Haneuse S, Vanderweele TJ, Arterburn D. Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. JAMA. 2019 Feb 12;321(6):602–3.

199. WHO Collaborating Centre for Drug Statistics Methodology. DDD alterations from 2005-2022. 2023. Available from: https://www.whocc.no/atc_ddd_alterations__cumulative/ddd_alterations

200. National Institute for Health and Care Excellence. Antimicrobial stewardship: systems and processes for effective antimicrobial medicine use. 2015. Available from: https://www.nice.org.uk/guidance/ng15

201. Davey P, Sneddon J, Nathwani D. Overview of strategies for overcoming the challenge of antimicrobial resistance. Expert Rev Clin Pharmacol. 2010 Sep;3(5):667–86.

202. Dyar OJ, Huttner B, Schouten J, Pulcini C. What is antimicrobial stewardship? Clinical Microbiology and Infection. 2017 Nov 1;23(11):793–8.

203. Huebner C, Flessa S, Huebner NO. The economic impact of antimicrobial stewardship programmes in hospitals: a systematic literature review. J Hosp Infect. 2019 Aug 1;102(4):369–76.

204. Nathwani D, Varghese D, Stephens J, Ansari W, Martin S, Charbonneau C. Value of hospital antimicrobial stewardship programs [ASPs]: a systematic review. Antimicrob Resist Infect Control. 2019 Feb 12;8(1):35.

205. Karanika S, Paudel S, Grigoras C, Kalbasi A, Mylonakis E. Systematic Review and Meta-analysis of Clinical and Economic Outcomes from the Implementation of Hospital-Based Antimicrobial Stewardship Programs. Antimicrob Agents Chemother. 2016 Aug;60(8):4840–52.

206. Baur D, Gladstone BP, Burkert F, Carrara E, Foschi F, Döbele S, et al. Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and Clostridium difficile infection: a systematic review and meta-analysis. Lancet Infect Dis. 2017 Sep 1;17(9):990–1001.

207. Davey P, Marwick CA, Scott CL, Charani E, McNeil K, Brown E, et al. Interventions to improve antibiotic prescribing practices for hospital inpatients. Cochrane Database Syst Rev. 2017 Feb 9;2(2):CD003543.

208. Stuart B, Hounkpatin H, Becque T, Yao G, Zhu S, Alonso-Coello P, et al. Delayed antibiotic prescribing for respiratory tract infections: individual patient data meta-analysis. BMJ. 2021 Apr 28;373.

209. Spurling GKP, Del Mar CB, Dooley L, Foxlee R, Farley R. Delayed antibiotic prescriptions for respiratory infections. Cochrane Database Syst Rev. 2017 Sep 7;9(9):CD004417.

210. Little P, Moore M, Kelly J, Williamson I, Leydon G, McDermott L, et al. Delayed antibiotic prescribing strategies for respiratory tract infections in primary care: pragmatic, factorial, randomised controlled trial. BMJ. 2014 Mar 6;348.

211. Little P, Stuart B, Moore M, Coenen S, Butler CC, Godycki-Cwirko M, et al. Amoxicillin for acute lower-respiratory-tract infection in primary care when pneumonia is not suspected: a 12-country, randomised, placebo-controlled trial. Lancet Infect Dis. 2013 Feb;13(2):123–9.

212. Van Staa TP, Palin V, Brown B, Welfare W, Li Y, Ashcroft DM. The Safety of Delayed Versus Immediate Antibiotic Prescribing for Upper Respiratory Tract Infections. Clinical Infectious Diseases. 2021 Jul 15;73(2):E394–401.

213. Royal College of General Practitioners. TARGET antibiotics toolkit hub. 2021. Available from: https://elearning.rcgp.org.uk/course/view.php?id=553

214. Scobie A, Budd EL, Harris RJ, Hopkins S, Shetty N. Antimicrobial stewardship: an evaluation of structure and process and their association with antimicrobial prescribing in NHS hospitals in England. J Antimicrob Chemother. 2019 Apr 1;74(4):1143–52.

215. Schweitzer VA, van Werkhoven CH, Rodríguez Baño J, Bielicki J, Harbarth S, Hulscher M, et al. Optimizing design of research to evaluate antibiotic stewardship interventions: consensus recommendations of a multinational working group. Clinical Microbiology and Infection. 2020 Jan 1;26(1):41–50.

216. McGregor JC, Furuno JP. Optimizing research methods used for the evaluation of antimicrobial stewardship programs. Clin Infect Dis. 2014 Oct 15;59(3):S185–92.

217. de Kraker MEA, Abbas M, Huttner B, Harbarth S. Good epidemiological practice: a narrative review of appropriate scientific methods to evaluate the impact of antimicrobial stewardship interventions. Clinical Microbiology and Infection. 2017 Nov 1;23(11):819–25.

218. Gillespie D, Francis NA, Carrol ED, Thomas-Jones E, Butler CC, Hood K. Use of co-primary outcomes for trials of antimicrobial stewardship interventions. Lancet Infect Dis. 2018 Jun 1;18(6):595–7.

219. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014 Mar 7;348.

220. Matuluko A, MacDonald J, Ness V, Currie K. Interventions to improve the review of antibiotic therapy in acute care hospitals: a systematic review and narrative synthesis. JAC Antimicrob Resist. 2020 Sep 1;2(3):dlaa065.

221. Duncan EM, Charani E, Clarkson JE, Francis JJ, Gillies K, Grimshaw JM, et al. A behavioural approach to specifying interventions: what insights can be gained for the reporting and

implementation of interventions to reduce antibiotic use in hospitals? J Antimicrob Chemother. 2020 May 1;75(5):1338.

222. Schuts E, Hulscher M, Mouton J, Verduin C, Stuart J, Overdiek H, et al. Current evidence on hospital antimicrobial stewardship objectives: a systematic review and meta-analysis. Lancet Infect Dis. 2016;16(7):847–56.

223. Hulscher MEJL, Prins JM. Antibiotic stewardship: does it work in hospital practice? A review of the evidence base. Clinical Microbiology and Infection. 2017 Nov 1;23(11):799–805.

224. Mowbray F, Sivyer K, Santillo M, Jones N, Peto TEA, Walker AS, et al. Patient engagement with antibiotic messaging in secondary care: a qualitative feasibility study of the "review and revise" experience. Pilot Feasibility Stud. 2020 Jan 6;6(1).

225. Santillo M, Sivyer K, Krusche A, Mowbray F, Jones N, Peto TEA, et al. Intervention planning for Antibiotic Review Kit (ARK): a digital and behavioural intervention to safely review and reduce antibiotic prescriptions in acute and general medicine. J Antimicrob Chemother. 2019 Nov 1;74(11):3362–70.

226. Walker AS, Budgell E, Laskawiec-Szkonter M, Sivyer K, Wordsworth S, Quaddy J, et al. Antibiotic Review Kit for Hospitals (ARK-Hospital): study protocol for a stepped-wedge cluster-randomised controlled trial. Trials. 2019 Dec 11;20(1):421.

227. British Society for Antimicrobial Chemotherapy. Infection Learning Hub: ARK - The Antibiotic Review Kit. 2022. Available from: https://infectionlearninghub.co.uk/ark-the-antibiotic-review-kit/

228. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials. 1986;7(3):177–88.

229. Deeks J, Higgins J, Altman D, editors. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors) Cochrane Handbook for Systematic Reviews of Interventions version 63 (updated February 2022) Cochrane. 2022. Available from: https://training.cochrane.org/handbook/current/chapter-10

230. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021 Sep 30;374.

231. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350:h1258.

232. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. Trials. 2013 Jan 12;14(1):15.

233. Hulscher MEJL, Laurant MGH, Grol RPTM. Process evaluation on quality improvement interventions. Qual Saf Health Care. 2003;12(1):40.

234. Rzewuska M, Charani E, Clarkson JE, Davey PG, Duncan EM, Francis JJ, et al. Prioritizing research areas for antibiotic stewardship programmes in hospitals: a behavioural perspective consensus paper. Clinical Microbiology and Infection. 2019 Feb 1;25(2):163–8.

235. Pouly E, Coppry M, Rogues AM, Dumartin C. Systematic review of factors promoting behaviour change toward antibiotic use in hospitals. Clinical Microbiology and Infection. 2022 Jul 1;28(7):911–9.

236. Schweder T, Spjotvøll E. Plots of P-values to evaluate many tests simultaneously. Biometrika. 1982;69(3):493–502.

237. Strom BL, Schinnar R, Aberra F, Bilker W, Hennessy S, Leonard CE, et al. Unintended effects of a computerized physician order entry nearly hard-stop alert to prevent a drug interaction: a randomized controlled trial. Arch Intern Med. 2010 Sep 27;170(17):1578–83.

238. Eljaaly K, Elarabi S, Alshehri S, Nix DE. Impact of requiring re-authorization of restricted antibiotics on day 3 of therapy. J Antimicrob Chemother. 2018 Feb 1;73(2):527–30.

239. Thom KA, Tamma PD, Harris AD, Dzintars K, Morgan DJ, Li S, et al. Impact of a Prescriber-driven Antibiotic Time-out on Antibiotic Use in Hospitalized Patients. Clinical Infectious Diseases. 2019 Apr 24;68(9):1581–4.

240. Lee TC, Frenette C, Jayaraman D, Green L, Pilote L. Antibiotic self-stewardship: trainee-led structured antibiotic time-outs to improve antimicrobial use. Ann Intern Med. 2014 Nov 18;161(10 Suppl):S53–8.

241. Mohayya SM, Narayanan N, Cimilluca D, Malanowski A, Vaidya P, Bhowmick T. Effectiveness of an Electronic Automated Antibiotic Time Out Alert in the Setting of Gram-Negative Bacteremia. Antibiotics. 2021 Sep 1;10(9).

242. Graber CJ, Jones MM, Glassman PA, Weir C, Butler J, Nechodom K, et al. Taking an Antibiotic Time-out: Utilization and Usability of a Self-Stewardship Time-out Program for Renewal of Vancomycin and Piperacillin-Tazobactam. Hosp Pharm. 2015;50(11):1011.

243. Centers for Disease Control and Prevention. Core Elements of Hospital Antibiotic Stewardship Programs. Atlanta, GA: US Department of Health and Human Services, CDC. 2019.

244. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. Stat Med. 2004 Jun 15;23(11):1663–82.

245. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine--reporting of subgroup analyses in clinical trials. N Engl J Med. 2007 Nov 22;357(21):2189–94.

246. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005 Jul 23;2(8):2–8.

247. Goeman JJ, Solari A. Multiple hypothesis testing in genomics. Stat Med. 2014 May 20;33(11):1946–78.

248. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. Nat Hum Behav. 2018 Jan 1;2(1):6–10.

249. Bland J, Altman D. Multiple significance tests: the Bonferroni method. BMJ. 1995 Jan 21;310(6973):170.

250. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics. 1979;6(2):65–70.

251. Romano J, Wolf M. Control of generalized error rates in multiple testing. The Annals of Statistics. 2007;35(4):1378–1408.

252. Lehmann E, Romano J. Generalizations of the familywise error rate. The Annals of Statistics. 2005;33(3):1138–1154.

253. European Medicines Agency. Guideline on multiplicity issues in clinical trials. 2017. Available from: https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf

254. Food and Drug Administration. Multiple Endpoints in Clinical Trials Guidance for Industry. 2017. Available from: https://www.fda.gov/files/drugs/published/Multiple-Endpoints-in-Clinical-Trials-Guidance-for-Industry.pdf

255. Rothman K. No adjustments are needed for multiple comparisons. Epidemiology. 1990;1(1):43–6.

256. Perneger T V. What's wrong with Bonferroni adjustments. BMJ. 1998 Apr 18;316(7139):1236–8.

257. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B. 1995;57(1):289–300.

258. Mc Cord KA, Al-Shahi Salman R, Treweek S, Gardner H, Strech D, Whiteley W, et al. Routinely collected data for randomized trials: promises, barriers, and implications. Trials. 2018 Jan 11;19(1).

259. Raman SR, Curtis LH, Temple R, Andersson T, Ezekowitz J, Ford I, et al. Leveraging electronic health records for clinical research. Am Heart J. 2018 Aug 1;202:13–9.

260. Dutey-Magni PF, Gill MJ, McNulty D, Sohal G, Hayward A, Shallcross L, et al. Feasibility study of hospital antimicrobial stewardship analytics using electronic health records. JAC Antimicrobial Resistance. 2021 Jan 18;3(1):dlab018.

261. Rezel-Potts E, Gulliford M. Electronic Health Records and Antimicrobial Stewardship Research: a Narrative Review. Curr Epidemiol Rep. 2022 Jul 21;1-10.

262. Mc Cord KA, Ewald H, Agarwal A, Glinz D, Aghlmandi S, Ioannidis JPA, et al. Treatment effects in randomised trials using routinely collected data for outcome assessment versus traditional trials: meta-research study. BMJ. 2021 Mar 3;372(1):n450.

263. Rawson TM, Charani E, Moore LSP, Hernandez B, Castro-Sánchez E, Herrero P, et al. Mapping the decision pathways of acute infection management in secondary care among UK medical physicians: a qualitative study. BMC Med. 2016 Dec 12;14(1):208.

264. Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice? Int J Epidemiol. 2020;49(3):1043.

265. Baio G, Copas A, Ambler G, Hargreaves J, Beard E, Omar RZ. Sample size calculation for a stepped wedge trial. Trials. 2015 Aug 17;16(1):354.

266. Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. J Clin Epidemiol. 2016 Jan 1;69:137-46.

267. Hemming K, Eldridge S, Forbes G, Weijer C, Taljaard M. How to design efficient cluster randomised trials. BMJ. 2017 Jul 14;358:j3064.

268. Hemming K, Taljaard M, Forbes A. Analysis of cluster randomised stepped wedge trials with repeated cross-sectional samples. Trials. 2017 Mar 4;18(1):101.

269. Bion J, Richardson A, Hibbert P, Beer J, Abrusci T, McCutcheon M, et al. "Matching Michigan": a 2-year stepped interventional programme to minimise central venous catheter-blood stream infections in intensive care units in England. BMJ Qual Saf. 2013 Feb;22(2):110–23.

270. NHS Digital. Hospital Admitted Patient Care Activity, 2021-22: Provider-level analysis. 2022. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity

271. NHS Digital. Hospital Episode Statistics Monthly Activity Report: 2018-19 to 2021/22. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/provisional-monthly-hospital-episode-statistics-for-admitted-patient-care-outpatient-and-accident-and-emergency-data/april-2022---november-2022

272. NHS England. Monthly Hospital Activity Data. Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/hospital-activity/monthly-hospital-activity/mar-data/

273. NHS England. Urgent and Emergency Care Daily Situation Reports. Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/uec-sitrep/

274. NHS England. Bed Availability and Occupancy. Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/critical-care-and-general-acute-beds-urgent-and-emergency-care-daily-situation-reports/

275. Goodman KE, Pineles L, Magder LS, Anderson DJ, Ashley ED, Polk RE, et al. Electronically Available Patient Claims Data Improve Models for Comparing Antibiotic Use Across Hospitals: Results From 576 US Facilities. Clin Infect Dis. 2021 Dec 1;73(11):E4484–92.

276. Yu KC, Moisan E, Tartof SY, Nguyen HM, Rieg G, Ramaprasad C, et al. Benchmarking Inpatient Antimicrobial Use: A Comparison of Risk-Adjusted Observed-to-Expected Ratios. Clin Infect Dis. 2018 Nov 13;67(11):1677–85.

277. Wattier RL, Thurm CW, Parker SK, Banerjee R, Hersh AL, Brogan T V., et al. Indirect Standardization as a Case Mix Adjustment Method to Improve Comparison of Children's Hospitals' Antimicrobial Use. Clin Infect Dis. 2021 Sep 1;73(5):925–32.

278. Kanerva M, Ollgren J, Lyytikäinen O, Agthe N, Möttönen T, Kauppinen M, et al. Benchmarking antibiotic use in Finnish acute care hospitals using patient case-mix adjustment. J Antimicrob Chemother. 2011 Nov;66(11):2651–4.

279. Polk RE, Hohmann SF, Medvedev S, Ibrahim O. Benchmarking risk-adjusted adult antibacterial drug use in 70 US academic medical center hospitals. Clin Infect Dis. 2011 Dec 1;53(11):1100–10.

280. O'Leary EN, Edwards JR, Srinivasan A, Neuhauser MM, Webb AK, Soe MM, et al. National Healthcare Safety Network Standardized Antimicrobial Administration Ratios (SAARs): A Progress Report and Risk Modeling Update Using 2017 Data. Clin Infect Dis. 2020 Dec 12;71(10):e702.

281. O'Leary EN, Van Santen KL, Edwards EM, Braun D, Buus-Frank ME, Edwards JR, et al. Using NHSN's Antimicrobial Use Option to Monitor and Improve Antibiotic Stewardship in Neonates. Hosp Pediatr. 2019 May 1;9(5):340.

282. Israelsen SB, Fally M, Tarp B, Kolte L, Ravn P, Benfield T. Short-course antibiotic therapy for hospitalized patients with early clinical response in community-acquired pneumonia: a multicentre cohort study. Clin Microbiol Infect. 2023 Jan 1;29(1):54–60.

283. Zhou Z, Rahme E, Abrahamowicz M, Pilote L. Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. Am J Epidemiol. 2005 Nov;162(10):1016–23.

284. Thorlacius-Ussing L, Sandholdt H, Nissen J, Rasmussen J, Skov R, Frimodt-Møller N, et al. Comparable Outcomes of Short-Course and Prolonged-Course Therapy in Selected Cases of Methicillin-Susceptible Staphylococcus aureus Bacteremia: A Pooled Cohort Study. Clin Infect Dis. 2021 Sep 7;73(5):866–72.

285. Abbas M, Rossel A, de Kraker MEA, von Dach E, Marti C, Emonet S, et al. Association between treatment duration and mortality or relapse in adult patients with Staphylococcus aureus bacteraemia: a retrospective cohort study. Clin Microbiol Infect. 2020 May 1;26(5):626–31.

286. Kim CJ, Song KH, Park KH, Kim M, Choe PG, Oh MD, et al. Impact of antimicrobial treatment duration on outcome of Staphylococcus aureus bacteraemia: a cohort study. Clin Microbiol Infect. 2019 Jun 1;25(6):723–32.

287. Chong YP, Moon SM, Bang KM, Park HJ, Park SY, Kim MN, et al. Treatment Duration for Uncomplicated Staphylococcus aureus Bacteremia To Prevent Relapse: Analysis of a Prospective Observational Cohort Study. Antimicrob Agents Chemother. 2013 Mar;57(3):1150.

288. Kempley S, Kapellou O, McWilliams A, Banerjee J, McCorqodale A, Millar M. Antibiotic treatment duration and prevention of complications in neonatal Staphylococcus aureus bacteraemia. J Hosp Infect. 2015 Oct 1;91(2):129–35.

289. Havey TC, Fowler RA, Pinto R, Elligsen M, Daneman N. Duration of antibiotic therapy for critically ill patients with bloodstream infections: A retrospective cohort study. Can J Infect Dis Med Microbiol. 2013;24(3):129–37.

290. Ruiz-Ruigómez M, Fernández-Ruiz M, San-Juan R, López-Medrano F, Orellana MÁ, Corbella L, et al. Impact of duration of antibiotic therapy in central venous catheter-related bloodstream infection due to Gram-negative bacilli. J Antimicrob Chemother. 2020 Oct 1;75(10):3049–55.

291. Chotiprasitsakul D, Han JH, Cosgrove SE, Harris AD, Lautenbach E, Conley AT, et al. Comparing the Outcomes of Adults With Enterobacteriaceae Bacteremia Receiving Short-Course Versus Prolonged-Course Antibiotic Therapy in a Multicenter, Propensity Score-Matched Cohort. Clin Infect Dis. 2018 Jan 15;66(2):172–7.

292. Giannella M, Pascale R, Toschi A, Ferraro G, Graziano E, Furii F, et al. Treatment duration for Escherichia coli bloodstream infection and outcomes: retrospective single-centre study. Clin Microbiol Infect. 2018 Oct 1;24(10):1077–83.

293. Nelson AN, Justo JA, Bookstaver PB, Kohn J, Albrecht H, Al-Hasan MN. Optimal duration of antimicrobial therapy for uncomplicated Gram-negative bloodstream infections. Infection. 2017 Oct 1;45(5):613–20.

294. Rodrigues RD, Garcia RCL, Bittencourt GA, Waichel VB, Garcia ECL, Rigatto MH. Antimicrobial Therapy Duration for Bloodstream Infections Caused by Pseudomonas aeruginosa or Acinetobacter baumannii-calcoaceticus complex: A Retrospective Cohort Study. Antibiotics. 2023 Mar 8;12(3):538.

295. Sousa A, Pérez-Rodríguez MT, Suárez M, Val N, Martínez-Lamas L, Nodar A, et al. Short- versus long-course therapy in gram-negative bacilli bloodstream infections. Eur J Clin Microbiol Infect Dis . 2019 May;38(5):851-857.

296. Daneman N, Rishu AH, Xiong W, Bagshaw SM, Dodek P, Hall R, et al. Duration of Antimicrobial Treatment for Bacteremia in Canadian Critically Ill Patients. Crit Care Med. 2016 Feb 1;44(2):256–64.

297. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009 Jan;20(1):3–5.

298. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. Int J Obes. 2008;32 Suppl 3:S8–14.

299. Pereira JM, Gonçalves-Pereira J, Ribeiro O, Baptista JP, Froes F, Paiva JA. Impact of antibiotic therapy in severe community-acquired pneumonia: Data from the Infauci study. J Crit Care. 2018 Feb 1;43:183–9.

300. Langan SM, Schmidt SA, Wing K, Ehrenstein V, Nicholls SG, Filion KB, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ. 2018 Nov 14;363:k3532.

301. Danaei G, Tavakkoli M, Hernán MA. Bias in Observational Studies of Prevalent Users: Lessons for Comparative Effectiveness Research From a Meta-Analysis of Statins. Am J Epidemiol. 2012 Feb 15;175(4):250.

302. Vandenbroucke J, Pearce N. Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology? Am J Epidemiol. 2015 Oct 12;182(10):826–33.

303. Shapiro DJ, Hall M, Lipsett SC, Hersh AL, Ambroggio L, Shah SS, et al. Short- Versus Prolonged-Duration Antibiotics for Outpatient Pneumonia in Children. J Pediatr. 2021 Jul 1;234:205-211.e1.

304. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, Staa T van, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun 1;44(3):827–36.

305. Ahmed H, Farewell D, Francis NA, Paranjothy S, Butler CC. Impact of antibiotic treatment duration on outcomes in older men with suspected urinary tract infection: Retrospective cohort study. Pharmacoepidemiol Drug Saf. 2019;28(6).

306. Amoah J, Stuart EA, Cosgrove SE, Harris AD, Han JH, Lautenbach E, et al. Comparing Propensity Score Methods Versus Traditional Regression Analysis for the Evaluation of Observational Data: A Case Study Evaluating the Treatment of Gram-Negative Bloodstream Infections. Clin Infect Dis. 2020 Nov 1;71(9):E497–505.

307. Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):253–9.

308. Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. Clinical trials. 2012 Feb;9(1):48.

309. Caniglia EC, Zash R, Fennell C, Diseko M, Mayondi G, Heintz J, et al. Emulating Target Trials to Avoid Immortal Time Bias - An Application to Antibiotic Initiation and Preterm Delivery. Epidemiology. 2023 May 21;34(3).

310. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. Basic Clin Pharmacol Toxicol. 2006 Mar;98(3):237–42.

311. Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. Stat Med. 2008 Oct 15;27(23):4678–721.

312. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling. 1986;7(9–12):1393–512.

313. Robins JM, Blevins D, Ritter G, Wulfsoh M. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology. 1992;3(4):319–36.

314. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol. 2011 Apr 1;173(7):761–7.

315. Newsome SJ, Keogh RH, Daniel RM. Estimating long-term treatment effects in observational data: A comparison of the performance of different methods under real-world uncertainty. Stat Med. 2018 Jul 10;37(15):2367–90.

316. Lin PID, Daley MF, Boone-Heinonen J, Rifas-Shiman SL, Bailey LC, Forrest CB, et al. Comparing Prescribing and Dispensing Data of the PCORnet Common Data Model Within PCORnet Antibiotics and Childhood Growth Study. eGEMs. 2019 Apr 12;7(1):11.

317. Gardarsdottir H, Souverein PC, Egberts TCG, Heerdink ER. Construction of drug treatment episodes from drug-dispensing histories is influenced by the gap length. J Clin Epidemiol. 2010 Apr;63(4):422–7.

318. Petersen ML, Porter KE, Gruber S, Wang Y, Van Der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res. 2012 Feb;21(1):31–54.

319. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008 Sep 15;168(6):656–64.

320. European Medicines Agency. Data Analysis and Real World Interrogation Network (DARWIN EU). 2022. Available from: https://www.ema.europa.eu/en/about-us/how-we-work/big-data/data-analysis-real-world-interrogation-network-darwin-eu

321. Platt RW, Henry DA, Suissa S. The Canadian Network for Observational Drug Effect Studies (CNODES): Reflections on the first eight years, and a look to the future. Pharmacoepidemiol Drug Saf. 2020 Jan 1;29 Suppl 1(S1):103–7.

322. Brown JS, Maro JC, Nguyen M, Ball R. Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system. J Am Med Inform Assoc. 2020 May 1;27(5):793–7.

323. Platt RW, Platt R, Brown JS, Henry DA, Klungel OH, Suissa S. How pharmacoepidemiology networks can manage distributed analyses to improve replicability and transparency and minimize bias. Pharmacoepidemiol Drug Saf. 2019 Jan 1;29(S1):3–7.

324. Maro JC, Toh S. Go BIG and Go Global: Executing Large-Scale, Multi-Site Pharmacoepidemiologic Studies Using Real-world Data. Am J Epidemiol. 2022 Jul 23;191(8):1368-1371.

325. Carnahan RM, Kuntz JL, Wang S V., Fuller C, Gagne JJ, Leonard CE, et al. Evaluation of the US Food and Drug Administration sentinel analysis tools in confirming previously observed drug-outcome associations: The case of clindamycin and Clostridium difficile infection. Pharmacoepidemiol Drug Saf. 2018 Jul 1;27(7):731–9.

326. Daneman N, Chateau D, Dahl M, Zhang J, Fisher A, Sketris IS, et al. Fluoroquinolone use for uncomplicated urinary tract infections in women: a retrospective cohort study. Clin Microbiol Infect. 2020 May 1;26(5):613–8.

327. Ernst P, Dah M, Chateau D, Daneman N, Quail J, Sketris IS, et al. Comparative Effectiveness Of Fluoroquinolone Antibiotic Use In Uncomplicated Acute Exacerbations Of COPD: A Multi-Cohort Study. Int J Chron Obstruct Pulmon Dis. 2019;14:2939.

328. Maro JC, Platt R, Holmes JH, Strom BL, Hennessy S, Lazarus R, et al. Design of a national distributed health data network. Ann Intern Med. 2009 Sep 1;151(5):341–4.