ARK Statistical Analysis Plan

Version Number and Date: Final 2.0 5 March 2020

Supersedes version: 1.0 23 October 2018

Author	**Position**	**Signature**	**Date**
Professor A. Sarah Walker	Trial Statistician	Authorised by email given COVID	
Approved by			
Professor Tim Peto	Chief Investigator	Authorised by email given COVID	
Professor Neil French	Chair, Data Monitoring Committee	Authorised by email given COVID	
Professor Alison Holmes	Chair, Trial Steering Committee	Authorised by email given COVID	

Revision History

Version	Author	Date	Reason for Revision
Draft 0.1			Protocol version 1.0
Draft 0.2	Sarah Walker	28 August 2018	First main draft
Draft 0.3	Sarah Walker	31 August 2018	Incorporating comments from Eric Budgell, Martin Llewelyn
Draft 0.4	Sarah Walker	4 September 2018	Incorporating comments from co-applicants and co-investigators
1.0	Sarah Walker	23 October 2018	Incorporating comments from TSC (Karla Hemming) and DMC (James Lewis)
2.0	Sarah Walker	5 March 2020	Minor amendments to definition of broad-spectrum antibiotics; clarification that the PHE interpretation of the WHO-definitions of Access, Watch, Reserve antibiotics will be used, and other minor amendments suggested by DMC at their meeting in March 2019

Contents

1 Trial design

1.1 Design

ARK is a stepped wedge cluster (Trust) randomised controlled trial to evaluate a multi-faceted behavioural intervention for healthcare professionals designed to increase compliance with, and acceptability of, the 'review and revise' component of the Department of Health 'Start Smart then Focus' strategy, and specifically stopping antibiotics in those who no longer need them at 'review and revise'.

ARK contains three phases: phase I was a feasibility study in Brighton, phase II was a non-randomised pilot study in 3 Trusts, and phase III is the main stepped wedge cluster randomised trial.

The stepped wedge cluster randomised trial uses calendar-time blocked randomisation, whereby Trusts are put into blocks of 6 when they join the trial, and then randomisation is stratified by these blocks. This is feasible because all research data on which the intervention is evaluated is routine electronic health record data that is completely independent of the trial, so is available historically, collected in the same way, regardless of the date of randomisation/implementation – there is no research-specific outcome data in the trial. Trusts have no control over which block they are allocated to or to their randomisation date. This was a pragmatic decision based on the fact that Trusts in the main trial would be randomised over 18 months, and in the current NHS climate it is impossible to get NHS Trusts to agree to join a programme to reduce antibiotic use where they might not get the intervention for 18 months. All comparisons will be made before and after implementation in the same Trust (see section 5 below), since any comparison across Trusts pooling time will be irretrievably confounded by various organisational changes internally and externally.

This Statistical Analysis Plan was written in conjunction with protocol v5.0.

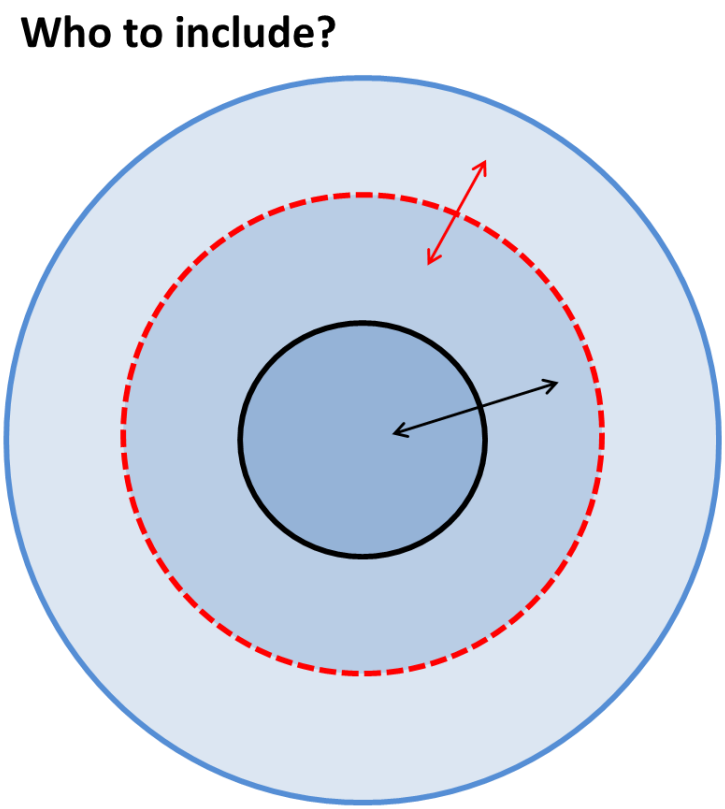
1.2 Patient population

The intended population is acute/general medical inpatients.

The research data on which the intervention will be evaluated are obtained from routine electronic health record data. All patients can request that their NHS records are not released for secondary use (for example, in Hospital Episode Statistics). Patients who have opted out of sending their hospital records to the NHS Digital will not have information submitted for intervention evaluation.

The patient population is therefore defined by either the treatment speciality code (under which the patient is treated) (Hospital Episode Statistics (HES) "tretspef") or the main speciality code (of the consultant) (HES "mainspef") of either the first or second consultant episode for an inpatient spell, where the inpatient spell has the primary admission date (HES "admidate") during the time periods in section 1.4 below.

Acute/general medical inpatients are not always identified by the use of the 300 (general medicine) speciality code, but may also be treated under several different adult specialities. See Figure below for summary of the issues.



Ideal study population: Patients who go through the acute medical unit (AMU) and hence are nominally affected by ARK intervention. However, hospital administration systems do not record whether patient passed through AMU, so we need to select our patient sample based on specialty codes. There are two approaches that could be used:

Option 1. Include patients with mainspef = 300 OR tretspef = 300 in either 1st or 2nd consultant episode. This will be a subset of those who pass through AMU (ie, will miss patients who go from AMU straight to other specialities).

Option 2. Include patients with mainspef OR tretspef in either 1st or 2nd consultant episode from any of the following (from Nuffield Trust):

300 General medicine
301 Gastroenterology
430 Gerontology/Geriatric Medicine
340 Respiratory/Thoracic Medicine
302 Endocrinology
350 Infectious Diseases
326 Acute Internal Medicine
307 Diabetic medicine
plus other smaller specialities (see text)

Note: Option 2 will lead us to include EXTRA patients who did not pass through AMU, and are therefore not managed using the decision aid (eg. cystic fibrosis patients). Including them will lead to some exposure misclassification (less than perfect specificity), but will likely include all the patients we want (high sensitivity). The extent of misclassification will depend on the share of patients in these wards that come from AMU.

In 2017 the Society for Acute Medicine and the Nuffield Trust (Martin Bardsley, personal communication) completed an exercise identifying the most commonly used speciality codes under which adult acute/general medical inpatients are admitted, and we follow their definition. Paediatric admissions have different speciality codes and are not included.

The included codes (as EITHER the treatment speciality code OR the main speciality code of EITHER the first OR second consultant episode for an inpatient spell) are

- 300 (General medicine)
- 301 (Gastroenterology)
- 430 (Gerontology/Geriatric Medicine)
- 340 (Respiratory Medicine/Thoracic Medicine)
- 302 (Endocrinology)
- 350 (Infectious Diseases)
- 326 (Acute Internal Medicine)
- 307 (Diabetic Medicine)
- 303 (Clinical Haematology)
- 320 (Cardiology)
- 410 (Rheumatology)
- 400 (Neurology)

Because of the potential for exposure misclassification (which may vary by Trust), sensitivity analyses of the primary outcomes will be conducted excluding cardiology (320), rheumatology (410), haematology (303) and neurology (400). If results provide a qualitatively different interpretation, these sensitivity analyses will be conducted for all outcomes.

The primary admission specialty (used for adjustment in secondary analyses, see section 5.4 below) will be calculated as follows, based on the prevalence of the different categories in an English-wide analysis of Hospital Episode Statistics:

- Acute/general medicine if mainspef=300 or tretspef=300 using the first episode meeting criteria for inclusion in the population of acute/general medical inpatients.
- If not defined, then repeat for gastroenterology (301), then geriatric medicine (430), then respiratory/thoracic medicine (340), then endocrinology (302), then infectious diseases (350), then acute internal medicine (326), then diabetic medicine (307), clinical haematology (303), then cardiology (320), then rheumatology (410), then neurology (400)
- For adjustment, admission specialties comprising <10% of the total will be grouped as "other".

1.3 Analysis population

Analyses will be presented for all Trusts. However, the primary analysis will include only the pilot and main trial trusts (including the pilot trusts to maximise power and because the intervention was essentially unchanged between the pilot study and the main trial and the pilot study sites did not choose their implementation dates which occurred monthly between Sept-Nov 2017). A sensitivity analysis of the primary outcomes will include the main trial sites only. If results provide a qualitatively different interpretation, these sensitivity analyses will be conducted for all outcomes.

1.4 Time periods for analysis

Data are requested for varying time periods depending on the study phase. Time periods relate to primary admission date (HES "admidate") for patients as defined in section 1.2 above. Approvals for data sharing in the protocol cover up to 24 months before the start of each phase, and therefore this data is requested wherever possible in order to maximise power. In practice, not all Trusts are able to provide data from this timepoint, in which case at least 6 months data is requested prior to the start of the study phase. All available data pre-implementation will be compared with all available data post-implementation.

- **Phase I feasibility study** (Brighton): implementation date 1 April 2017

	From	To
Ideal	12 months prior to 1 April 2017, ie from 1 April 2016	end of Phase III study
Minimum	6 months prior to 1 April 2017, ie from 1 October 2015	end of Phase III study
Maximum	24 months prior to 1 April 2017, ie from 1 April 2015	end of Phase III study

- **Phase II internal pilot**: start of Phase II pilot 1 September 2017

	From	To
Ideal	Start of Phase I data collection, ie from 1 April 2016	end of Phase III study
Minimum	6 months prior to start of Phase II, ie from 1 March 2017	end of Phase III study
Maximum	24 months prior to start of Phase II, ie from 1 September 2015	end of Phase III study

- **Phase III main trial**: start of Phase III main trial 1 February 2018

	From	To

Ideal	Start of Phase I data collection, ie from 1 April 2016	end of Phase III study
Minimum	6 months prior to start of Phase III, ie from 1 August 2017	end of Phase III study
Maximum	24 months prior to start of Phase III, ie from 1 February 2016	end of Phase III study

2 Outcome measures

2.1 Co-primary outcomes

- 30-day mortality post-admission, including deaths out of hospital (non-inferiority)
- Defined daily doses (DDDs) of antibiotics per acute/general medical admission (superiority)

The intervention will be considered successful only if it significantly reduces antibiotic usage with no evidence of increased mortality.

2.2 Secondary outcomes

- ICU admission during the current admission
- Total length-of-stay (hours)
- Emergency re-admission in the 30 days after discharge (to any speciality)
- *C. difficile* diarrhoea in the 90 days after admission
- 90-day mortality post-admission, including deaths out of hospital
- DDD per occupied bed-day (per admission is the primary outcome)
- Carbapenem DDD per admission and per bed-day
- Broad-spectrum DDD per admission and per bed-day
- IV and oral DDD per admission and per bed-day
- WHO-defined "Access", "Watch" and "Reserve" DDD per admission and per bed-day, and "Access" as a percentage of all antibiotic use

For Trusts with electronic antibiotic prescribing:

- Days on antibiotics per admission and bed-day (length-of-therapy (LOT)), antibiotic days per admission and bed-day (days-of-therapy, (DOT)), carbapenem DOT and LOT (per admission and per bed-day), broad-spectrum DOT and LOT (per admission and per bed-day) , IV/oral DOT and LOT (per admission and per bed-day), WHO-defined "Access", "Watch" and "Reserve" DOT and LOT (per admission and per bed-day)
- Antibiotic restart after discontinuation for >48h

Faecal substudy:

- Proportion of discarded faecal samples from medical inpatients from which extended spectrum beta-lactamase (ESBL)-carrying *Enterobacteriaceae* can be isolated (this SAP does not cover this analysis)

2.3 Tertiary outcomes

- Resource-utilisation and costs (this SAP does not cover this analysis)
- Proportion of locally identified and prespecified essential individuals who drive prescribing decisions for acute/general medical inpatients at the Trust who complete the online training (part of the intervention) by 12 weeks from randomised implementation date
- Proportion of regularly audited antibiotics prescriptions (part of the intervention) which document the ARK classification criteria through 12 weeks from randomised implementation date

- Proportion of regularly audited antibiotics prescriptions (part of the intervention) which are stopped at 'review and revise' through 12 weeks from randomised implementation date

3 Analysis models

3.1 Binary outcomes

The primary analysis method for the binary outcomes above will be logistic regression as specified in the protocol. This is because for mortality, data will be considered to be completely ascertained (patients not recorded as dying in the national system checked by each Trust will be assumed to be alive). Further for all binary outcomes, event rates are relatively low (5-10%), so there is little difference between analysing a cause-specific hazard or a competing risks sub-hazard. Analysing the outcome using logistic regression also facilities visualisation of the estimated pre- and post-implementation trends, and step-change at implementation, against the observed percentages for each outcome per Trust. For interim analyses, admissions in the 30 days before data submission will not be included, since their 30-day outcomes cannot be determined completely. This will also be done at the final analysis for Trusts which are not able to provide complete data to the common enddate.

Sensitivity analyses will consider these outcomes as time-to-event, counting time

- from admission for mortality, length-of-stay to discharge alive, length-of-stay to medically fit for discharge, ICU admission and *C. difficile* diarrhoea,
- from discharge for re-admission (restricted to patients discharged alive)
- from last dose of antibiotics (where this occurs for >48h, see below) for antibiotic restart after discontinuation

Inpatient death will be considered as a competing risk for

- length-of-stay (ie, analysis of time to discharge alive or medically fit for discharge, in addition to crude analysis of length-of-stay as a numeric value regardless of outcome below)

Death out of hospital will be considered as a competing risk for

- emergency re-admission within 30 days post-discharge (as only deaths within 90 days of admission are requested from Trusts, this analysis will also be restricted to patients who were discharged alive within 60 days of admission in whom mortality 30 days post-discharge can therefore be ascertained)

Inpatient death and discharge alive will be considered as a competing risk for

- ICU admission

Death (in or out of hospital) will be considered as a competing risk for

- *C. difficile* diarrhoea within 90 days of admission

In these sensitivity analyses, all other patients not experiencing the event of interest or a competing event will be censored at the relevant time period

- 30 or 90 days for mortality
- 30 days for re-admission
- 90 days for *C. difficile* diarrhoea

This implicitly assumes that vital status is known at these timepoints – as hospital records are routinely updated from national death reporting this is a reasonable assumption.

For length-of-stay and antibiotic restart after discontinuation, time will be counted in hours and mins from last dose. Where only date and not time of restart is known, it will be assumed to have occurred at midday (midpoint imputation).

For all other time-to-event outcomes, time will be counted in days, setting time to 0.1 for any patients experiencing the event (or the competing event) on day 1 (eg some patients may die on the day of admission).

3.2 Other outcomes

Negative binomial regression will be used to model antibiotic defined daily doses (DDDs) incorporating over-dispersion and using different offsets as defined in the outcome section above.

Other antibiotic outcomes are planned to be analysed using ordinal logistic regression; however, if, for example, data are approximately normally distributed, either before or after transformation, this regression model may be used instead.

4 Derivation of data to be analysed

4.1 Population

The population is defined by speciality and treatment codes as described in section 1.2 above. Any individual aged <16 years at admission will be dropped from analyses, even if these codes have been used (incorrectly), since the population the ARK intervention applies to is adult acute/general medical inpatients. Any admission without one of the codes in section 1.2 above, but with an A&E code instead, will also not be included in the analysis.

4.2 Emergency re-admissions

These will be defined by '2' as the first character in admission method.

4.3 Truncation and non-linearity assessment for continuous variables

The distribution of all continuous measurements will be visually inspected, including all values from all Trusts. Values larger than the 99th percentile will be truncated to the 99th percentile (pre-defined based on skew observed in previous analyses).

Non-linear relationships between continuous explanatory variables and outcomes will be assessed using natural cubic splines, fixing knots at pre-defined percentiles of the distribution across all Trusts (10th, 90th, then equally distributed within these exterior knots), choosing the number of knots based on minimising the AIC or 5 knots, whichever is the smaller.

For length-of-stay as an outcome regardless of inpatient mortality vs discharge alive, the BoxCox transformation such that the transformed variable is approximately normal will be used as the outcome in regression models.

4.4 Missing data

Other than antibiotics, outcome data and variables for adjustment (see section 5.4 below) are based on routine electronic health records linked by NHS numbers within participating Trusts. As these records are used for patient management, data in the variables being requested should not be missing. As above, for mortality, data will be considered to be completely ascertained (patients not recorded as dying in the national system checked by each Trust will be assumed to be alive); similarly for other outcomes, as follow-up time is relatively short, observed data will be assumed to represent all events within the prescribed follow-up time (30-90 days).

For adjustment variables, first if demographics are missing for only some records for a patient, these will be carried forwards and backwards to any other records with missing values. Afterwards, any records with missing values for any of the key adjustment variables (section 5.4 item 3) will be enumerated. If these comprise <0.1% of records from a Trust, they will be assumed to represent incorrect records in the underlying data sources and checked with local Trusts, but will be dropped from all analyses (adjusted and unadjusted). Where records with missing data comprise >0.1% of records from a Trust, this will be queried since it is not anticipated that this should ever occur with the requested data items, excepting those described explicitly in the next paragraph.

The exceptions are Index of Multiple Deprivation (IMD) and ethnicity which may be missing either because the patient is not resident in the UK (no postcode) or for other reasons, or because ethnicity is recorded as "unknown", respectively. Any known ethnicity for a patient will first be replaced with known ethnicity for a patient, and similarly any missing IMD will be replaced with the IMD from the closest admission. Otherwise analyses adjusting for IMD and ethnicity will use complete cases, given potential for bias using missing indicator methods and the challenges of imputation in these complex large datasets.

4.5 Antibiotics

Antibiotic usage will be transformed into defined-daily-doses (DDD) using standard WHO formulae (www.whocc.no). They are designed to transform mg of different antibiotics into standard units reflecting a typical dose. These standardised doses were defined historically, often on the basis of less severe infections. For example, the DDD for amoxicillin is 1g (1 DDD), although 3g (3 DDD) is the standard dose used for serious infections in hospitals. WHO are in the process of updateing the DDD conversion factors to incorporate modern dosing regimes. We intend to use the updated WHO DDDs for the final analysis, but until these are available will use conversion factors (derived from consultations with clinical microbiologists) for typical daily doses used in hospitals for all calculations involving DDDs (Appendix I).

DDDs have recognised limitations, the most important of which is that reducing prescribing of broad-spectrum antibiotics often involves moving to a combination multiple agents of narrower spectrum. A single DDD of a broad-spectrum antibiotic may therefore be replaced with multiple DDD of narrower spectrum antibiotics, but this is probably better for future antimicrobial resistance. For this reason, where Trusts are able to provide individual level antibiotic prescribing data, we will also analyse

- Length of therapy (LOT), defined as the time (days/hours/mins) between the last and first administration of antibiotic treatment (lumping all antibiotics together)
- Days of therapy (DOT), defined as the sum of the time (days/hours/mins) between the last and first administration of each separate antibiotic

Note that DOT suffers from some of the same limitations as DDDs in that moving from one broad-spectrum antibiotic to two or more narrow-spectrum agents leads to an increase in DOT, but it does calculate per-patient daily exposure.

Admission (rather than bed-day) is the denominator for antibiotic usage (similarly to the DH Antimicrobial Prescribing Quality Measures) because bed-day can be strongly influenced by non-medical reasons for not discharging inpatients. However, this provides a challenge since it is not straightforward to identify these admissions from the electronic inpatient admission records (see section 1.2 above). Where hospitals provide data on bulk prescriptions only (e.g. using the JAC or Rx-info systems), they are able to provide both total DDDs and DDDs per admission to the relevant hospital areas. The primary analysis will use this denominator wherever available. Where this data is not available, we will use estimates of the relevant admissions from the population defined in section 1.2. Analysis per bed-day will use overnight stays as calculated from the population defined in section 1.2.

For Trusts that are only able to provide bulk (DDD) data by area, ie do not have individual electronic antibiotic prescribing, we will consider two further adjustments.

The first is to analyse only antibiotics used as first or second-line for key medical indications. This will enable us to exclude patients being managed for surgical infections, prophylaxis etc in clinical areas which have implemented the ARK intervention. These antibiotics will be locally defined according to each Trust's antimicrobial prescribing guidelines.

The second is an adjustment for how often antibiotics are used alone or in combination, since this is not possible to determine from DDD, yet it is LOT that ARK is actually trying to change. If a drug is always used alone the relationship between DDD and LOT would be 1. If it is used alone half the time and the other half of the time in a combination of two drugs this would be 0.5 + 0.25 = 0.75. If used 50% of the time as a single drug, 25% as 1 of 2 drugs and 25% as 1 of 3 drugs this would be 0.5 + 0.125 + 0.075 = 0.7 etc. This will require validation either from existing point prevalence data, national (e.g. (Klein, Van Boeckel et al. 2018)) or local (e.g. as contributed to European Center for Disease Control) or from local antimicrobial prescribing guidelines within the trial.

Broad-spectrum is defined as co-amoxiclav; piperacillin/tazobactram; second (eg cefuroxime), third (eg ceftriaxone ceftazidime) or fourth (eg cefepime) generation cephalosporins or cephalosporin-beta-lactamase inhibitor combinations (eg ceftazidime-avibactam, ceftolozane-tazobactam); carbapenems; quinolones; azithromycin; tigecycline; aztreonam; telithromycin;and polymyxins. Cefaclor is not included as a second generation cephalosporin because it is administered orally and is not well absorbed.

Public Health England (PHE)interpretation of the WHO definitions of "Access", "Watch" and "Reserve" antibiotics will be used (Budd, Sharland et al. 2018, NHS Improvement 2018). This is because WHO definitions follow the 2017 Essential Medicines List and define antibiotics that individuals should reliably able to access, whereas the (minor) PHE amendments reflect what antibiotics hospitals should be prioritising for human use, and also reflect the fact that this specific indication is not known for bulk data.

4.6 Per-protocol population

All analyses will be intention-to-treat, using the randomised date of implementation as the time of implementation. Sites which withdraw and never implement the intervention will be excluded, and therefore the analysis will be formally described as modified intention-to-treat (mITT). This is because no change in either antibiotic use or clinical outcomes would be expected in sites that never implemented the intervention, and as the clinical outcome is a non-inferiority comparison, it is more important to replace these sites, than use resources collecting data which would show no effect. Numbers withdrawn will be reported.

It is not possible to specify an individual-level per-protocol population since the precise method of management of each individual inpatient is unknown. All results will be interpreted in the light of the changes in antibiotic use achieved, both overall and by Trust.

At the Trust-level, a per-protocol population will be defined at the Trust-level by >50% of locally identified and prespecified essential individuals who drive prescribing decisions for acute/general medical inpatients having completed the online training.

5 Main Analysis

All analysis will be included in the final report, but only analysis in bold below will be included in the DMC report.

5.1 Recruitment and baseline characteristics

- **For each Trust, date randomised, randomised date of implementation, current status (randomised but not yet implemented, during 12 week implementation phase, in maintenance/sustainability phase), availability of electronic research data, months from initial request to receiving data for each interim report**
- **For each Trust, region, type of trust, total number of admissions and beddays in 2016 (to reflect approximate total size), estimated number of acute beds (from KHO3 statistics https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-overnight/, adjusted for whether ARK was implemented at one of multiple acute sites), functional role of local Champion, prescribing system at implementation**
- **Withdrawals and reasons**

5.2 Follow-up

- **Figures showing periods for which electronic data is available from each Trust on admissions and on prescribing**

5.3 Implementation

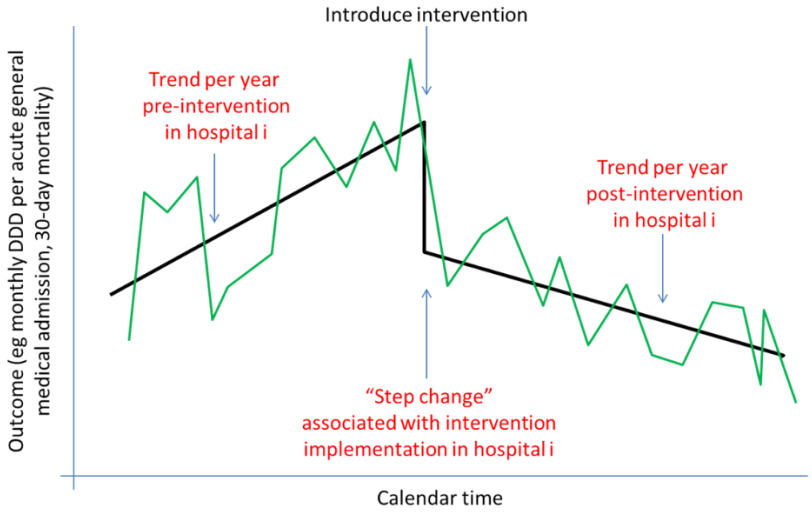
- **For each Trust, number of pre-specified essential people, number (%) completing training by 12 weeks from implementation date, number (%) completing training after 12 weeks from implementation date, total numbers completing training by 12 weeks from implementation date, mean numbers completing training by 12 weeks from implementation per 100 acute beds, total numbers completing training after 12 weeks from implementation date**
- **Proportion of regularly audited antibiotics prescriptions (part of the intervention) which document the ARK classification criteria through 12 weeks from randomised implementation date**
- **Proportion of regularly audited antibiotics prescriptions (part of the intervention) which are stopped at 'review and revise' through 12 weeks from randomised implementation date**

5.4 Binary outcome analyses

Analyses of each of the binary outcomes listed below will follow the same approach.

For analyses comparing periods before and after introduction of the intervention, we hypothesise that the intervention could produce either a step-change in each outcome, or ongoing changes over time (i.e. an interaction between intervention vs control and calendar time), or both. Analysis will therefore use an interrupted time series (ITS) analysis which not only estimates whether the intervention has a direct immediate impact, but also whether it has any impact on year-on-year trends after implementation (and compared with year-on-year trends pre-implementation) (Figure). Other reasons for using the ITS approach within Trust, and then pooling Trust-level estimates, are that many other things are likely to be changing differently in different Trusts, and within Trusts, over calendar time; these will affect within Trust comparisons less than between Trust comparisons. Further rates may not be constant before or after the intervention. Further the calendar time blocked randomization makes the assumptions underlying the standard vertical comparisons (between Trusts already randomised to intervention vs not at each step) more questionable.

The date of introduction of the intervention will be the date that each Trust was randomised to implement the intervention (non-randomised planned date of implementation feasibility/pilot study). The protocol v1.0 specified that a sensitivity analysis for the main trial would consider the date of introduction as the date on which the original invitation to complete online training was sent to the locally identified and prespecified essential individuals who drive prescribing decisions for acute/general medical inpatients at the Trust(s). However, in practice these individuals comprise both those in the core team preparing for implementation and key members of staff on the ground, so has been a range of dates (in some cases very wide) rather than a single date, and therefore this will not be done (removed from protocol v5.0). A sensitivity analysis will use the actual date that implementation occurred (delayed in some sites due to delays obtaining R&D approval or for other logistic reasons relating to structural factors).



(Then use meta-analytic techniques to combine hospital-specific estimates weighted by the precision with which they can be estimated)

Specifically therefore

1. **Logistic regression models will be fit to each outcome separately within each Trust. The unit of analysis will be admission. Patient will be used as a clustering variable for robust variance adjustment (but not otherwise adjusted for).**
    - **a.** This approach is similar to a multilevel model, but has more flexibility in modelling individual Trusts and particularly controlling for confounding (case-mix) in individual Trusts.
2. **All analyses will include a trend over calendar day pre-implementation, an immediate step-change at implementation, and a trend over calendar day post-implementation. The date of implementation will be that randomised (modified intention-to-treat, as defined above).**
    - a. Sensitivity analyses for primary outcomes will assume a linear trend over the three months immediately following implementation rather than a step-change. If results provide a qualitatively different interpretation, these sensitivity analyses will be conducted for all outcomes.
    - b. The primary comparisons will be the 'step-change' associated with intervention implementation and change in calendar trends post-vs pre-implementation: secondary analyses will consider absolute magnitude of calendar trends post-implementation. The primary comparison will be a 2 degree of freedom test, jointly testing that the 'step-change' is zero and that there is no change in calendar trends post-vs pre-implementation. This joint test will be fitted on model parameters and explicitly account for correlation between the two estimands. It will be fitted at the individual Trust level, and also overall using multivariate meta-analysis (as described

below). At an individual Trust level, we will also estimate the likelihood ratio test comparing a single trend model with the model including both a step-change and a change in trend post vs pre-implementation.

3. **Analyses will be conducted unadjusted for other variables (primary analysis) and adjusted for the following admission-level covariates regardless of statistical significance (with non-linear relationships between covariates and outcomes assessed and incorporated using natural cubic splines).**
    a. Sex (female, male; setting intersex to female)
    b. Age at admission, truncated at the 99th percentile
    c. Immunosuppression (yes, no), defined as any of HIV, severe liver disease, or metastatic cancer in primary or secondary diagnosis codes
    d. Charlson Co-Morbidity Index (defined using secondary diagnosis codes associated with the first episode meeting criteria for inclusion in the population of acute/general medical inpatients as defined above), truncated at the 99th percentile
    e. Interaction between age at admission and Charlson Co-Morbidity Index, as in previous analysis of weekend mortality (Walker, Mason et al. 2017)
    f. Admission method (A&E, Elective and other non-emergency, Emergency via GP or other source)
    g. Admission source (usual/other place of residence, NHS general ward/other case provider)
    h. Admission specialty (e.g. general medicine, gastroenterology, haematology, cardiology, geriatric medicine, respiratory medicine, other)
    i. Patient classification (ordinary admission, day case admission, regular day attender) (note: this was inadvertently missed out of data specification documents up to 1.5 but will be added from 1.6 onwards)
    j. Admission day of the week (weekend, weekday)
    k. Admission time of day (hour/minute) and day of year (both modelled using a sin() + cos() function (2df) to ensure a smooth transition in risk from minute to minute and day to day respectively)
    l. Interaction between admission day of the week and time of day, as in previous analysis of weekend mortality (Walker, Mason et al. 2017)
    m. Number of admissions in previous year (excluding as day case), truncated at the 99th percentile
    n. Ever had a complex admission (>1 consultant episode) in the previous year (excluding as day case), truncated at the 99th percentile
    o. Analyses of re-admission will also adjust for the length of the admission the discharge is being counted from.

The major concern about unadjusted analyses is the potential for variation in case-mix, particularly over a calendar year, to affect estimates of intervention effect. These variables for adjustment have been chosen based on their significant effects on 30-day mortality in previous analyses of emergency admissions from routinely collected data (Walker, Mason et al. 2017). Effect estimates for each will be compared with those from previous analyses to assess potential for over-fitting (e.g. opposite effects in these adjusted analyses compared with previous much larger studies) and any covariates with evidence of co-linearity will be removed from models.

Note that these analyses do not directly adjust for the randomisation stratification factor (calendar time via randomisation blocks), because this factor is fixed for each Trust, and randomisation block is completely confounded with the randomised implementation date. See section 5.7 below for subgroup analyses including randomisation block.

4. Other possible variables which could be adjusted for are Index of Multiple Deprivation (IMD) score, ethnicity, Clinical Classifications Software (CCS) group of the primary diagnosis in the first episode meeting criteria for inclusion in the population of acute/general medical inpatients as defined above, and laboratory test results.
    a. IMD score and ethnicity are not available for Trusts from Northern Ireland, so cannot be part of the primary analysis. A secondary analysis will additionally adjust for these two variables, categorising ethnic category (white, asian, black, other, unknown) and IMD score (no truncation as limited to 0-100 by definition). These adjusted estimates can only be pooled over Trusts excluding Northern Ireland.
    b. CCS group is only assigned at the end of the admission, and therefore reflects the final diagnosis, not necessarily the one for which antibiotics were given. Analyses also adjusting for CCS group in addition to the variables above will therefore be conducted as secondary analyses.
    c. Not all Trusts are able to provide laboratory test results. A sensitivity analysis will restrict to Trusts providing laboratory test results and to patients with CRP, neutrophils, albumin and haemoglobin within a ±48h window of admission time (taking the closest value to admission within this window)
5. **Estimates of pre-implementation trend, implementation step-change, post-implementation trend, and change in trend post- vs pre-implementation, will then be combined across Trusts using meta-analytic techniques, weighted by their standard errors, using random effects models, and presented as forest plots. Heterogeneity between Trusts will be summarised using I2 statistics. The intervention estimates (implementation step-change, and change in trend post- vs pre-implementation) will be plotted against each other (with their 95% CI) and analysed using multivariate meta-analysis techniques to estimate overall changes in the two different components of intervention effectiveness. We will also use meta-analysis techniques to combine the combined estimate of intervention effectiveness from the 2df test described above. For tests of superiority outcomes, this has a straightforward interpretation. For non-inferiority outcomes we will interpret the results in comparison with a non-inferiority margin as pre-defined on the log(RR) scale for the step-change in mortality.**
6. **For each Trust, monthly proportions achieving each outcome will also be plotted.**

Binary outcomes

- **30-day mortality post-admission, including deaths out of hospital**
- **90-day mortality post-admission, including deaths out of hospital**
- **Emergency re-admission in the 30 days after discharge (to any speciality)**
- ***C. difficile* diarrhoea in the 90 days after admission**
- ICU admission during the current admission
- Antibiotic restart after discontinuation for >48h

Sensitivity analyses will treat these outcomes as time-to-event using either Cox proportional hazards regression or equivalent methods for the competing risks sub-hazard (Fine and Gray 1999) (see section 3.1 for competing risks/censoring). Total length-of-stay will also be considered as a time-to-event outcome treating death as a competing risk (hours) in these analyses.

5.5 Antibiotic prescribing

Negative binomial regression will be used following the strategy described in section 5.4 above to estimate pre-implementation, at implementation, and post-implementation changes in antibiotic DDDs incorporating over-dispersion.

DOT and LOT are only available for Trusts with electronic prescribing, and their distribution is currently unknown in this population. At present, it is planned to analyse them using ordinal logistic

regression, but if for example data are approximately normally distributed, either before or after transformation, this regression model may be used instead.

Antibiotic outcomes (see section 4.4 for definitions) are as follows

- **DDD of antibiotics per acute/general medical admission**
- DDD per bed-day, carbapenem DDD (per admission and per bedday)
- Days on antibiotics per admission and bed-day (LOT)
- Antibiotic days per admission and bed-day (DOT)
- Carbapenem DDD, DOT and LOT (per admission and per bedday)
- Broad-spectrum DDD per admission and per bed-day
- IV and oral DDD per admission and per bed-day
- WHO-defined "Access", "Watch" and "Reserve" DDD per admission and per bed-day, and "Access" as a percentage of all antibiotic use

5.6 Continuous outcomes

Normal linear regression of Boxcox-transformed length-of-stay will be used following the strategy described in section 5.4 above to estimate pre-implementation, at implementation, and post-implementation changes in total length-of-stay in all patients regardless of outcome (discharge alive vs in-hospital death) (hours). A sensitivity analysis will include stay only up to the point a patient was declared medically fit for discharge. A second sensitivity analysis will use quantile regression of the 90th percentile to assess effects at the extremes of the distribution.

5.7 Subgroup analyses

For the primary endpoint, subgroup analyses will be conducted by investigating heterogeneity in the Trust-level intervention effect estimates across the following factors. These analyses will use meta-regression techniques.

Subgroup analyses will be conducted for

- Randomisation block (categorical stratification factor)
- Calendar period of randomised implementation date (reflecting different NHS pressures) (Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec)
- Trust type (small, medium, large, teaching)
- UK region
- Functional role of champion (acute medicine, microbiology/infectious diseases, pharmacist)
- Paper vs electronic prescribing systems
- Percentage of essential people completing training by 12 weeks
- Total number completing training by 12 weeks per 100 acute beds

6 Other Analyses

Analysis described above will be conducted on routine electronic health record data submitted from participating Trusts. However, for English Trusts data on both antibiotic prescribing and inpatient admissions (not *C. difficile*) is available through the English Surveillance Programme on Antibiotic Utilisation and Resistance (ESPAUR), through purchasing data provided to Public Health England (PHE) from a commercial company, and through Hospital Episode Statistics (HES). This would enable us to conduct analyses of mortality, re-admission, length-of-stay and DDDs for all Trusts in England, according to whether or not they had implemented ARK within the trial, to compare pre-implementation trends in all Trusts, and step-change and post-implementation trends in ARK hospitals vs non-ARK hospitals for these outcomes.

Secondary analyses of the main trial will consider how much of any effect of the intervention could be mediated through different aspects of the implementation of review and revise, including training completion rates and acceptability, evidence of implementation from audit records and success of implementation in each hospital as categorised based on mixed methods evaluation and documentary evidence.

REFERENCES

Budd, E., M. Sharland, k. Hand, P. Howard, P. Wilson, M. H. Wilcox and S. Hopkins (2018). Adaptation of the WHO essential medicines list for national antibiotic stewardship policy (P1706). 28th European Conference for Clinical Microbiology and Infectious Diseases. Madrid, 21-24 April.

Fine, J. P. and R. J. Gray (1999). "A proportional hazards model for the subdistribution of a competing risk." Journal of the American Statistical Association **94**(446): 496-509.

Klein, E. Y., T. P. Van Boeckel, E. M. Martinez, S. Pant, S. Gandra, S. A. Levin, H. Goossens and R. Laxminarayan (2018). "Global increase and geographic convergence in antibiotic consumption between 2000 and 2015." Proceedings of the National Academy of Sciences of the United States of America **115**(15): E3463-E3470.

NHS Improvement (2018). Reducing the impact of serious infections CQUIN 2018/19 FAQs, parts 2c and 2d (https://improvement.nhs.uk/documents/2689/7_CQUIN_FAQs_1819_FINAL.pdf).

Walker, A. S., A. Mason, T. P. Quan, N. J. Fawcett, P. Watkinson, M. Llewelyn, N. Stoesser, J. Finney, J. Davies, D. H. Wyllie, D. W. Crook and T. E. A. Peto (2017). "Mortality risks associated with emergency admissions during weekends and public holidays: an analysis of electronic health records." Lancet **390**(10089): 62-72.

Appendix I Conversion factors from WHO DDD to typical daily doses used in hospitals for serious infections

atccode	Antibiotic agent	WHO DDD	Typical Daily dose	Adjustment
J01GB06	Amikacin IV	1	1	1.00
J01CA04	Amoxicillin PO	1	3	0.33
J01CA04	Amoxicillin IV	1	3	0.33
J01CA01	Ampicillin PO	2	2	1.00
J01CA01	Ampicillin IV	2	2	1.00
J01FA10	Azithromycin PO	0.3	0.5	0.60
J01FA10	Azithromycin IV	0.5	0.5	1.00
J01CE01	Benzyl-penicillin	3.6	4.8	0.75
J01AA02	Doxicycline PO	0.1	0.2	0.50
J01DB01	Cephalexin PO	2	2	1.00
J01DD04	Ceftriaxone IV	2	2	1.00
J01DC02	Cefuroxime IV	1.5	1.5	1.00
J01DD02	Ceftazidime IV	4	6	0.67
J01MA02	Ciprofloxacin PO	1	1	1.00
J01MA02	Ciprofloxacin IV	0.5	1.2	0.42
J01FA09	Clarythromyin PO	0.5	1	0.50
J01FA09	Clarythromyin IV	1	1	1.00
J01FF01	Clindamycin PO	1.2	1.8	0.67
J01FF01	Clindamycin IV	1.8	2.4	0.75
J01CR02	Co-amoxiclav PO	1	1.5	0.67
J01CR02	Co-amoxiclav IV	3	3	1.00
J01FA01	Erythromycin PO	1	1	1.00
J01FA01	Erythromycin IV	1	1	1.00
J01CF05	Flucloxacillin PO	2	2	1.00
J01CF05	Flucloxacillin IV	2	4	0.50
J01GB03	Gentamicin	0.24	0.35	0.69
J01MA12	Levofloxacin PO	0.5	1	0.50
J01MA12	Levofloxacin IV	0.5	1	0.50
J01XD01	Metronidazole PO	2	1.2	1.67
J01XD01	Metronidazole IV	1.5	1.5	1.00
J01DH02	Meropenem IV	2	3	0.67
J01MA14	Moxifloxacin PO	0.4	0.4	1.00
J01MA14	Moxifloxacin IV	0.4	0.4	1.00
J01XE01	Nitrofuantoin	0.2	0.2	1.00
J01CE02	Phenoxymethylpenicillin	2	2	1.00
J01CR05	Tazocin IV	14	12	1.17
J01CA17	Temocillin IV	2	4	0.50
J01XA02	Teicoplanin IV	0.4	0.4	1.00
J01CR03	Timentin IV	15	12	1.25
J01EA01	Trimethoprim PO	0.4	0.4	1.00
J01XA01	Vancomycin IV	2	2	1.00

Note: if antibiotic not listed, then the conversion factor is 1.00