Optimising design of research to evaluate antibiotic stewardship interventions; consensus recommendations of a multinational working group

Article (Published Version)

Clinical Microbiology and Infection 26 (2020) 41–50

Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Consensus statement

Optimizing design of research to evaluate antibiotic stewardship interventions: consensus recommendations of a multinational working group

V.A. Schweitzer [1], C.H. van Werkhoven [1], J. Rodríguez Baño [2], J. Bielicki [3], S. Harbarth [4], M. Hulscher [5], B. Huttner [4], J. Islam [6], P. Little [7], C. Pulcini [8], A. Savoldi [9, 10], E. Tacconelli [9, 10], J.-F. Timsit [11, 12], M. van Smeden [13], M. Wolkewitz [14], M.J.M. Bonten [15], A.S. Walker [16, 17], M.J. Llewelyn [7, *], on behalf of Joint Programming Initiative on Antimicrobial Resistance (JPIAMR) Working Group on Design of Antimicrobial Stewardship Evaluations

1) *Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, the Netherlands*
2) *Unit of Infectious Diseases, Clinical Microbiology and Preventive Medicine, Department of Medicine, Hospital Universitario Virgen Macarena, Universidad de Sevilla and Biomedicine Institute of Sevilla (IBiS), Seville, Spain*
3) *Paediatric Infectious Disease Research Group, St George's University of London, London, UK*
4) *Department of Infectious Diseases and Infection Control, Hôpitaux Universitaires de Genève, Geneva, Switzerland*
5) *Scientific Centre for Quality of Healthcare, Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, the Netherlands*
6) *Department of Global Health and Infection, Brighton and Sussex Medical School, Falmer, UK*
7) *Department of Primary Care Research, University of Southampton, Southampton, UK*
8) *Infectious Diseases Department, Université de Lorraine, CHRU-Nancy, APEMAC, Université de Lorraine, Nancy, France*
9) *Infectious Diseases, Department of Diagnostic and Public Health, Verona, Italy*
10) *University Hospital, Internal Medicine, Tuebingen University, Germany*
11) *Medical and Infectious Diseases ICU, Bichat University Hospital, AP-HP, Paris, France*
12) *UMR 1137, Infection Antimicrobials Modelling Evolution, Paris Diderot University, Paris, France*
13) *Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, the Netherlands*
14) *Institute for Medical Biometry and Statistics, University of Freiburg, Freiburg, Germany*
15) *Department of Medical Microbiology, University Medical Centre Utrecht, Utrecht, the Netherlands*
16) *MRC Clinical Trials Unit, University College London, London, UK*
17) *Nuffield Department of Medicine, University of Oxford, Oxford, UK*

A R T I C L E I N F O

Article history:
Received 19 June 2019
Received in revised form
20 August 2019
Accepted 22 August 2019
Available online 4 September 2019

Editor: L Leibovici

Keywords:
Antimicrobial resistance
Antimicrobial stewardship
Appropriate antimicrobial use
Methodology
Quality
Research design

A B S T R A C T

Background: Antimicrobial stewardship interventions and programmes aim to ensure effective treatment while minimizing antimicrobial-associated harms including resistance. Practice in this vital area is undermined by the poor quality of research addressing both what specific antimicrobial use interventions are effective and how antimicrobial use improvement strategies can be implemented into practice. In 2016 we established a working party to identify the key design features that limit translation of existing research into practice and then to make recommendations for how future studies in this field should be optimally designed. The first part of this work has been published as a systematic review. Here we present the working group's final recommendations.
Methods: An international working group for design of antimicrobial stewardship intervention evaluations was convened in response to the fourth call for leading expert network proposals by the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR). The group comprised clinical and academic specialists in antimicrobial stewardship and clinical trial design from six European countries. Group members completed a structured questionnaire to establish the scope of work and key issues to develop ahead of a first face-to-face meeting that (a) identified the need for a comprehensive systematic review of study designs in the literature and (b) prioritized key areas where research design considerations restrict translation of findings into practice. The working group's initial outputs were reviewed by

* Corresponding author. M. J. Llewelyn, Department of Global Health and Infection, Brighton and Sussex Medical School, University of Sussex, Falmer BN1 9PS, United Kingdom.
E-mail address: m.j.llewelyn@bsms.ac.uk (M.J. Llewelyn).

https://doi.org/10.1016/j.cmi.2019.08.017

independent advisors and additional expertise was sought in specific clinical areas. At a second face-to-face meeting the working group developed a theoretical framework and specific recommendations to support optimal study design. These were finalized by the working group co-ordinators and agreed by all working group members.
Results: We propose a theoretical framework in which consideration of the intervention rationale the intervention setting, intervention features and the intervention aims inform selection and prioritization of outcome measures, whether the research sets out to determine superiority or non-inferiority of the intervention measured by its primary outcome(s), the most appropriate study design (e.g. experimental or quasi- experimental) and the detailed design features. We make 18 specific recommendation in three domains: outcomes, objectives and study design.
Conclusions: Researchers, funders and practitioners will be able to draw on our recommendations to most efficiently evaluate antimicrobial stewardship interventions. **V.A. Schweitzer, Clin Microbiol Infect 2020;26:41**

Background and context

Antimicrobial resistance is a rapidly growing and major threat to human health [1]. Overuse of antimicrobials drives resistance at the individual [2] and population level [3]). The term antimicrobial stewardship refers to interventions and programmes that aim to optimize antimicrobial use, achieving effective treatment while minimizing antimicrobial-associated harms including resistance [4].

Despite the large and exponentially increasing number of studies published since the term antimicrobial stewardship was coined [5–7], evidence remains remarkably weak both for *what* specific antimicrobial use interventions are effective (in terms of mortality, length of stay, adverse events, resistance rates) and *how* antimicrobial use improvement strategies can be implemented to deliver the desired antimicrobial use in daily clinical practice [8]. A 2016 systematic review of evidence supporting key antimicrobial use interventions (e.g. prescribing according to guidelines, de-escalation of therapy, intravenous to oral switching) identified predominantly low-quality and highly heterogenous supporting evidence [9]. The evidence around improvement strategies is similarly weak, dominated by uncontrolled before–after studies and inadequately performed interrupted time series analyses, mostly performed within single hospitals [10].

We recently reported a broad systematic review of antimicrobial stewardship intervention studies which highlighted key frequent design weaknesses [7]. Studies which aim to assess effectiveness of antimicrobial use interventions are typically underpowered and fail to provide evidence on safety or even do not report clinical outcome data at all. Improvement strategy studies are often multifaceted with inadequate process evaluation to allow mediators of impact to be assessed [11]. Generally, the field of antimicrobial stewardship research is dominated by single-centre observational and quasi-experimental studies which fail to deal optimally with risks of different forms of bias and that lack external validity [7,8].

Building on this work we established a working group of investigators in this field that used a consensus-building iterative process over 12 months to build a conceptual framework and develop specific recommendations for the design of stewardship evaluations, which were then reviewed and amended by an expert advisory committee. This guidance is the final result of that process and aims to support investigators when making key design decisions and funders assessing proposals for studies of antimicrobial stewardship interventions and hopefully enhances the quality and impact of research in this crucial area.

Methods

An international working group for design of antimicrobial stewardship intervention evaluations was convened in response to the fourth call for leading expert network proposals by the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR). The study sponsor was the UK Medical Research Council. The working group co-ordinators (M.J.M.B., M.J.L.) and co-applicants (V.A.S., A.S.W. and C.H.v.W.) purposively selected an additional eight leading clinical and academic specialists in antimicrobial stewardship and clinical trial design from six European countries (France, Germany, Italy, The Netherlands, Spain, Switzerland and the UK) to contribute. Selection secured input from the diversity of professionals involved in antimicrobial stewardship practice (infection, internal medicine, intensive care medicine) and research (trial design, statistics and qualitative research) disciplines. Consensus was sought through a nominal group process. Group members completed a structured questionnaire to establish the scope of work, key study designs used in antimicrobial stewardship, identify the major limitations on different study designs and key issues to develop ahead of a first face-to-face meeting. The group met in March 2017 and anonymized responses were used for feedback to the whole group and relevant literature was presented (V.A.S., C.H.v.W., M.J.L.). This identified the need for a comprehensive systematic review of study designs in the literature. In parallel, in moderated small group work, candidate solutions were proposed to address the limitations identified, and in a final round-table moderated discussion the group prioritized four key areas where research design considerations restrict translation of findings into practice: features of the intervention under evaluation; appropriate selection of outcome measures; demonstration of superiority/non-inferiority of the intervention according to the outcome measures selected; and strategies to minimize bias within experimental and quasi-experimental study designs. The working group's initial outputs were reviewed by two independent advisory experts, both senior, clinically active antimicrobial stewardship experts in different European countries. Their input prompted widening the group to bring in additional expertise in the field of implementation research, primary care and paediatrics. A second face-to-face meeting the working group used the findings of the systematic review to develop a theoretical framework through which researchers can address these four key research design considerations. The group proposed a series of key questions researchers can use to highlight the major issues they need to address to arrive at an optimal design for their specific research project. Final agreement of recommendations presented here by all 18 members of the working group was achieved by email.

A theoretical framework for designing antimicrobial stewardship evaluations

The impact of intervention design

Detailed discussion of how antimicrobial stewardship interventions are designed is beyond the scope of this guidance. However, the design of the scientific evaluation of an intervention depends on how that intervention was designed, and this then may depend on a set of interdependent considerations (Fig. 1A). The intervention rationale should include its basis in theory and existing evidence (Table 1 is a glossary of terms used in this guidance.). The existing evidence that informed the research question should be clearly explained on an efficacy-effectiveness-implementation spectrum [12], as these considerations will determine how outcomes are selected and prioritized (Fig. 1B). Detailed characterization of the intervention setting is required to allow assessment of external validity and to minimize selection bias. Stewardship interventions are typically multifaceted and each intervention feature must be specified precisely. The same holds for how the intervention's impact will be determined; this will influence definition and selection of outcomes, selection of clusters/sites and feasibility of blinding. The intervention aims will be informed by the rationale and setting and will also be key to selecting the primary and secondary outcomes; whether these will determine effectiveness and safety or how implementation results change antimicrobial use and what data are required to support translation of study findings into practice. These considerations will inform whether the research sets out to determine superiority or non-inferiority of the intervention measured by its primary outcome(s) against standard practice and the detectable effect sizes/non-inferiority margins, the most appropriate study design (e.g. experimental or quasi- experimental) and the detailed design features.

Recommendations regarding selection of outcome measures

When assessing the impact of a stewardship intervention, researchers should aim to consider all intended and potential unintended effects [13–15]. Outcome measures can be helpfully

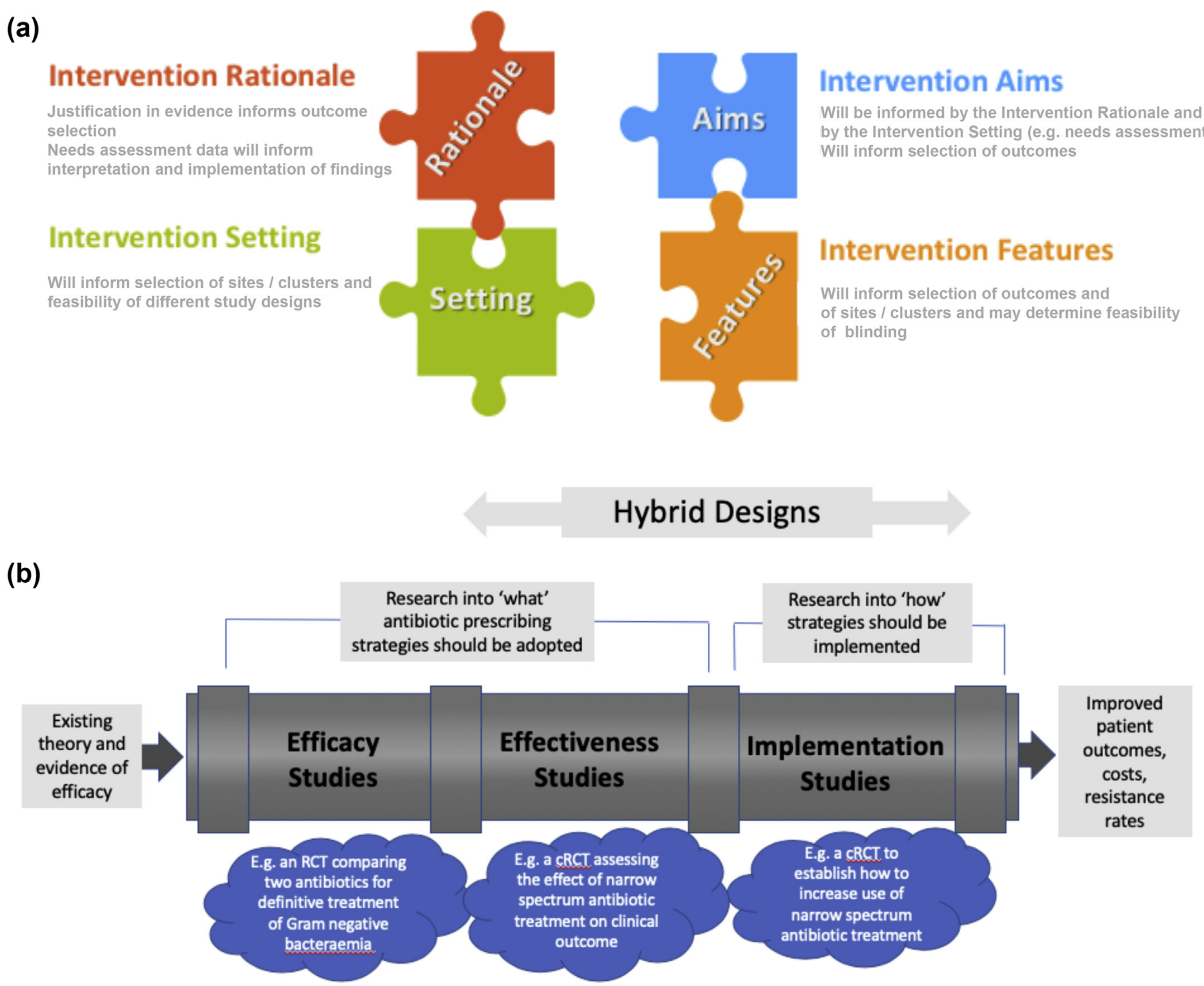

Fig. 1. (A) Interacting considerations relating to the intervention to be evaluated and their impact on study design. (B) An evaluation pipeline for antimicrobial stewardship intervention. Adapted from [12].

Table 1
Glossary of terms

Term	Explanation
Intervention rationale	The theory and evidence behind the stewardship intervention which is to be evaluated encompassing external factors (e.g. behavioural theory, evidence from previous research) and the clinical setting
Clinical setting	The environment in which the intervention is evaluated, both physical (e.g. ICU, emergency room, hospital type, primary care, long-term care) and practical (e.g. prescribing practice, team structures, staffing, behaviour)
Intervention aim(s)	The improvement being sought (e.g. reduction in inappropriate antimicrobial prescribing, reduction in use of specific antimicrobial classes or reduced *Clostridium difficile* infection)?
Features of the intervention	The different elements which make up a multifaceted intervention (e.g. education, decision support)
Cluster	A unit representing a group of smaller components, at which an intervention is delivered (e.g. a hospital ward representing all the doctors working in it, a group of primary care physicians working in a practice)
Outcomes of interest	The outcomes measured to determine effectiveness, safety and costs of the intervention
Experimental design studies	Studies which use randomization to allocate the stewardship intervention and control, either to individual patients/ professionals or clusters of patients/professionals
Quasi-experimental design studies	Studies which don't use randomization to allocate the stewardship intervention but rather use as controls different time period(s) and/or site (s), either external (controlled before-after studies) or internal (interrupted time series analyses, before-after studies)
Contamination	Unintended exposure of patients in the control phase or cluster to some or all of the intervention
Efficacy study	A study which assesses whether an antimicrobial use intervention produces the expected result under ideal and controlled conditions
Effectiveness study	A study which assesses whether an antimicrobial use intervention produces the expected result under 'real-world' pragmatic conditions
Implementation Study	A study which assesses the impact of an antimicrobial use improvement strategy in daily practice
Mediator analyses	Techniques to investigate mechanisms through which complex interventions achieve an observed effect
Superiority analysis	An analysis which sets out to determine if the intervention or strategy being assessed is better than comparator
Non-inferiority analysis	An analysis which sets out to determine whether the intervention or strategy being assessed not worse (by a prespecified amount, the non-inferiority margin) than comparator
Process Indicators	Measures of the care that is actually delivered to the patients (e.g., empirical regimen according to guideline)
Structure indicators	Measures of the organization of the healthcare system (e.g., the availability of a stewardship team)
Ecological assessment (of antimicrobial resistance)	Measurement of burden if antimicrobial resistant organism(s) or gene(s) in the environment or aggregated patient samples

grouped into three domains as clinical (typically to assess safety of an antimicrobial-sparing intervention in terms of patient outcome), microbiological (resistance) and care related (processes and structures of care, sometimes referred to as quality or performance outcomes) [16] (Table 2). Whether the study is primarily assessing effectiveness, implementation or a combination of both, will determine how outcomes are selected and prioritized, but, in general, appropriate outcome measures should be prospectively defined from each of the three domains. It is essential to recognize that although individually randomized efficacy trials aim to avoid selection bias, the inevitably restricted populations that enter such trials can potentially lead to generalisability bias, making extrapolation to wider populations challenging. While stewardship studies typically assess interventions made at the cluster level, assessment of clinical, microbiological and care-related outcomes is often possible at an individual patient level and should be included where possible to address this.

Clinical outcomes are missing from many published stewardship studies. In fact, most of these studies were not sufficiently powered to exclude clinically meaningful harm. Concern that this prevents adoption of antimicrobial reduction strategies into practice has led some to call for routine use of co-primary clinical outcomes in stewardship evaluations [17]. The working group felt that clinical outcome measures should always be prespecified and reported. Exceptions could be implementation studies of interventions for which concerns over safety will not be a barrier to adoption of their findings.

Microbiological outcomes address the impact of the intervention on antimicrobial resistance and/or rates of *Clostridium difficile* infection. A central rationale for antimicrobial stewardship interventions is that reducing antimicrobial exposure should reduce harm to a patient's microbiome and selection for antibiotic resistance. However, the evidence base remains sparse, and mostly of low quality, with lack of reliable pre-intervention data a particular limitation [9,18,19]. Incorporating assessment of colonization/infection by resistant organisms within a stewardship study can be challenging as event rates are often low and the relationship between antimicrobial exposure and resistance may be temporally distant and complicated by interactions with exposure to resistant pathogens and infection control measures. The working group agreed that while reductions in antimicrobial resistance should not be the primary outcome of stewardship studies, measurement of prevalence or incidence of *C. difficile* infection and of antimicrobial resistance should be included in the design where possible, and it should be clear whether measured resistance is in relation to the infecting pathogen and type of infection or among colonizing strains.

Care provision outcome measures (sometimes called quality or performance measures) include process indicators, prescribing behaviours and antimicrobial use data. These are usually relatively straightforward to obtain and are important to gather and report since clinical outcomes can only be interpreted meaningfully if it is clear that patient management has truly changed. Process indicators may address prescribing quality (e.g. guideline adherence or documentation practice) and reveal mediators of observed results. They are particularly important in implementation research to assess how the intervention under evaluation was actually delivered across the study (fidelity). This allows distinction between strategies that do and do not change the behaviours they aim to change and identification of those elements of an intervention that are impactful and of barriers for implementation [11]. Gathering appropriate qualitative data (e.g. from service managers, care providers and patients as appropriate) will allow an intervention's impact on cultural aspects of antibiotic use to be evaluated. Process outcomes are needed to assess organizational impact, of both implementation and long-term sustainability. Sustainability assessment is particularly important when an intervention has significant organizational-level impact through diversion of activity or cost [20]. For detailed consideration of these issues researchers should consult current guidance on development and evaluation of complex interventions [21].

Table 2
Outcome measures in antimicrobial stewardship evaluations

Examples	Notes
Clinical outcome measures	
Clinical cure, clinical failure, time to clinical response, recurrence rate.	Typically used to determine the safety of the intervention in terms of patient treatment outcome.
Mortality, length of stay, need for escalation of care (e.g. from ward to high dependency or critical care), (re)admission to hospital, revisits.	May include microbiological evidence of clinical outcome (e.g. microbiological cure or recurrence).
Patient-reported outcomes (e.g. quality of life measures).	Most are directly relevant to the individual patient.
	Important safety outcomes which are relatively easy to gather at cluster-level, but may only be linked partially to the intervention and may be a long way down the patient pathway
Adverse drug reactions, drug–drug interactions.	Gathering relevant data may require individual consent but could be from a subset of patients or use anonymized electronic records.
Microbiological (resistance) outcome measures	
Colonization by antimicrobial resistant pathogens (e.g. MRSA or multi-drug resistant (MDR) Enterobacteriaceae)	Valuable as short-term surrogate measures of antimicrobial resistance-related harm but relevance to individual patients is indirect through risk of antimicrobial resistant infection in the future or through transmission.
	Ecological assessments may be more feasible than individual patient-level measurement
Infection by specific organisms (C. *difficile*, antimicrobial resistant bacteria)	Outcome directly relevant to the impact of the antimicrobial intervention on the individual patient but uncommon and may require long follow-up beyond that needed for clinical outcomes
Care provision (quality or performance) outcome measures	
Drug use (e.g. Defined daily doses (DDD) or Days of Therapy (DOT) per admission or per bed-day.	Measurement of antimicrobial use (e.g. volume, range of agents) used to determine whether the intervention has potential to have an effect on clinical or microbiological outcomes (if no impact on process, then no clinical/microbiological impact by definition).
Appropriateness of treatment (e.g. proportion of prescriptions in accordance with guidelines).	Can be selected to measure appropriateness of antimicrobial selection.
Measures of intervention (e.g. recommendations given, use of clinical decision support)	Important for health-economic analyses and assessment of sustainability.
Resource requirements (e.g. staff time, clinical consultations, diagnostic testing)	Important for mediator analyses.
Costs measures.	

Timing of outcome measurements

Within each domain of outcome measure, consideration must be given to appropriate timing depending on the nature of the intervention and population (e.g. long- and short-term mortality, clinical complications during hospitalization or after discharge). Timing of measurement of microbiological outcomes should be considered to assess impact on resistance including *C. difficile* and timing of process outcome measurements should be considered to assess long-term sustainability.

Establishing superiority or non-inferiority

Where a stewardship study sets out to establish the effectiveness of an intervention, incorporation of appropriate controls is essential if the results are to inform practice, irrespective of whether an experimental or non-experimental design is used (see below). Researchers need to decide whether their primary objective is to determine superiority or non-inferiority of the intervention vs control.

Interventions aiming to improve treatment outcome

In some situations, a relevant clinical benefit can be hypothesized for an intervention (e.g. an intervention that focuses on increasing earlier targeted treatment based on test results or preventing under-treatment) and a study assessing the effectiveness of the intervention would seek superiority of the intervention vs. control for an appropriate primary clinical outcome.

Interventions aiming to reduce antimicrobial exposure

In most situations, stewardship interventions aim to preserve clinical outcome while reducing unnecessary antimicrobial exposure (e.g. less inappropriate initiation of antibiotics, choice of narrower spectrum or shorter duration) and improving quality of prescribing. As a result there is often some degree of real or perceived risk of patient-level harm, which may be specific to the intervention, patient population, setting and disease. Researchers designing effectiveness evaluations should consider what potential for patient harm would prevent adoption of the intervention *even if it were effective in reducing antimicrobial exposure*. Researchers should select appropriate secondary clinical end-point(s) to address this concern. Ideally in this situation the research should seek both superiority for an appropriate process measure *and* non-inferiority (i.e. not qualitatively worse than control) for a co-primary clinical outcome. The key measure to assess non-inferiority is the non-inferiority margin, being the smallest outcome difference for which the intervention would be considered no worse than control. The size of the non-inferiority margin strongly influences the sample size required to demonstrate non-inferiority with sufficient power. What margin is chosen depends on the outcome selected. The margin needs to be small enough to exclude relevant harm, which would prevent intervention implementation into practice. Researchers should justify the non-inferiority margin chosen with regard to severity and frequency of the outcome in the control group (which may, for example be affected by case mix [22].

Naturally, trials designed for demonstrating non-inferiority of clinical outcomes usually require large sample sizes. In such trials an interim analysis of a process outcome could be used to determine futility; if the intervention does not lead to the pursued process change continuing that intervention may not be logical, as non-inferiority will be the inevitable outcome.

Recognizing that achieving adequate power to exclude clinically relevant non-inferiority will not always be feasible, the group felt that researchers should at least specify and report point estimates and confidence intervals for a single prespecified lead clinical outcome. Bayesian analyses may be helpful to directly estimate the probability that intervention is more than 2.5%, 5%, 7.5%, etc., inferior to control [23]. Researchers should also prespecify the clinical outcomes they will use to assess the safety of the intervention, and all available clinical outcome data should be reported, in order to allow future meta-analysis. Unavailability of data should be explained. Unplanned exploratory analyses of clinical outcomes should be reported as such.

In studies addressing how interventions with established efficacy should be implemented, the quantitative outcome measures will be predominantly process measures and comparisons will seek to determine superiority of the intervention over comparator.

Sample size calculations

Studies evaluating effectiveness of an antimicrobial intervention need to be powered to demonstrate clinically relevant non-inferiority. In a superiority trial, detecting a large effect with high probability is almost always possible at a feasible sample size. Whereas demonstrating superiority only requires the confidence interval for the effect estimate to exclude zero, regardless of its width, determining non-inferiority requires the entire confidence interval to lie below the non-inferiority margin [24]. As a result, much larger participant numbers are usually required to demonstrate non-inferiority within clinically relevant margins which may be very small and difficult to define for outcomes such as mortality [25]. This difference lies in that superiority trials tend to be powered on an expected effect, which is often larger than what would be deemed a clinically relevant effect, whereas non-inferiority trials need to be powered on a clinically relevant effect.

One proposed solution to this issue is the Desirability of Outcome Ranking (DOOR)/Response Adjusted for Days of Antibiotic Risk (RADAR) approach, which uses investigator-ranked composite outcomes. This approach is based on the assumption that the same outcome with less antimicrobial exposure is desirable [26]. Yet, problems with clinical interpretation and sensitivity to the clinical outcomes chosen have been reported [27,28]. It remains to be determined to what extent the RADAR approach can robustly establish the effectiveness of novel stewardship interventions.

Interrupted time series studies require enough sequential measures before and after the intervention; the study's power will depend on the number of data points, their distribution, variability, the expected strength of the intervention effect and confounding factors such as seasonality [29], and therefore there are no straightforward sample size formulae. Researchers should consider the minimal requirements set out in the Cochrane Effective Practice and Organisation of Care (EPOC) resources [30].

Study design

Stewardship interventions typically target prescribers or other healthcare professionals rather than individual patients. As a consequence, evaluations involving individual patient randomization are usually not possible because of contamination. Instead, intervention allocation must be clustered (e.g. hospital, ward, primary care practice or physician). An important advantage of allocation at the cluster level is that it is more representative of real-life clinical practice. It is therefore more suited to studying both antimicrobial use interventions and antimicrobial improvement strategies rather than efficacy. Whereas in individual patient trials, randomization can be expected to control for confounding bias and maximize internal validity, with cluster randomized controlled trials (cRCTs), researchers need to give careful consideration to how clusters are defined and characterized. Clusters should be defined at the lowest level (e.g. clinical team, ward, practice, hospital) where contamination is unlikely as this will maximize the number of available clusters and hence study power. However, with the small number of clusters typically available in stewardship evaluations, randomization cannot be relied on to avoid imbalance between intervention and control clusters. Therefore baseline imbalances which may influence the intervention's impact (e.g. antimicrobial use, antimicrobial resistance rates, infection control standards, antimicrobial stewardship structures and processes, case mix of patients) should be specified *a priori* and data on these should be gathered for inclusion in multivariate analyses. Baseline imbalance in factors which a strong association with outcome or that could potentially modify the effect of the intervention can be addressed through stratified randomization (e.g. putting clusters into similar pairs and allocating one of each pair randomly to intervention vs. control), or use of a crossover design (see below). Cluster characterization is also essential to understand any observed heterogeneity of the intervention's effect between clusters. It optimizes external validity by allowing others to judge the representativeness for their clinical practice and to understand the logistical challenges of implementation.

Experimental study designs

Three main forms of cluster-randomized design may be appropriate depending on the intervention (Table 3). As above, *parallel cRCTs*, in which each cluster is randomized to either the intervention or control, minimize risk of contamination and maximize independence of the intervention from cluster-level characteristics. In some situations, perceptions of the intervention may influence whether clusters are willing to be randomized to control or intervention arms and hamper participation or introduce bias. *Stepped-wedge cRCTs (swcRCTs)* overcome this issue since all clusters receive the intervention during the trial, and allow estimation of the intervention effect within each cluster. swcRCTs can be logistically challenging to deliver since some clusters may have to wait to introduce the intervention and exposure should be avoided. Furthermore, the analysis of swcRCT is more complex [31]. Randomization of time of implementation is crucial to ensure independence of the timing of introduction from cluster-level factors. *Crossover cRCTs* offer the potential to estimate intervention effects in both directions—i.e. introducing and withdrawing, but may not be practicable (e.g. it may not be feasible to withdraw an educational intervention. Alternatively, the washout phase of a crossover study may be considered an assessment of sustainability for some forms of intervention. Assessment of carried antimicrobial resistance in crossover designs may need to consider the potential for resistance selection to persist.

A particular challenge with evaluation of interventions made at a cluster rather than patient level is intracluster correlation [32]. This must be incorporated into the sample size calculation otherwise a trial may be underpowered. Intracluster correlation is the extent to which patients are more similar to each other within a cluster than they would be if selected at random. The intracluster correlation coefficient (ICC) of an outcome is a measure of the relatedness of clustered data by comparing the variance within

Table 3
Design recommendations for experimental evaluations antimicrobial stewardship Interventions

Feature	Recommendations		
	Parallel cRCTs	Stepped-wedge cRCTs	Crossover cRCTs
Cluster selection	Randomized implementation at the lowest level (e.g. prescriber, ward, hospital, primary care practice) at which contamination can be minimized Define eligibility criteria and document representativeness of included clusters with respect to system from which they are drawn (e.g. size, case mix)		
Cluster allocation and randomization, timing of intervention	Ensure allocation concealment until the intervention is implemented (as complete blinding to allocation after randomization is often not feasible)	Conceal timing and order of intervention/crossover as much as possible Timing of intervention should be determined externally and at random, where possible	
Cluster balance	Pursue good/excellent balance between clusters (e.g. matching, stratified randomization based on factors likely to be associated with the outcome under study). No lower limit above which randomization will ensure balance but particularly problematic if there are fewer than 20 clusters per randomized group Collect data to document balance between clusters	Good/excellent balance between clusters achieved through design	
Blinding	Consider the objectivity of the selected outcomes and the extent to which patients and assessors of outcomes can be blinded to the cluster allocation		
Outcomes	Specify a primary or co-primary process outcome Specify a co-primary clinical outcome or at minimum one lead clinical outcome, and specify and report secondary clinical outcomes even if not powered on these Specify and analyse outcomes in each domain—clinical, microbiological, process (quantity or quality of antimicrobial use) Within implementation research, process outcomes should be selected with regard to complex intervention methodology [21], e.g. measures of fidelity, mediators and modifiers of the intended effect and measures of organizational impact Consider all important harms/unintended effects including 'squeezing the balloon' effects in which achieving the intended reduction in antimicrobial overuse results in an unintended increase in harmful overuse elsewhere [14,15,38] Define timing of different cluster-level and individual-level outcomes		
Power calculation	Provide sample size calculations to demonstrate study power—for the primary/co-primary outcome(s), and taking intra-cluster correlation into account		
Analysis	Adjust for secular trends (particularly for stepped-wedge cRCTs)		
Selection of patients for outcome evaluation	Ensure robust consistent inclusion of patients in control and intervention clusters/phases Report denominators from whom included patients were selected wherever possible		
Follow-up of patients	Timing of patient follow-up to assess patient-level outcomes should consider relevant time scales for both effectiveness and harms		
Follow-up of clusters	Consider duration of follow-up both for immediate effect of the intervention and sustainability		Only possible with short-term interventions with rapid loss of effect post withdrawal
Reporting	Report according to CONSORT criteria for cluster RCTs, stepped-wedge cRCTs, and other CONSORT guidelines as appropriate (e.g. pragmatic trials, non-inferiority trials). Consider using the TiDier checklist to clearly describe any behavioural intervention [39]		

CRCT, cluster randomized controlled trial.

clusters (e.g. hospitals) with the variance between clusters. A high ICC means that observations within clusters are much more similar to each other than to observations in other clusters, while an ICC of zero means that observations within one cluster are equally similar to each other than to observations in other clusters. In general, if the ICC is large, research designs with crossover are more efficient, while if the ICC is low, parallel cluster designs are more efficient [32].

Quasi-experimental study designs

In situations where randomization is not feasible or ethically not acceptable (see below), quasi-experimental, before–after studies have the potential to deliver robust evidence of a causal relationship between an intervention and measured outcomes if they incorporate appropriate controls and analyses which account for time trends (Table 4). Where control is provided through comparison with centre(s) where the intervention is not introduced, the term controlled before–after (CBA) study is used. Where control is provided by use of pre-intervention observations within centres, and secular time trends in the outcomes are specifically accounted for, the term interrupted time series (ITS) study is used. In practice, ITS reflects a method of analysis, being used for before and after studies and CBA, rather than a specific study type and can also be applied to CBA studies. CBA studies which do not control for time trends are unlikely to provide reliable evidence, regardless of external control [19]. The working group agreed that design of quasi-experimental evaluations of stewardship interventions must always account for changes in time [33,34]. Such analyses require sufficient pre-intervention time points to incorporate segmented regression analysis, and should consider adjustment for autocorrelation (e.g. using ARIMA models). Such analyses should report immediate effects on outcome and trends before and after the implementation, and assess whether trends are non-linear [29,35]. Furthermore, the timing of intervention implementation must be externally set to avoid the problem of regression to the mean which occurs when sites introduce a stewardship intervention in response to deterioration in the chosen outcome measure. Detailed guidance on conduct of Interrupted Time Series analyses are available through EPOC [30] and described in a recent review [36].

Ethical considerations

Antimicrobial stewardship measures which balance immediate and individual risks against future and societal access to effective antimicrobials raise challenging ethical issues around intergenerational justice, global distributive justice and protection of public health [37]. A key ethical issue in stewardship research is that, by gathering evidence for safety through clinical outcome measures, the possibility of individual harm is acknowledged. Individual patient consent may not be feasible in studies of interventions which act on prescribers or structures such as hospitals or clinics. This may set a higher ethical barrier than for individually randomized studies in which informed consent can be obtained. In this situation the research design process should involve patients to ensure that

Table 4
Design recommendations for quasi-experimental evaluations antimicrobial stewardship Interventions

Feature	Recommendations
Control	Even in situations where randomization is not possible (e.g. too few available clusters) allocation to intervention or control group should be made externally if at all possible, i.e. not depending on known factors or clinician preference Consider trying to match controls to minimize risk of bias arising from intrinsic differences between control and intervention groups
Timing	Timing of intervention should be externally set OR if this is not possible timing must be explained and described
Data	Data from automated electronic data recording (e.g. antimicrobial use data, routine electronic patient data) can be used retrospectively for pre-intervention data providing that collection/entry is consistent over calendar time, otherwise all data should be collected prospectively Measure, report and analyse any concurrent changes in case-mix, changes in methodology of outcome assessment, and care practices
Analysis	Include at least 12 monthly time points before and after the intervention to allow for anticipated secular trends [36,40] Use segmented regression or ARIMA models to account for secular trends Include at least 100 observations per time point [40] Check and, if necessary, correct for autocorrelation
Outcomes	See Table 3
Follow-up of patients	Timing of patient follow-up to assess patient-level outcomes should consider relevant timescales for both effectiveness and harms
Follow-up of clusters	Consider duration of follow up both for immediate effect of the intervention and sustainability
Reporting	Report according to relevant recommendations; STROBE-AMS [41] or STROBE [42] and the TiDier checklist [39], SQUIRE to describe in detail quality improvement component of study [43], TREND statement for nonrandomized evaluations of behavioural and public health interventions [44]

independent non-research views from the relevant patient population about these trade-offs are heard, actively considered, and incorporated into the final design. Additionally, researchers should be able to justify why the interventions under examination are reasonable choices of practice which could also be made outside the study setting. Studies in which the intervention is made at a cluster level will often still use individual patient data. Any requirement for individual patient consent to collect data may lead to loss of representativeness and a biased assessment of the intervention effect. Because consent is acquired with knowledge of the intervention, there is an increased risk of selection bias, e.g. if investigators are more motivated to enrol patients during the intervention period. Depending on the national regulations, in some countries study designs can address this issue through use of de-identified or anonymous data (e.g. through electronic patient records) of parameters collected routinely in clinical practice without the need for individual patient consent.

Key design decisions

The consensus group considered that researchers planning antimicrobial stewardship evaluations must make a set of key decisions (Table 5) that will ultimately determine optimal study design. We have classified these decisions based on whether they apply to the *intervention itself*, the *evaluation setting*, the *outcomes of interest*, the *research objective* and *type of study*. Detailed explanation of the decisions are presented (please see supplementary materials).

Discussion and conclusions

The theoretical framework and design recommendations we present have been developed by a diverse international working group with broad and substantial expertise in antimicrobial stewardship research and practice. They address aspects of study design which are crucial to translation of research into practice and will, we believe, increase the impact of future research in this field. By drawing on wide expertise and building on our comprehensive systematic review we consider our recommendations relevant across diverse settings of care. Our work has some notable limitations. Although we gave careful consideration to the breadth of expertise required on the group and sought external advice, we did not seek lay input. We cannot discount the possibility that this would have changed our emphasis, around patient reported outcome or experience measures for example. Given the technical nature of our guidance we think it unlikely this would have changed our conclusions. An inherent risk of the consensus group design is 'group think' in which members trying to reach consensus fail to critically evaluate alternative views. To address this we sought critical evaluation by two highly eminent international experts in this field. Although these were also, of necessity, experts in antimicrobial stewardship research, the impact of their input on our thinking, the breadth and seniority of expertise in our group make it unlikely we have failed to consider major alternative viewpoints. Notwithstanding these caveats, we believe that application of this guidance has the potential to greatly improve the quality and impact of antimicrobial stewardship research.

Summary recommendations

Outcomes

- Researchers should determine whether their study aims to investigate, effectiveness or implementation ('what or 'how'). This will determine the priority and nature of outcomes.
- All antimicrobial stewardship studies should define process, clinical and microbiological outcomes and specify a primary

Table 5
Key design decisions. A detailed explanation of the rationale and how these address different aspects of design is set out in the supplementary materials

Question	Design aspect addressed
Where does knowledge gap the study aims to address lie on a spectrum between 'what' and 'how' questions?	Selection and prioritization of outcomes
What are the risks of contamination?	How clusters will be defined within the study
Is it possible to remove the intervention after it has been implemented?	What study design will be most appropriate.
Is the intervention impact threatened by sustainability?	Selection and timing of study outcomes
What forms of bias threaten the validity of the study?	Cluster selection; feasibility of blinding; data collection
What features of the evaluation setting will impact on external validity?	Cluster selection; feasibility of blinding; data collection
Is it possible to blindly assess the outcome?	Feasibility of blinding

process outcome(s) to measure effectiveness of the intervention.
- Unless there is pre-existing evidence that a stewardship intervention cannot or will not compromise treatment outcome, an evaluation should attempt to prespecify a co-primary clinical/microbiological efficacy outcome on which the study is adequately powered or, at minimum, a single, lead clinical outcome.
- Clinical and microbiological data documenting treatment outcome should be collected and reported as prespecified secondary outcomes even if the study is not powered on them
- Measurement of incidence of infections/colonization due to multidrug-resistant bacteria and infections due to *C. difficile* infection should be included in the design of stewardship interventions whenever possible. Studies assessing resistance should clarify whether this is related to the infecting pathogen or among colonizers.

Objectives

- If a relevant clinical benefit can be hypothesized for an intervention, then the research objective should seek superiority for an appropriate primary clinical outcome.
- If not, researchers should seek both superiority for an appropriate process measure and ideally non-inferiority for a co-primary clinical/clinically relevant microbiological outcome.
- Researchers should justify how the non-inferiority margin has been selected and balanced against research costs and feasibility.
- Where this is not possible, as a minimum, researchers should specify, and report point estimates and confidence intervals for, at minimum, a single prespecified lead clinical outcome.
- In situations where the study size is determined by a co-primary non-inferiority safety outcome, an interim futility analysis of the superiority process outcome should be considered to confirm a relevant change in treatment/management.

Study design

- Cluster randomized controlled trials (including crossover and stepped-wedge designs) are preferable to quasi-experimental before/after studies.
- The threshold for defining clusters should be as low as possible to minimize contamination, allowing the maximum number of clusters to be studied.
- In a parallel cluster RCT, randomization should not be relied on to control for imbalance between study arms if the number of clusters is <20 per arm and stratified or matched randomization should be considered
- Designs using within-cluster comparisons (stepped-wedge cRCT, crossover cRCT or quasi-experimental approaches) are indicated where there are fewer than ten clusters per arm.
- Quasi-experimental studies should incorporate appropriate controls and analyses to account for time trends.
- In quasi-experimental studies, timing of the intervention should be externally set or if this is not possible timing should be explained and described.
- Segmented regression analysis of interrupted time series studies should include 12 time points with at least 100 observations per time point before and after the intervention to allow for anticipated secular trends and test or correct for autocorrelation.
- Single centre studies using a robustly designed and analysed interrupted time series approach including observations before and after the intervention should be considered the lowest quality research design which will impact on clinical practice.

Transparency declaration

The authors declare no conflicts of interest. The presented work was supported by a grant (JPIAMRWG-010) from the Joint Programming Initiative on Antimicrobial Resistance. A.S.W. is supported by the NIHR Oxford Biomedical Research Centre and the Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with Public Health England (PHE) [HPRU-2012-10041], and is an NIHR Senior Investigator. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

Acknowledgements

The authors would like Prof Dilip Nathwani and Prof Jan Prins for providing independent critical advice during the project and to Sandy Gray for administration support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cmi.2019.08.017.

References

[1] The review on antimicrobial resistance. Antimicrobial resistance: tackling a crisis for the health and wealth of nations. 2014. Available at: https://amr-review.org.
[2] 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;340:c2096.
[3] 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. Lancet 2005;365:579–87.
[4] Dyar OJ, Huttner B, Schouten J, Pulcini C, Esgap. What is antimicrobial stewardship? Clin Microbiol Infect 2017;23:793–8.
[5] McGowan Jr JE, Gerding DN. Does antibiotic restriction prevent resistance? New Horiz 1996;4:370–6.
[6] Molina J, Cisneros JM. Editorial commentary: a chance to change the paradigm of outcome assessment of antimicrobial stewardship programs. Clin Infect Dis 2015;61:807–8.
[7] 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. Clin Microbiol Infect 2018;22(5):555–61.
[8] Hulscher M, Prins JM. Antibiotic stewardship: does it work in hospital practice? A review of the evidence base. Clin Microbiol Infect 2017;23:799–805.
[9] Schuts EC, Hulscher M, Mouton JW, Verduin CM, Stuart J, Overdiek H, et al. Current evidence on hospital antimicrobial stewardship objectives: a systematic review and meta-analysis. Lancet Infect Dis 2016;16:847–56.
[10] Ramsay C, Brown E, Hartman G, Davey P. Room for improvement: a systematic review of the quality of evaluations of interventions to improve hospital antibiotic prescribing. J Antimicrob Chemother 2003;52:764–71.
[11] Hulscher ME, Laurant MG, Grol RP. Process evaluation on quality improvement interventions. Qual Saf Health Care 2003;12:40–6.
[12] Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012;50:217–26.
[13] Burke JP. Antibiotic resistance–squeezing the balloon? JAMA 1998;280:1270–1.
[14] Peterson LR. Squeezing the antibiotic balloon: the impact of antimicrobial classes on emerging resistance. Clin Microbiol Infect 2005;11:4–16.
[15] Toma M, Davey PG, Marwick CA, Guthrie B. A framework for ensuring a balanced accounting of the impact of antimicrobial stewardship interventions. J Antimicrob Chemother 2017;72:3223–31.
[16] McGregor JC, Furuno JP. Optimizing research methods used for the evaluation of antimicrobial stewardship programs. Clin Infect Dis 2014;59:S185–92.
[17] 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;18:595–7.
[18] Baur D, Gladstone BP, Burkert F, Carrara E, Foschi F, Dobele 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;17:990–1001.
[19] 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;2:CD003543.
[20] Lennox L, Maher L, Reed J. Navigating the sustainability landscape: a systematic review of sustainability approaches in healthcare. Implement Sci 2018;13:27.
[21] Medical Research Council. Developing and evaluating complex interventions. 2019.
[22] Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, Group C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012;308:2594–604.
[23] Laptook AR, Shankaran S, Tyson JE, Munoz B, Bell EF, Goldberg RN, et al. Effect of therapeutic hypothermia initiated after 6 hours of age on death or disability among newborns with hypoxic-ischemic encephalopathy: a randomized clinical trial. JAMA 2017;318:1550–60.
[24] Mauri L, D'Agostino Sr RB. Challenges in the design and interpretation of noninferiority trials. N Engl J Med 2017;377:1357–67.
[25] Cranendonk DR, Opmeer BC, Prins JM, Wiersinga WJ. Comparing short to standard duration of antibiotic therapy for patients hospitalized with cellulitis (DANCE): study protocol for a randomized controlled trial. BMC Infect Dis 2014;14:235.
[26] Evans SR, Rubin D, Follmann D, Pennello G, Huskins WC, Powers JH, et al. Desirability of outcome ranking (DOOR) and response adjusted for duration of antibiotic risk (RADAR). Clin Infect Dis 2015;61:800–6.
[27] Phillips PP, Morris TP, Walker AS. DOOR/RADAR: a gateway into the unknown? Clin Infect Dis 2016;62:814–5.
[28] Schweitzer VA, van Smeden M, Postma DF, Oosterheert JJ, Bonten MJM, van Werkhoven CH. Response Adjusted for Days of Antibiotic Risk (RADAR): evaluation of a novel method to compare strategies to optimize antibiotic use. Clin Microbiol Infect 2017;23:980–5.
[29] Lopez Bernal J, Soumerai S, Gasparrini A. A methodological framework for model selection in interrupted time series studies. J Clin Epidemiol 2018;103:82–91.
[30] EPOC. CEPaOoC. What study designs can be considered for inclusion in an EPOC review and what should they be called? Cochrane Effective Practice and Organisation of Care; 2017. EPOC resources for authors. available at: https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/EPOC%20Study%20Designs%20About.pdf.
[31] Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. BMJ 2018;363:k1614.
[32] Hemming K, Taljaard M. Sample size calculations for stepped wedge and cluster randomised trials: a unified approach. J Clin Epidemiol 2016;69:137–46.
[33] Boel J, Andreasen V, Jarlov JO, Ostergaard C, Gjorup I, Boggild N, et al. Impact of antibiotic restriction on resistance levels of Escherichia coli: a controlled interrupted time series study of a hospital-wide antibiotic stewardship programme. J Antimicrob Chemother 2016;71:2047–51.
[34] Taggart LR, Leung E, Muller MP, Matukas LM, Daneman N. Differential outcome of an antimicrobial stewardship audit and feedback program in two intensive care units: a controlled interrupted time series study. BMC Infect Dis 2015;15:480.
[35] Lawes T, Lopez-Lozano JM, Nebot CA, Macartney G, Subbarao-Sharma R, Wares KD, et al. Effect of a national 4C antibiotic stewardship intervention on the clinical and molecular epidemiology of Clostridium difficile infections in a region of Scotland: a non-linear time-series analysis. Lancet Infect Dis 2017;17:194–206.
[36] 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. Clin Microbiol Infect 2017;23:819–25.
[37] Littmann J, Rid A, Buyx A. Tackling anti-microbial resistance: ethical framework for rational antibiotic use. Eur J Public Health 2018;28:359–63.
[38] Davey P, Peden C, Charani E, Marwick C, Michie S. Time for action-Improving the design and reporting of behaviour change interventions for antimicrobial stewardship in hospitals: early findings from a systematic review. Int J Antimicrob Agents 2015;45:203–12.
[39] 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;348:g1687.
[40] Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002;27:299–309.
[41] Tacconelli E, Cataldo MA, Paul M, Leibovici L, Kluytmans J, Schroder W, et al. STROBE-AMS: recommendations to optimise reporting of epidemiological studies on antimicrobial resistance and informing improvement in antimicrobial stewardship. BMJ Open 2016;6:e010134.
[42] Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. PLoS Med 2007;4:e297.
[43] Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf 2016;25:986–92.
[44] Des Jarlais DC, Lyles C, Crepaz N, Group T. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004;94:361–6.