Systematic review:

The quality of studies evaluating antimicrobial stewardship interventions: a systematic review

Valentijn A. Schweitzer[1], Inger van Heijl[2], Cornelis H. van Werkhoven[1], Jasmin Islam[3], Kelly D. Hendriks-Spoor[2], Julia Bielicki[4], Marc J.M. Bonten[5], A. Sarah Walker[6], Martin J. Llewelyn[3], Consensus on Antimicrobial Stewardship Evaluations (CASE) study group*

[1]Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, The Netherlands

[2]Department of Clinical Pharmacy and Medical Microbiology, Tergooi Hospital, Hilversum/Blaricum, The Netherlands

[3]Department of Global Health and Infection, Brighton and Sussex Medical School, Falmer, United Kingdom

[4]Paediatric Infectious Disease Research Group, St George's University of London , London , United Kingdom

[5]Department of Medical Microbiology, University Medical Center Utrecht, The Netherlands

[6]MRC Clinical Trials Unit, University College London, London, United Kingdom

**Other members of the CASE study group are mentioned in the acknowledgements*

Keywords: antimicrobial stewardship, methodology, systematic review, quality, research design

Running title: Systematic review of antimicrobial stewardship design

Corresponding author:

Valentijn A. Schweitzer, MD

Julius Center for Health Sciences and Primary Care

University Medical Center Utrecht, Heidelberglaan 100

Post Box 85500, Utrecht 3508 GA, The Netherlands

V.A.Schweitzer-2@umcutrecht.nl

ABSTRACT

BACKGROUND:

Antimicrobial stewardship aims to optimise antibiotic use and minimise selection of antimicrobial resistance. The methodological quality of published studies in this field is unknown.

OBJECTIVES:

Our objective was to perform a comprehensive systematic review of antimicrobial stewardship research design and identify features which limit validity and translation of research findings into clinical practice.

DATA SOURCES:

The following online database was searched: PubMed.

STUDY ELIGIBILITY CRITERIA:

Studies published between January 1950 and January 2017, evaluating any antimicrobial stewardship intervention in the community or hospital setting, without restriction on study design or outcome.

METHODS:

We extracted data on pre-specified design quality features and factors that may influence design choices including: (1) clinical setting, (2) age group studied, (3) when the study was conducted, (4) geographical region and (5) financial support received.

RESULTS:

The initial search yielded 17,382 articles; 1,008 were selected for full-text screening, of which 825 were included. Most studies (675/825, 82%) were non-experimental and 104 (15%) used interrupted time series analysis, 41 (6%) used external controls and 19 (3%) used both. Studies in the community setting fulfilled a median of 5/10 quality features (IQR 3-7) and 3 (IQR 2-4) in the hospital setting. Community setting studies (25%, 205/825) were significantly more likely to use randomisation (OR 5.9 (95%CI 3.8-9.2)), external controls (OR 5.6 (95%CI 3.6-8.5)) and multiple centres (OR 10.5 (95%CI 7.1-15.7)). From all studies, only 48% (398/825) reported clinical and 23% (190/825) reported microbiological outcomes. Quality did not improve over time.

CONCLUSIONS:

Overall quality of antimicrobial stewardship studies is low and has not improved over time. Most studies do not report clinical and microbiological outcome data. Studies conducted in the community setting were associated with better quality. These limitations should inform the design of future stewardship evaluations so that a robust evidence base can be built to guide clinical practice.

INTRODUCTION

Antimicrobial resistance (AMR) is increasing globally and is a substantial threat to human health[1]. There is a clear relationship between antibiotic exposure and AMR both in populations [2] and individual patients [3]. An estimated 30% of human antibiotic use may be unnecessary and healthcare systems around the world are aiming to achieve substantial reductions in unnecessary antibiotic prescribing. The term 'antimicrobial stewardship' is used to describe use of antibiotics which balances the need for effective individual treatment against the longer-term, societal impact of antibiotic use on antibiotic resistance[4]. Interventions to improve antimicrobial stewardship are usually multifaceted and include periodic or individual patient audit and feedback, decision support, education (educational meetings, educational materials), and antimicrobial formulary restriction[5].

Recognition of the threat posed by AMR and the need to optimise antibiotic prescribing has driven an exponential increase in the publication of studies evaluating antimicrobial stewardship interventions over the last 20 years[6]. Previous systematic reviews have synthesised this evidence with the aim of making recommendations for practice[5,7–12]. These have, appropriately, considered studies with the lowest possible risk of bias but have excluded >50% of published studies in which methodological quality falls below Effective Practice and Organisation of Care (EPOC) criteria[13]. Because the minority of studies are of sufficient quality, many areas of practice rely on a weak evidence base [8]; conducting studies which do not inform practice is a waste of time and valuable resources [5,12].

Journals are beginning to report the minimum standards for antimicrobial stewardship studies to be published [14]; however, there remains a need for consensus on how to design, analyse and report studies evaluating interventions to improve antibiotic prescribing. This would optimise use of valuable resources and strengthen the evidence base in this field. Currently, no overview exists of how antimicrobial stewardship evaluations are designed. We conducted a systematic review of antimicrobial stewardship evaluations with the aim of identifying areas in stewardship evaluation most in need of improvement, to increase validity and translation of research findings into clinical practice.

METHODS

Search strategy

We searched PubMed for studies evaluating antimicrobial stewardship interventions between January 1950 and January 2017. The search strategy (appendix 1) was designed to be as broad as possible. Inclusion criteria were: any study evaluating an antimicrobial stewardship intervention, without restriction on the type of intervention studied and what outcomes were evaluated. Studies were excluded if they were (1) not in English, (2) case-reports, (3) focused mainly on HIV or (4) narrative or systematic reviews. All studies were screened by one author (VAS, CHvW, JI, KH or IvH). In case of duplicate publications, only the original article was included. A random selection of 700 (~4% of total) studies were assessed by a second author (VAS, CHvW, JI, KH or IvH). Uncertainties about the inclusion of studies was resolved by discussion.

Data extraction

Studies fulfilling the inclusion criteria were identified by screening on title and abstract. All selected studies then underwent full-text evaluation by one author (VAS, CHvW, JI, KH or IvH) against the inclusion and exclusion criteria and data was subsequently extracted using a standard data extraction template (appendix 2). Data were extracted on study characteristics (i.e. title, authors, year of publication), design quality features, and factors possibly associated with methodological quality. If no funding was reported it was assumed that studies received no financial support. Authors were not contacted in case data were missing or incomplete. A random selection of 10% of the studies were extracted by a second author. We followed the PRISMA criteria for the reporting of systematic reviews (appendix 3) [15].

Selection of quality features, and factors associated with quality

In February 2017 we established an international Consensus Working Group funded by the Joint Programming Initiative on Antimicrobial Resistance (JPIAMRWG-010) to develop recommendations on the design, analysis and reporting of antimicrobial stewardship evaluations. The working group coordinators (VAS, CHvW, ML, ASW, MB) invited members to the consensus group based on their expertise on antimicrobial stewardship and/or trial methodology, ensuring that all key clinical areas (primary care, secondary care, intensive care medicine and paediatrics) were represented. The consensus group agreed that a review to identify areas in stewardship evaluation most in need of improvement to increase validity and translation of research findings into clinical practice was required. For this purpose, existing quality scores were not applicable because these focus solely on methodological quality. The group selected quality features and factors likely to underlie design quality and features based on plenary group discussion during the consensus meeting. Ten quality features were selected for assessment: (1) randomized research design used, (2) external control group assessed, (3) multiple centres used, (4) sustainability of the intervention sufficiently assessed (≥12 months), (5) sample size calculation reported, (6) prospective data collection, (7) correction for confounding factors, (8) primary outcome defined and reported, (9) clinical outcome reported, and (10) microbiological outcome reported.. Selected factors likely to underlie design quality features were (a) the clinical setting (community versus hospital), (b) age group studied (studies including children versus adults, or both), (c) year when study was conducted (newer versus older studies, categorised at approximate quintiles: 1977-2004, 2005-2010, 2011-2013, 2014-2015, 2016-2017) (d) geographical region, and (e) financial support. The quality features and corresponding categorizations are shown in table S1.

Statistical analysis

Descriptive statistics were used to describe the quality features of included studies. Differences stratified by subgroup were displayed using spider graphs (Microsoft Excel, version 2010). To assess the independent relationship between factors and quality features we performed multivariable logistic regression models with backward stepwise selection (exit p-value >0.10), presenting odds ratios and 95% confidence intervals in a heat map displaying the strength of the association (Microsoft Excel, version

2010). Analyses were performed using the Statistical Package for the Social Sciences for Windows (Version SPSS 21.0.0.0).

RESULTS

The initial search yielded 17,382 articles. After title and abstract screening, 1,008 articles were selected for full-text screening (figure 1). Of these, 183 were excluded leaving 825 articles for full assessment (appendix 4). Among 700 randomly selected articles that were screened by a second author, 640/700 (91%) were excluded by both authors, 23/700 (3.3%) were selected for inclusion by both authors, 13/700 (1.8%) were selected for inclusion by only the first author, and 24/700 (3.4%) were selected for inclusion by only the second author, resulting in a percentage agreement after title/abstract screening of 95%, with a moderate interobserver agreement (Cohen's kappa: 0.53). After discussion of the full text articles, consensus about inclusion was reached in 99.5% (696/700) of the selected articles. From the 3.4% (24/700) studies that were newly identified by a second author, 7/700 (1.0%) were considered correct inclusions after discussion. Therefore, the low proportion of missed papers justified not screening in duplicate. Among the 83 articles selected for double data-extraction by a second author, the percentage agreement per variable ranged from 91%-100%, with all the quality features showing an agreement of ≥95%.

Characteristics of included studies

A total of 825 studies were included, In the community setting, most studies were multicentre (72%, 148/205) and the commonest study designs were before-after studies without an interrupted time-series (ITS) analysis (23%, 48/205), randomised controlled trials (14%, 28/205), and parallel cluster randomised trials (15%, 30/205)(Table 1). Among the included studies in the hospital setting, most were single centre (84%, 519/620) and the commonest study designs were before-after studies without an ITS analysis (59%, 365/620), before-after studies with an ITS analysis (13%, 82/620), and cohort studies without a control group (12%, 75/620). Among the before-after studies without an ITS analysis, the majority were

single centre (86%, 352/411), and retrospective (58%, 239/411). In the 86 (10%) cluster randomised studies a median of 28 clusters were randomised, with 57% (49/86) randomising ≥20 clusters, 28% (24/86) randomising <15 clusters, 21% (18/86) randomising <10 clusters, and 9% (8/86) randomising <5 clusters. In the hospital setting, 13% (2/13) randomised ≥20 clusters, while in the community setting 65% (64/71) randomised ≥20 clusters. A minority of studies were conducted in children (12%, 101/825). Both in the community and the hospital setting, most studies did not target a specific disease or syndrome (31%, 63/205, 55% 324/620, respectively) or specific antibiotic class (86%, 177/205, 69% 427/620, respectively). The majority of interventions were bundles (57%, 470/825). Commonest interventions in the community setting included education (73%, 149/205), audit with periodic feedback (29%, 59/205), and clinical decision support 15% (31/205). In the hospital setting, commonest interventions included education 42% (260/620), audit and feedback on an individual patient level (40%, 245/620), restriction (18%, 113/620), and clinical decision support (18%, 112/620). Both in the community and hospital setting, virtually all included studies reported process measure outcomes (99%, 818/825) (Table 2). Both in the community setting and the hospital the most commonly reported process measures included the proportion of patients treated with antibiotics (59%, 121/205, 21%, 131/620), costs/cost-effectiveness (18%, 36/205, 32%, 200/620, respectively), appropriateness (17%, 34/205, 29%, 178/620, respectively), and defined daily doses (17%, 34/205, 25%, 156/620, respectively). In the community setting, commonest reported clinical outcomes were revisits (11%, 22/205), clinical cure (6%,12/205), and infection (5%, 10/205), while in the hospital setting these were mortality (32%, 302/620), length of stay (32%, 201/620), and hospital readmissions (12%, 76/620).

Quality features

The percentage of studies including each quality feature is shown in Table 3. Studies in the community setting fulfilled a median of 5 quality features (IQR 3-7), while studies in the hospital setting fulfilled 3 (IQR 2-4). None fulfilled all 10 quality features. In the community setting 2% (4/205) fulfilled 9, 16% (33/205) fulfilled at least 8, and 35% (72/205) fulfilled at least 7 quality indicators. In the hospital setting 1% (4/620) fulfilled 9, 3% (19/620) fulfilled 8, and 6% (37/620) fulfilled 7 quality indicators. Of note, there were

substantial differences between studies which did and did not use randomised designs to the extent to which other quality features were present. Among the 150 randomised studies, all used an external control group, 71% (107/150) included multiple centres, 64% (97/150) reported a sample size calculation and 96% (144/150) collected data prospectively. In contrast, among the non-randomised studies, 12% (78/675) used an external control group, 21% (142/675) included multiple centres, 11% (76/675) reported a sample size calculation and 43% (288/675) collected data prospectively.

Factors associated with design quality

Design quality was considerably better in almost all quality aspects of community versus hospital setting studies (Figure S1A), with more use of randomised designs (46% vs. 9%), external controls (63% vs. 16%), sample size calculations (38% vs. 15%), prospective data collection (70% vs. 46%), correction for confounding (55% vs. 25%), use of a defined primary outcome (57% vs. 44%), and involvement of multiple centres (72% vs. 16%). However, community setting studies less often reported clinical (30% vs. 54%) and microbiological (8% vs. 28%) outcomes. Community setting remained significantly associated with all these factors in multivariable models (Table 4). From the studies that reported financial support, 20% (53/264) were industry funded, and 84% (221/264) were publicly funded. Studies with financial support were of higher methodological quality than studies without (Figure S1B), as they more frequently used randomised designs (31% vs. 7%), external controls (34% vs. 8%), sample size calculations (33% vs. 11%), prospective data collection (63% vs 43%), correction for confounding (46% vs. 21%), a defined primary outcome (56% vs. 39%), and involved multiple centres (46% vs. 16%). Financial support remained significantly associated with these factors in multivariable models (Table 4). In addition, financial support increased the frequency of reporting clinical outcomes in multivariable models. There was little change in design quality over time, other than a decrease in the proportion of studies with prospective data collection (77% in 1977-2004, 67% in 2005-2010, 42% in 2011-2013, 40% in 2014-2015, 39% in 2016-2017) and an increase in studies reporting a clinical outcome (39% in 1977-2004, 44% in 2005-2010, 44% in 2011-2013, 53% in 2014-2015, 59% in 2016-2017)) (Figure S1D). These outcomes were significantly associated with calendar time in multivariable models (Table 4), and sample

size calculations were independently reported more in later studies. The decrease in studies with prospective data collection is most prominent in studies in the hospital setting (Table S2). There were no large differences between studies performed in children versus adults (Figure S1C). Geographical region was independently associated with randomised designs, using an external control, prospective data collection, performing sample size calculations, reporting a primary, clinical or microbiological outcome, and being multicentre (Table 4).

DISCUSSION

In previous systematic reviews of antimicrobial stewardship evaluations, many studies have been excluded due to not fulfilling minimal methodological quality criteria. We have undertaken the first comprehensive systematic review focusing on describing quality, rather than excluding studies based on quality, to facilitate formulating recommendations for improvement. In addition, we evaluated quality features required for validity and translation into practice instead of focusing solely on methodological quality. Our systematic review revealed that the design quality of antimicrobial stewardship evaluations is low, with only a minority of studies reporting clinical and microbiological outcome data. Design quality is considerably better in studies performed in the community setting.

We find published evaluations provide a striking lack of evidence for the clinical and microbiological impacts of antimicrobial stewardship interventions. The majority of studies focus exclusively on process measures. While it is clearly essential to establish whether an intervention is effective in changing antibiotic use, reporting clinical outcomes is crucial to assess the safety of antimicrobial stewardship interventions [16,17]. The clinical outcomes reported often utilise routinely collected data, which may explain the differences between the community and hospital setting. In particular, in the hospital setting, commonly used clinical outcomes are mortality and length of hospital stay [18,19]. As indicated by the observed time trends, these outcomes are being used with increasing frequency, probably because extraction of relevant data from electronic health records is becoming more feasible. Such routinely available data are not the most sensitive and patient-relevant outcomes. In the hospital setting, markers

of early treatment response such as clinical stability may be preferable. In the community setting, repeat consultations, relapse of infection, and hospital admissions may be more relevant; yet data on these outcomes are not routinely collected [20–22].

Very few stewardship studies report microbiological outcomes. This is surprising given that reducing antimicrobial resistance is the ultimate goal of antimicrobial stewardship. However, this is consistent with a meta-analysis on the effect of stewardship interventions on infection and colonisation with antibiotic-resistant bacteria and *Clostridium difficile* infections that showed the literature on this topic is sparse and dominated by low quality research [23]. Some authorities have called for stewardship evaluations to routinely include consideration of the impact of stewardship on resistance [24] but studies generally lack power to determine this. Relationships between antimicrobial exposure and resistance may be more efficiently established through specific mechanistic studies rather than within stewardship evaluations.

Our analysis demonstrates that factors that would be expected to effect study design do, whiles others do not. The contrast between the community and hospital setting is striking in terms of the greater use of multicentre, randomised controlled designs. One explanation could be that clusters required for cluster randomisation are more readily available in the community setting. In contrast, clinical outcomes and microbiological data are less readily available in the community setting. Retrospective study designs are therefore less feasible in the community setting.

Financial support was associated with better design quality. In addition to the costs inherent to conducting multicentre, prospective studies with longer follow-up, the process of securing funding may drives careful consideration of study validity. Less than half of the stewardship studies reviewed reported external funding. However, our finding of an association between external funding and improved design quality underscore the necessity of external funding to support appropriate implementation and robust evaluation of antimicrobial stewardship programmes [25].

Our results show that there is no improvement of design quality over time, which is in contrast to previous reports[5,23]. This may be explained by our evaluation and inclusion of all studies without a pre-selection on study design, while previous reviews only included adequate studies with interpretable data[5]. .

In keeping with previous reports we find that the majority of antimicrobial stewardship studies used non-randomised research designs, with before-after studies being the most prevalent. This quasi-experimental research design is commonly used for quality improvement projects. However, such studies are at risk from multiple forms of bias [16,24,26] and the Effective Practice and Organisation of Care (EPOC) criteria strongly discourages the inclusion of before-after studies without an ITS analysis in systematic reviews[13]. Incorporation of properly conducted ITS analysis into before-after studies has the potential to robustly control for time-dependent bias [27] but only 20% (104/515) of before-after studies we identified used ITS analysis. Moreover, it has been shown that contemporary ITS analyses are often performed with an insufficient number of data points [11].

This systematic review has several strengths. First, our comprehensive search strategy gives a unique overview of the quality of studies evaluating antimicrobials stewardship interventions. Second, we used the PRISMA reporting guide for systematic reviews [15]. Third, the quality indicators and candidate factors were selected in a consensus procedure as part of a Joint Programming Initiative on Antimicrobial Resistance (JPIAMR) funded consensus group, which was selected to both include experts on the field of antimicrobial stewardship and trial methodology. Finally, this is the first comprehensive systematic review to determine the extent to which published antimicrobial stewardship evaluations include quality features required for validity and translation into practice .

The limitations of our review were firstly, we only searched PubMed and excluded non-English studies, which makes it possible that antimicrobial stewardship studies indexed elsewhere and non-English studies were missed. However, if we compare the studies identified by our searching strategy with the largest community and hospital stewardship systematic reviews, only 6 studies were missed due to being indexed elsewhere and 11 due to being non-English. Therefore, this is likely to have had a minimal impact on the total results[5,12]. Secondly, the screening, inclusion, and data extraction was performed by only one investigator that could have resulted in studies being missed, wrongly included or misclassification of the extracted data. To estimate the amount of studies that might be missed, a proportion of the studies were screened and data was extracted by a second author. In this second round, we showed that the percentage of agreement was high, with a moderate interobserver agreement.

Assuming that every disagreement in included studies would have inadvertently excluded a study (1% of 700 studies reviewed twice), we may have missed a maximum of 104 inclusions in the other studies not screened twice. In a systematic review with meta-analysis the consequence of missing or wrongly including a single study could have a large impact on the pooled effect estimate. However, as we did not focus on the outcome of individual studies but rather on the design quality of many studies, given the large number of studies included, it is unlikely that the missed studies would have changed the conclusions. And lastly, the quality indicator definition of a primary outcome was only based on what was described in the manuscript. It is possible that a primary outcome was defined retrospectively based on the observed data in a proportion of the studies.

Concerns about the methodological quality of antimicrobial stewardship studies have been raised before, with several publications making recommendations to improve their scientific methods [16,24,26]. However, have found no improvement in methodological quality over time except for more frequent inclusion of sample size calculation and clinical outcomes. Therefore, there is still a need for clear recommendations to improve antimicrobial stewardship design quality. Recommendations for improvement should especially consider: (1) emphasizing the importance of choosing appropriate clinical and microbiological outcomes, (2) focusing on robust methods to evaluate stewardship interventions in the hospital setting. Implementing these recommendations in future antimicrobial stewardship studies will help in the optimal use of resources to determine which stewardship interventions are most effective to change clinical practice. Building on the work from the systematic review we established a working group of expert investigators in this field. This systematic review identifies the limitations in design features that are most important for validity and translation into clinical practice. The results will be used to formulate recommendations in a white paper that will support investigators with key design decisions, support funders assessing proposals for stewardship studies and enhance the quality and impact of research in this crucial area.

SUPPORT STATEMENT

The presented work was supported by a grant (JPIAMRWG-010) from the Joint Programming Initiative on Antimicrobial Resistance.

ACKNOWLEDGMENTS

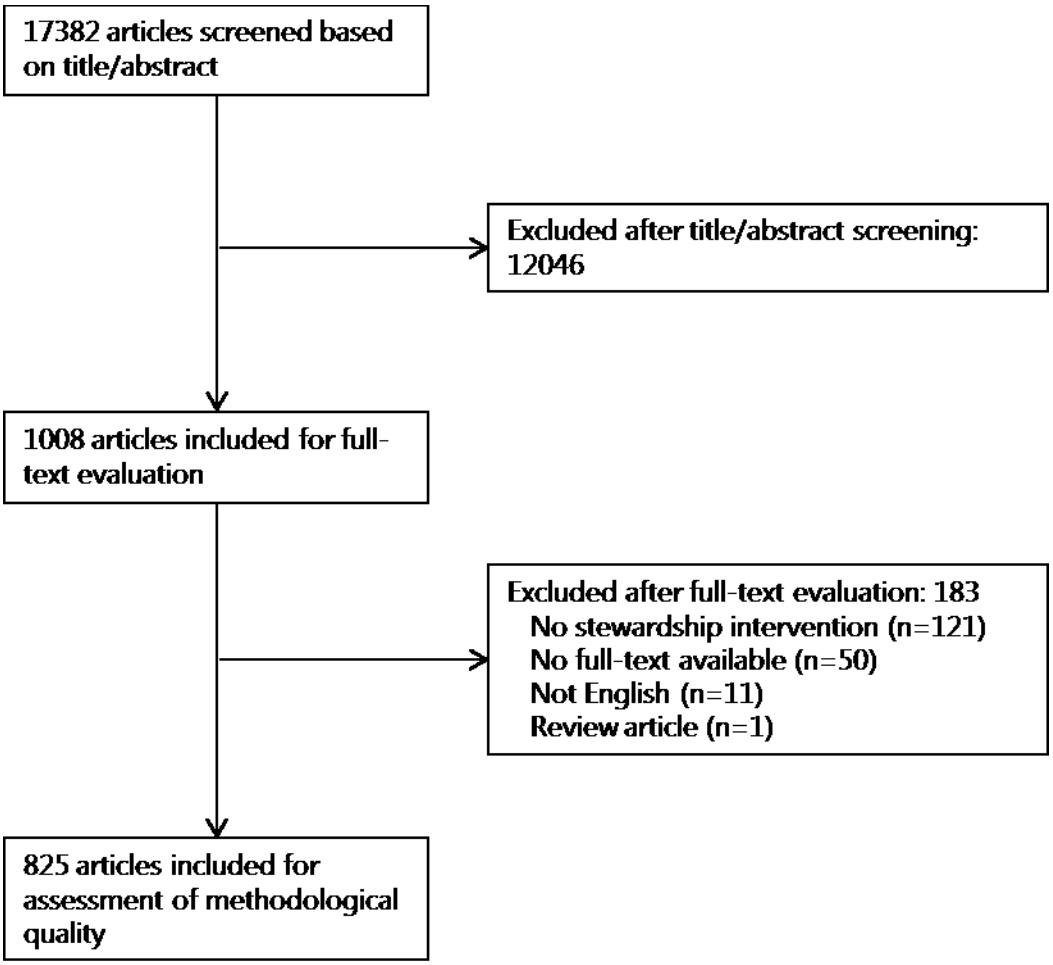
Other members of the Consensus on Antimicrobial Stewardship Evaluations (CASE) study group: S.Harbarth, B. Huttner, P. Little, J. Rodriguez-Baño, A. Savoldi, M. van Smeden, E. Tacconelli, J.F. Timsit, M. Wolkewitz.

Part of these data were presented at the 28th ECCMID 2018 in Madrid, Spain, 21-24 of April 2018 (#O0530).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

338 Figure 1. Flowchart of the studies included in the systematic literature review



339

340

341

342 Table 1. Characteristics of included studies stratified by studies performed in the community and the
343 hospital setting

Study characteristics	Community (n=205) n (%)	Hospital (n=620) n (%)
Number of patients included (median, IQR)	1255 (278-11230)	423 (186-1398)
Number of centres involved (median, IQR)	27 (8-90)	1 (1-1)
Age		
Adults	64 (31)	285 (46)
Children	33 (16)	68 (11)
Both	108 (53)	267 (43)
Specific disease targeted		
No specific disease targeted	63 (31)	342 (55)
Upper respiratory tract infections	106 (52)	25 (4)
Lower respiratory tract infections	57 (28)	70 (11)
Bacteraemia	0 (0)	42 (7)
Urinary tract infections	18 (9)	27 (4)
Prophylaxis	2 (1)	39 (6)
Sepsis	0 (0)	17 (3)
Skin and soft tissue infections	1 (1)	8 (1)
Abdominal infections	3 (1)	10 (2)
Other	16 (8)	62 (10)
Antibiotic class targeted		
No specific antibiotic class targeted	177 (86)	427 (69)
Cephalosporins	4 (2)	66 (11)
Fluoroquinolones	10 (5)	54 (9)
Carbapenems	0 (0)	48 (8)
Vancomycin	0 (0)	31 (5)
Aminoglycosides	0 (0)	28 (5)
Penicillins	10 (5)	15 (2)
Macrolides	8 (4)	5 (1)
Other	17 (8)	89 (14)
Antimicrobial stewardship interventions		
Education	149 (73)	260 (42)
Audit and feedback – individual patient	12 (6)	245 (40)
Audit and feedback – periodic	59 (29)	82 (13)

Restriction	13 (6)	113 (18)
Clinical decision support	31 (15)	112 (18)
Rapid diagnostic testing	24 (12)	68 (11)
Therapeutic drug monitoring	0 (0)	15 (2)
Guideline implementation	20 (10)	95 (15)
Delayed prescribing	9 (4)	0 (0)
Other	42 (20)	78 (13)
Research designs		
Before-after study	48 (23)	365 (59)
Before-after study (ITS*)	22 (11)	82 (13)
Cohort without control group	6 (3)	75 (12)
Controlled before-after study	22 (11)	19 (3)
Randomised controlled trial (RCT)	28 (14)	40 (6)
Parallel cluster randomised trial	30 (15)	6 (1)
Parallel cluster randomised trial with baseline period	26 (13)	2 (1)
Controlled before-after study (ITS)	7 (3)	12 (2)
Non-randomised parallel cluster study	5 (2)	8 (1)
Factorial cluster randomised trial	7 (3)	0 (0)
Cluster randomised cross-over trial	2 (1)	3 (1)
Stepped wedge cluster randomised trial	0 (0)	1 (1)
Non-randomised cluster cross-over study	0 (0)	3 (1)
Factorial randomised controlled trial (RCT)	1 (1)	2 (1)
Non-randomised stepped wedge study	0 (0)	2 (1)
Adaptive RCT	1 (1)	0 (0)

344 IQR: interquartile range, *ITS: interrupted time series

345 Table 2. Outcomes reported in the included antimicrobial stewardship studies stratified by studies
346 performed in the community and the hospital setting

Process measure outcomes	**Community (n=205) n (%)**	**Hospital (n=620) n (%)**
Costs/cost-effectiveness	36 (18)	200 (32)
Appropriateness	34 (17)	178 (29)
Defined daily doses (DDD)	34 (17)	156 (25)
Proportion treated with antibiotics	121 (59)	131 (21)
Recommendation acceptance	6 (3)	114 (18)
Guideline adherence	27 (13)	100 (16)
Duration of treatment	5 (2)	93 (15)
Days on therapy (DOT)	6 (3)	62 (10)
Time to appropriate therapy	3 (1)	71 (11)
Antibiotic knowledge	17 (8)	14 (2)
None	2 (1)	5 (1)
Other	58 (28)	166 (27)
Clinical outcome measures		
None	144 (70)	283 (46)
Mortality	9 (4)	203 (33)
Length of stay	5 (2)	201 (32)
Infection	10 (5)	75 (12)
Hospital readmission	0 (0)	76 (12)
Adverse effects	5 (2)	52 (8)
Revisits	22 (11)	0 (0)
Clinical cure	12 (6)	27 (4)
Intensive care unit admission	0 (0)	24 (4)
Hospital admission	9 (4)	0 (0)
Time to clinical stability	3 (1)	6 (1)
Other	16 (8)	49 (8)
Microbiological outcome measures		
None	188 (92)	447 (72)
Colonization/infection resistant pathogens	17 (8)	146 (24)
Clostridium difficile infections	5 (2)	62 (10)
Other	2 (1)	12 (2)

347

348 Table 3. Design quality features of the included studies stratified by studies performed in the community

349 and the hospital setting

Quality feature	**Community (n=205) n (%)**	**Hospital (n=620) n (%)**
Randomised research design	95 (46)	55 (9)
External control group	129 (63)	99 (16)
Multicentre	148 (72)	101 (16)
Sample size calculation reported	77 (38)	96 (15)
Prospective data collection	144 (70)	288 (46)
Correction for confounding factors	113 (55)	157 (25)
Primary outcome defined	116 (57)	272 (44)
Clinical outcome reported	61 (30)	337 (54)
Microbiological outcome reported	17 (8)	173 (28)
Sustainability assessed (≥12 months)	115 (56)	347 (56)

350

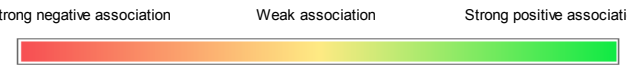
351

352

Table 4. Results of stepwise backward selection of multivariable model containing all the factors with the different design quality indicators as outcome. The colours indicate either a strong negative association (OR<1.0) in red, or a strong positive association (OR>1.0) in green. Numbers indicate odds ratio's with 95% confidence intervals

		Design quality indicators									
		Model 1 Randomised design	**Model 2** External control	**Model 3** Sample size calculation	**Model 4** Prospective data	**Model 5** Confounding correction	**Model 6** Primary outcome	**Model 7** Clinical outcome	**Model 8** Microbiological outcome	**Model 9** Multicentre	**Model 10** Sustainability assessed
Factors											
Clinical setting	Community (n=205)	5.9 (3.8 - 9.2)	5.6 (3.6 - 8.5)	2.9 (1.9 - 4.5)	1.7 (1.2 - 2.6)	3.0 (2.1 - 4.3)	1.4 (1.0 - 2.0)	0.3 (0.2 - 0.5)	0.2 (0.1 - 0.4)	10.5 (7.1 - 15.7)	
Financial support	Yes (n=385)	4.7 (2.9 - 7.5)	4.1 (2.6 - 6.3)	3.1 (2.1 - 4.7)	2.2 (1.6 - 3.0)	2.5 (1.8 - 3.4)	1.8 (1.4 - 2.5)	1.9 (1.4 - 2.6)		3.0 (2.0 - 4.3)	
Age group studied**	Children (n=101)	1.1 (0.6 - 2.2)	0.8 (0.4 - 1.6)	0.8 (0.5 - 1.4)	1.1 (0.7 - 1.9)		1.0 (0.6 - 1.6)	0.6 (0.4 - 1.0)			1.4 (0.9 - 2.2)
	Both (n=375)	0.6 (0.4 - 1.0)	0.6 (0.4 - 0.9)	0.4 (0.2 - 0.6)	0.5 (0.4 - 0.7)		0.5 (0.4 - 0.7)	0.4 (0.3 - 0.5)			1.6 (1.2 - 2.2)
Year study conducted***	2005-2010 (n=160)	1.0 (0.5 - 1.9)	1.1 (0.6 - 1.9)	2.5 (1.3 - 4.8)	0.4 (0.2 - 0.6)	0.6 (0.4 - 1.0)		0.9 (0.6 - 1.5)	2.1 (1.2 - 3.7)		1.1 (0.7 - 1.8)
	2011-2013 (n=156)	0.5 (0.3 - 1.1)	0.5 (0.2 - 0.9)	1.8 (0.9 - 3.5)	0.2 (0.1 - 0.3)	0.7 (0.4 - 1.2)		1.0 (0.6 - 1.6)	1.7 (1.0 - 3.0)		1.6 (1.0 - 2.5)
	2014-2015 (n=173)	0.4 (0.2 - 0.8)	0.5 (0.2 - 0.9)	2.5 (1.3 - 4.9)	0.1 (0.1 - 0.2)	1.1 (0.7 - 1.8)		1.3 (0.8 - 2.1)	1.2 (0.7 - 2.1)		1.8 (1.1 - 2.8)
	2016-2017 (n=176)	0.8 (0.4 - 1.5)	0.9 (0.5 - 1.7)	2.5 (1.3 - 4.9)	0.1 (0.1 - 0.2)	1.2 (0.7 - 1.9)		1.7 (1.0 - 2.7)	1.4 (0.8 - 2.4)		1.2 (0.8 - 1.8)
Geographical region****	Europe (n=260)	2.5 (1.5 - 4.2)	2.2 (1.4 - 3.5)	1.6 (1.1 - 2.5)	2.8 (1.9 - 4.1)		1.1 (0.8 - 1.5)	1.1 (0.8 - 1.6)	0.9 (0.6 - 1.4)	2.0 (1.3 - 3.1)	
	Asia (n=128)	2.6 (1.3 - 5.0)	2.4 (1.3 - 4.4)	1.1 (0.6 - 2.0)	1.7 (1.1 - 2.7)		0.4 (0.3 - 0.7)	1.8 (1.1 - 2.8)	1.2 (0.8 - 2.0)	1.0 (0.6 - 1.8)	
	Oceania (n=36)	1.6 (0.6 - 4.2)	1.5 (0.6 - 3.7)	0.5 (0.1 - 1.4)	1.8 (0.8 - 3.9)		0.5 (0.2 - 1.0)	0.5 (0.2 - 1.0)	0.2 (0.1 - 0.8)	0.9 (0.4 - 2.2)	
	Africa (n=29)	4.0 (1.1 - 14.5)	3.3 (0.9 - 11.8)	1.8 (0.5 - 5.8)	2.0 (0.6 - 6.2)		0.9 (0.3 - 2.5)	1.8 (0.6 - 5.3)	0.2 (0.1 - 1.7)	7.2 (2.0 - 25.1)	
	South America (n=17)	1.8 (0.4 - 7.4)	2.3 (0.6 - 8.4)	2.0 (0.6 - 6.5)	2.2 (0.8 - 6.4)		1.1 (0.4 - 2.9)	1.1 (0.4 - 3.0)	1.3 (0.4 - 3.8)		

*Reference category: studies performed in the hospital setting (n=620), **Reference category: studies performed in adults (n=349), ***Reference category: studies performed before 2005 (n=160), ****Reference category: studies performed in North America (n=368)



REFERENCES

[1] O'Neill J. Review on Antimicrobial Resistance. Tackling Drug-Resistant Infect Glob Final Rep Recomm 2016. https://amr-review.org/sites/default/files/160518_Final paper_with cover.pdf.

[2] Goossens H, Ferech M, Vander Stichele R, Elseviers M, ESAC Project Group H, Sprenger M, et al. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. Lancet (London, England) 1998;365:579–87. doi:10.1016/S0140-6736(05)17907-0.

[3] 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. doi:10.1136/bmj.c2096.

[4] Dyar OJ, Huttner B, Schouten J, Pulcini C, ESGAP (ESCMID Study Group for Antimicrobial stewardshiP). What is antimicrobial stewardship? Clin Microbiol Infect 2017;0. doi:10.1016/j.cmi.2017.08.026.

[5] 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;2017. doi:10.1002/14651858.CD003543.pub4.

[6] Molina J, Cisneros JM. A chance to change the paradigm of outcome assessment of antimicrobial stewardship programs. Clin Infect Dis 2015;61:807–8. doi:10.1093/cid/civ496.

[7] Wagner B, Filice GA, Drekonja D, Greer N, MacDonald R, Rutks I, et al. Antimicrobial stewardship programs in inpatient hospital settings: a systematic review. Infect Control Hosp Epidemiol 2014;35:1209–28. doi:10.1086/678057.

[8] Schuts EC, Hulscher MEJL, Mouton JW, Verduin CM, Stuart JWTC, Overdiek HWPM, et al. Current evidence on hospital antimicrobial stewardship objectives: A systematic review and meta-analysis. Lancet Infect Dis 2016;16:847–56. doi:10.1016/S1473-3099(16)00065-7.

[9] Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback : effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012;6:Art. No.: CD000259. DOI: 10.1002/14651858.CD000259. doi:10.1002/14651858.CD000259.pub3.

[10] Charani E, Edwards R, Sevdalis N, Alexandrou B, Sibley E, Mullett D, et al. Behavior change strategies to influence antimicrobial prescribing in acute care: A systematic review. Clin Infect Dis 2011;53:651–62. doi:10.1093/cid/cir445.

[11] 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. doi:10.1093/jac/dkg460.

[12] Drekonja DM, Filice GA, Greer N, Olson A, MacDonald R, Rutks I, et al. Antimicrobial stewardship in outpatient settings: A systematic review. Infect Control Hosp Epidemiol 2015;36:142–52. doi:10.1017/ice.2014.41.

[13] Effective Practice and Organisation of Care. No Title. EPOC Resour Rev Authors 2015:http://epoc.cochrane.org/resources/epoc-resources-.

[14] Pulcini C, Huttner A. CMI policy on antimicrobial stewardship research. Clin Microbiol Infect 2017. doi:10.1016/j.cmi.2017.11.006.

[15] Moher D, Liberati A, Tetzlaff J ADPG. Preferred reporting items for systematic reviews and meta-

analyses: the PRISMA statement. Bmj 2009.

[16] McGowan JE. Antimicrobial Stewardship—the State of the Art in 2011 Focus on Outcome and Methods. Infect Control Hosp Epidemiol 2012;33:331–7. doi:10.1086/664755.

[17] Gillespie D. Use of co-primary outcomes for trials of antimicrobial stewardship interventions. n.d.

[18] Horikoshi Y, Higuchi H, Suwa J, Isogai M, Shoji T, Ito K. Impact of computerized pre-authorization of broad spectrum antibiotics in Pseudomonas aeruginosa at a children's hospital in Japan. J Infect Chemother 2016;22:532–5. doi:10.1016/j.jiac.2016.05.001.

[19] Hohn A, Heising B, Hertel S, Baumgarten G, Hochreiter M, Schroeder S. Antibiotic consumption after implementation of a procalcitonin-guided antimicrobial stewardship programme in surgical patients admitted to an intensive care unit: a retrospective before-and-after analysis. Infection 2015;43:405–12. doi:10.1007/s15010-014-0718-x.

[20] Vellinga A, Galvin S, Duane S, Callan A, Bennett K, Cormican M, et al. Intervention to improve the quality of antimicrobial prescribing for urinary tract infection: a cluster randomized trial. CMAJ 2016;188:108–15. doi:10.1503/cmaj.150601.

[21] Butler CC, Simpson SA, Dunstan F, Rollnick S, Cohen D, Gillespie D, et al. Effectiveness of multifaceted educational programme to reduce antibiotic dispensing in primary care: Practice based randomised controlled trial. BMJ 2012;344. doi:10.1136/bmj.d8173.

[22] Persell SD, Doctor JN, Friedberg MW, Meeker D, Friesema E, Cooper A, et al. Behavioral interventions to reduce inappropriate antibiotic prescribing: A randomized pilot trial. BMC Infect Dis 2016;16. doi:10.1186/s12879-016-1715-8.

[23] Baur D, Gladstone BP, Burkert F, Carrara E, Foschi F, Döbele S, et al. Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and Clostridium difficile infection: a systematic review and meta-analysis. Lancet Infect Dis 2017;17:990–1001. doi:10.1016/S1473-3099(17)30325-0.

[24] 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. doi:10.1016/j.cmi.2017.05.019.

[25] Pulcini C, Morel CM, Tacconelli E, Beovic B, de With K, Goossens H, et al. Human resources estimates and funding for antibiotic stewardship teams are urgently needed. Clin Microbiol Infect 2017;23:785–7. doi:10.1016/j.cmi.2017.07.013.

[26] McGregor JC, Furuno JP. Optimizing research methods used for the evaluation of antimicrobial stewardship programs. Clin Infect Dis 2014;59:S185–92. doi:10.1093/cid/ciu540.

[27] Cantey JB, Wozniak PS, Pruszynski JE, Sánchez PJ. Reducing unnecessary antibiotic use in the neonatal intensive care unit (SCOUT): a prospective interrupted time-series study. Lancet Infect Dis 2016;16:1178–84. doi:10.1016/S1473-3099(16)30205-5.

SUPPLEMENT

Table S1. Definition of the design quality determinants used and corresponding categorisation if applicable

Design quality indicator	**Definition**	**Categorisation (if applicable)**
Randomised research design	Allocation of the antimicrobial stewardship intervention or comparator was random. Randomisation is defined as control by means of random allocation at any level (individual or cluster).	Randomised: RCT, c-RCT, cx-RCT, sw-CRT Non-randomised: BA(with or without ITS analysis), cBA, cohort studies, non-randomised trials
External control group	The outcome was also assessed in an external control group without antimicrobial stewardship intervention. External indicates that a contemporaneous cluster or group was included in which the intervention under study is not implemented.	Parallel control group: cBA, RCT, non-randomised trials, c-RCT, cx-RCT, sw-CRT No parallel control group: BA, cohort studies
Number of centres	Amount of centres involved in the study, either as control or intervention.	Single centre: 0-1 centre Multicentre: >1 centre
Sustainability of the intervention sufficiently assessed	The duration of follow-up of the cluster after the intervention was introduced to assess the sustainability of the intervention	Yes: Duration of follow-up ≥12 months No: Duration of follow-up <12 months
Sample size calculation reported	A sample size calculation was performed to ensure sufficient power for the primary outcome.	-
Prospective data collection	The data was prospectively collected. If not reported we assumed the data collection to be retrospective.	-
Confounding correction	The intervention effect was corrected for confounding bias, either by randomisation, matching, stratification or correction.	-
Primary outcome	A primary outcome was clearly defined.	-

defined		
Clinical outcome reported	Any clinical outcome was reported. Clinical outcomes include mortality, length of stay, readmissions, revisits, etc.	-
Microbiological outcome reported	Any microbiological outcome was reported. Microbiological outcomes include CDI, colonisation or infection with antimicrobial resistant bacteria	-

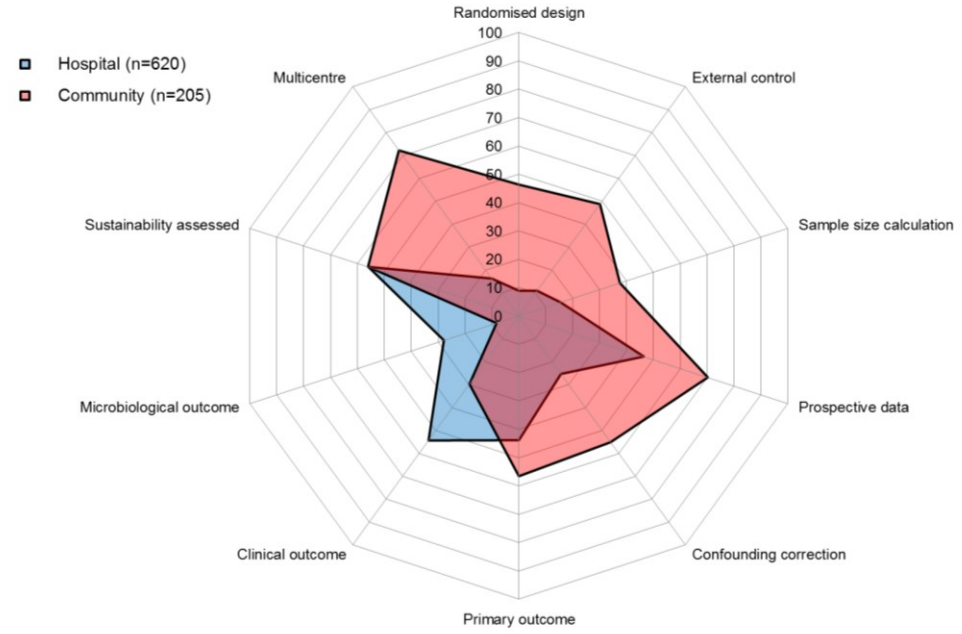
RCT: randomised controlled trial, c-RCT: parallel cluster randomised controlled trial, cx-RCT: cluster cross-over randomised controlled trial, sw-CRT: stepped wedge cluster randomised trial, BA: before-after study, cBA: controlled before-after study, ITS: interrupted time-series, CDI: *Clostridium difficile* infection

Table S2. Design quality features changes over time of the included studies stratified by studies performed in the community and the hospital setting

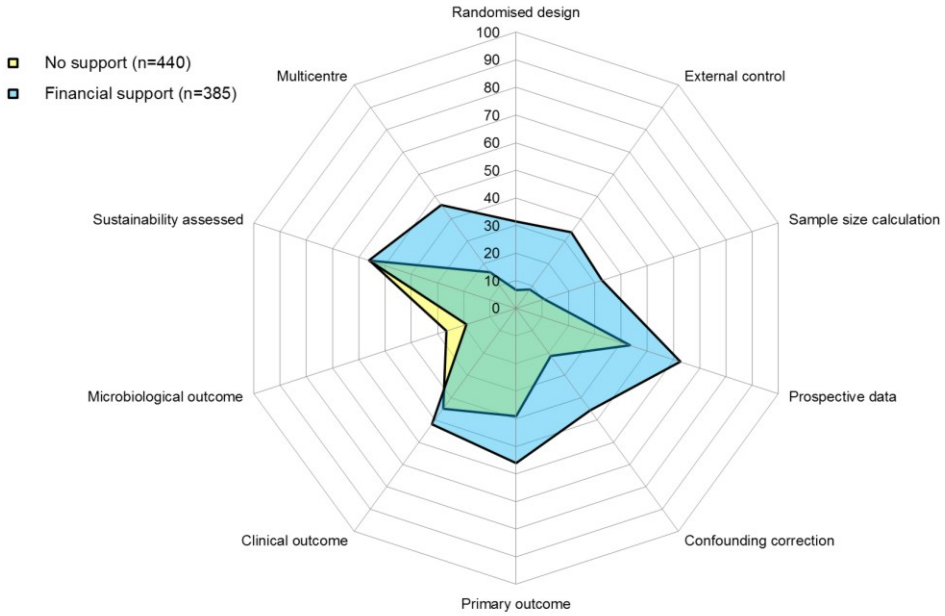
Hospital	**1977-2004**	**2005-2010**	**2011-2013**	**2014-2015**	**2016-2017**
Randomised design	10%	17%	6%	6%	8%
External control	12%	20%	6%	8%	11%
Sample size calculation	8%	21%	13%	18%	17%
Multicentre	9%	14%	18%	20%	18%
Prospective data	71%	68%	37%	36%	31%
Confounding correction	24%	22%	19%	29%	30%
Primary outcome	38%	50%	37%	45%	48%
Clinical outcome	43%	51%	50%	58%	66%
Microbiological outcome	21%	38%	32%	23%	27%
Sustainability assessed	53%	50%	55%	65%	54%
Community					
Randomised design	43%	47%	41%	41%	61%
External control	45%	51%	41%	45%	65%
Sample size calculation	19%	44%	35%	48%	52%
Multicentre	68%	75%	78%	69%	71%
Prospective data	89%	65%	57%	59%	74%
Confounding correction	58%	42%	54%	66%	65%
Primary outcome	53%	53%	59%	66%	58%
Clinical outcome	32%	29%	27%	31%	29%
Microbiological outcome	6%	5%	8%	17%	10%
Sustainability assessed	43%	56%	78%	55%	52%

Figure S1. Design quality indicators stratified by factors: (A) community versus hospital setting, (B) financial support versus no financial support, (C) age setting: children, adults or both, (D) old versus new studies, (E) geographical region

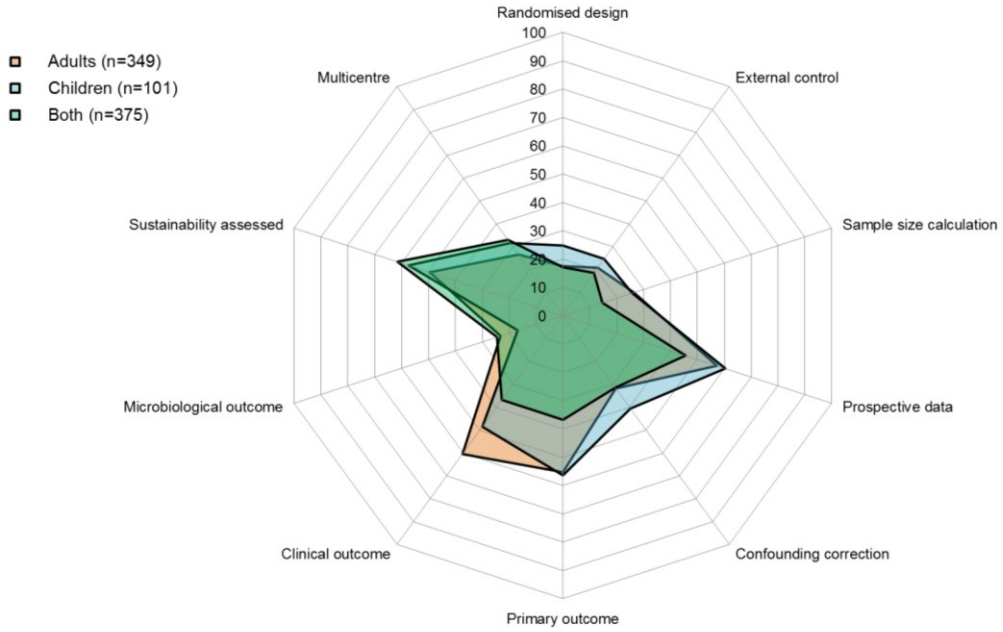
A



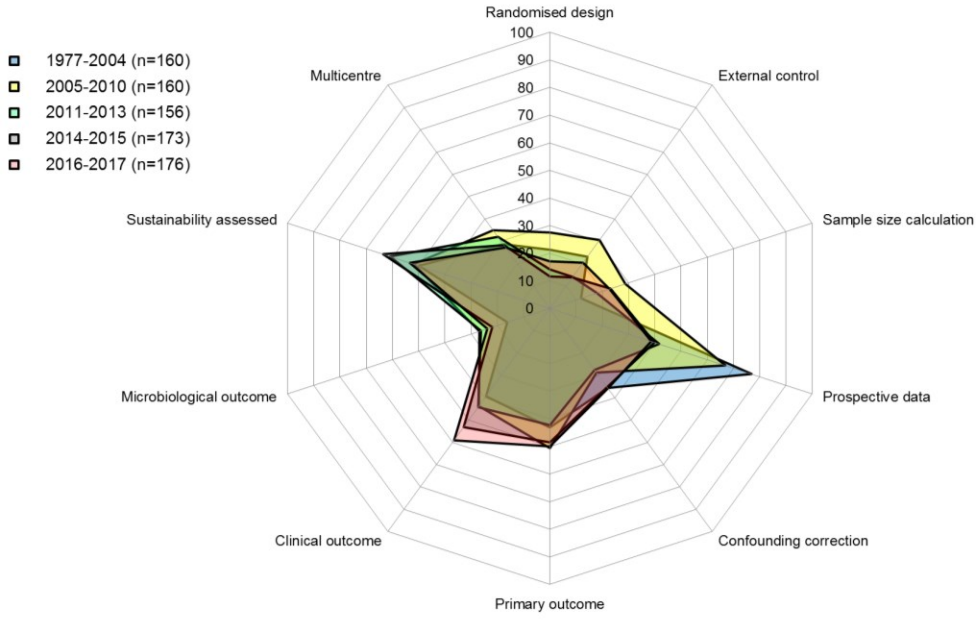
B



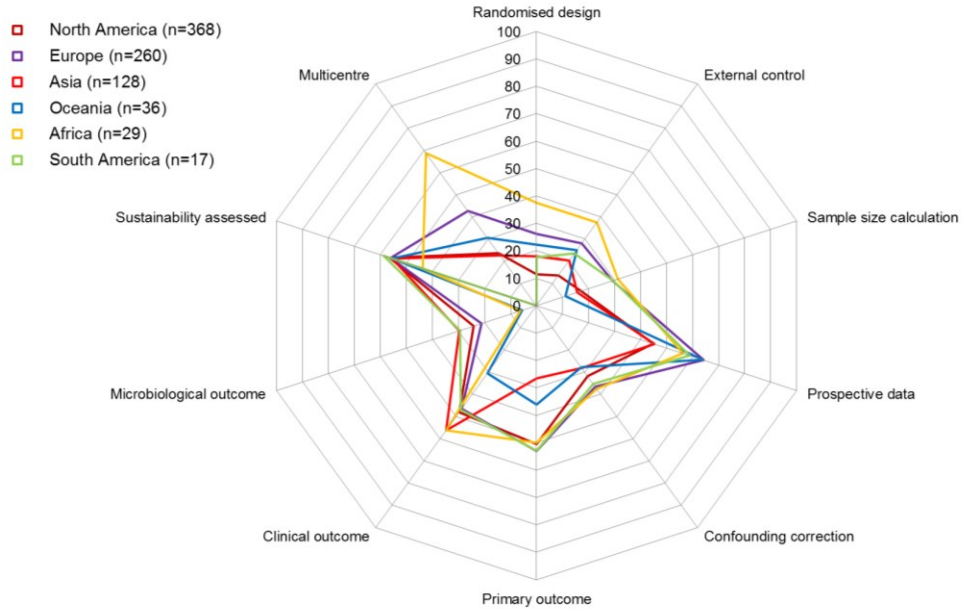
C


Adults (n=349)
Children (n=101)
Both (n=375)
Randomised design
External control
Sample size calculation
Prospective data
Confounding correction
Primary outcome
Clinical outcome
Microbiological outcome
Sustainability assessed
Multicentre
100
90
80
70
60
50
40
30
20
10
0

D


1977-2004 (n=160)
2005-2010 (n=160)
2011-2013 (n=156)
2014-2015 (n=173)
2016-2017 (n=176)
Randomised design
External control
Sample size calculation
Prospective data
Confounding correction
Primary outcome
Clinical outcome
Microbiological outcome
Sustainability assessed
Multicentre
100
90
80
70
60
50
40
30
20
10
0

E


North America (n=368)
Europe (n=260)
Asia (n=128)
Oceania (n=36)
Africa (n=29)
South America (n=17)
Randomised design
External control
Sample size calculation
Prospective data
Confounding correction
Primary outcome
Clinical outcome
Microbiological outcome
Sustainability assessed
Multicentre
100
90
80
70
60
50
40
30
20
10
0