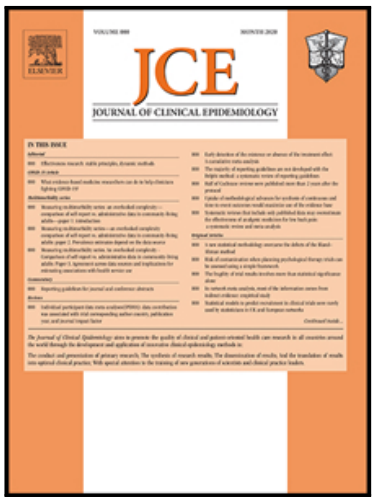


Sankey diagrams can clarify 'evidence attrition': a systematic review and meta-analysis of the effectiveness of rapid diagnostic tests for antimicrobial resistance

Rebecca E. Glover , Mustafa Al-Haboubi , Mark P. Petticrew , Elizabeth Eastmure , Sharon J Peacock , Nicholas Mays

PII: S0895-4356(21)00390-5
DOI: https://doi.org/10.1016/j.jclinepi.2021.11.032
Reference: JCE 10701

To appear in: *Journal of Clinical Epidemiology*

Accepted date: 22 November 2021

Please cite this article as: Rebecca E. Glover , Mustafa Al-Haboubi , Mark P. Petticrew , Elizabeth Eastmure , Sharon J Peacock , Nicholas Mays , Sankey diagrams can clarify 'evidence attrition': a systematic review and meta-analysis of the effectiveness of rapid diagnostic tests for antimicrobial resistance, *Journal of Clinical Epidemiology* (2021), doi: https://doi.org/10.1016/j.jclinepi.2021.11.032

Highlights

- We developed a novel method to identify, group, and analyse included studies in a systematic review using a Sankey diagram
- Sankey diagrams can help compare patterns of methodological quality and variation in outcomes within primary studies across sectors and topics within a review. They provide a visual way of identifying methodological concerns in the evidence included in systematic reviews.
- We demonstrated this technique in an area where systematic review and meta-analysis is underused, namely the clinical effectiveness of rapid diagnostic tests for bacterial identification and antibiotic susceptibility testing.
- This is the first systematic review and MA of the effectiveness of RDTs for bacterial identification and antibiotic susceptibility testing which shows that, despite their widespread use and claims about their value, they do not appear to be effective.
- While there appears to be evidence of reporting bias (publication bias, small study effects), the paucity of studies included in our systematic review means that Egger's test is underpowered so the influence of publication bias on the summary estimates is difficult to determine.

Sankey diagrams can clarify 'evidence attrition': a systematic review and meta-analysis of the effectiveness of rapid diagnostic tests for antimicrobial resistance

Rebecca E. Glover (ORCID ID 0000-0001-9150-9977), Mustafa Al-Haboubi, Mark P. Petticrew, Elizabeth Eastmure, Sharon J Peacock, Nicholas Mays (ORCID ID 0000-0001-9808-8466)
Affiliations:

REG, MAH, EE, and **NM** are members of the Policy Innovation and Evaluation Research Unit, in the Department of Health Services Research and Policy at the London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, WC1H 9SH, London, UK
MPP is a member the Policy Innovation and Evaluation Research Unit based in the Department of Public Health, Environments, and Society at the London School of Hygiene & Tropical Medicine, 15-17 Tavistock Place, WC1H 9SH, London, UK
SJP is a Professor of public health and microbiology at the University of Cambridge

*Corresponding author.

REG, Tel: +44 (0)2079272710, Rebecca.glover@lshtm.ac.uk

Abstract

Objectives

To demonstrate, using the example of a new systematic review of rapid diagnostic tests, how Sankey diagrams, alongside the PRISMA guidelines, can (i) facilitate reporting of the quality of the evidence base and (ii) help assess evidence syntheses when studies use heterogeneous outcomes.

Study design

Systematic review and meta-analysis of experimental and observational studies which included at least one prescribing or clinical outcome of RDTs in hospital in-patients. Sub-group analysis was used to assess heterogeneity in summary effect estimates. A Sankey diagram was then used to show the pattern and quality of evidenceon RDT outcomes.

Results

57 studies from 14 countries were included. The introduction of RDTs did not significantly reduce in-hospital mortality (RR 0.83, 95% CI 0.60 - 1.15) or length of stay (weighted mean difference = -0.36, 95% CI -1.67 to 0.96). There was high heterogeneity in outcomes.

Conclusion

There is no clear evidence that the routine use of RDTs for bacterial identification and antibiotic susceptibility testing improves clinical outcomes in hospital inpatients. Sankey diagrams may be a useful further way succinctly to present the pattern and quality of evidence in systematic reviews, especially when it is heterogeneous and not easily amenable to meta-analysis.

Key words

Antibiotic resistance; antimicrobial resistance; systematic review; meta-analysis; health technology appraisal; rapid diagnostic tests; Sankey diagrams

Introduction

The identification and synthesis of evidence on outcomes of interventions is a key step in systematic reviews, and a focus of methodological research in clinical epidemiology.(1) Selection – and selective reporting - of outcomes is also a major source of bias in primary studies and thus reviews, and can lead to overestimates of the effectiveness of interventions, and under-reporting of harms. It can also involve the reporting of outcomes that represent no clinical benefit to patients, and for this reason there is an increasing emphasis on the incorporation of patients' views into the development of outcome measures, as a way of ensuring the utility and credibility of trial findings: "Clinical trials are only as credible as their endpoints".(2) Guidance from the Cochrane Handbook is that reviewers should choose only outcomes that are critical or important to users of the review, such as patients, health professionals and policy makers, and outcome measures should be defined in advance.(3) In a mature field, where there are many trials reporting on direct patient benefit, this often involves selecting and synthesising evidence on a narrow set of outcomes. However, in fields where new technologies are rapidly emerging, it may be more useful to incorporate a wider range of outcomes, to help assess the claims being made about the balance of costs and benefits of the intervention, and to help make judgements (sometimes in the absence of patient-level outcomes) about the potential effects of the intervention, drawing on evidence from different parts of the care pathway.

Synthesising and reporting on a heterogeneous and complex set of outcomes is challenging, however. Common approaches used in systematic reviews such as summary tables and forest plots often do not make full use of the data – for example, they cannot show clearly how different studies contribute to understanding how interventions work at different points along the care pathway. This is particularly the case for diagnostic tests related to antimicrobial resistance (AMR). Diagnostic test accuracy, but not clinical effectiveness, has often been

used to justify the routine use of these tests.(4–8) This is because it is difficult to perform meta-analysis using the evidence on diagnostic tests for three reasons: its relative paucity;(9) different proprietary technologies with different functions in the bacteriology care pathway; and different outcomes measured in each study.

In the absence of a previous evidence synthesis, we undertook a systematic review and meta-analysis of the evidence on whether RDTs for bacterial identification and antibiotic susceptibility testing confer clinical advantages over standard tests. We were aware in advance that the available evidence was likely to be heterogeneous and difficult-to-interpret, covering different RDT technologies, and using different definitions of the same outcome. For example, some tests have been described as "rapid" when they take 14 hours, while others are considered rapid when they take 15 minutes.

We were also aware of a proliferation of different outcomes in studies, which may in itself be a reason why no previous systematic review exists. We therefore used a Sankey diagram as a way of presenting the current state of the evidence on RDTs in AMR and to show how much of the evidence can be robustly synthesised.(10) Sankey diagrams represent flows (e.g. flows of information, or of any property) within a process, in this case the review process. They are frequently used in industrial processes and in engineering.(10) The overall aim of this paper, then, is to demonstrate, using the example of a new systematic review of RDTs, how Sankey diagrams, alongside the PRISMA guidelines, can (i) facilitate reporting of the quality of the evidence base and (ii) help assess evidence syntheses when studies use heterogeneous outcomes.

Methods

(i) *The systematic review:* We conducted a comprehensive systematic review and meta-analysis of the outcomes of introducing rapid molecular diagnostic tests for bacterial identification and antibiotic susceptibility testing, following PRISMA guidelines.(9) . The systematic review aimed to synthesise the evidence on effectiveness of RDTs in terms of clinical and prescribing outcomes compared with standard care in acute hospitals. The technologies included in the review are: multiplex, real-time, and quantitative polymerase chain reaction (PCR); matrix-assisted laser desorption ionisation time-of-flight mass spectrometers (MALDI-TOF MS); peptide nucleic acid florescent in situ hybridisation; and rapid procalcitonin testing. We registered our protocol on PROSPERO (CRD 42017060566) in 2017.

We searched (with no language restrictions) Ovid Medline [1950-2017], Ovid Embase [1947-2017], PubMed [1950-2017], Web of Science [1970-2017], Open Grey [1997-2017] and Cochrane CENTRAL [1997-2017]. (see Appendix 1). Our search was conducted in April 2017 and updated in April 2018. Two reviewers double-screened 20,592 titles, 1,445 abstracts and 319 full-text studies. We included 57 studies in our final analyses. The Kappa statistic for inter-rater reliability of inclusion and exclusion decisions was 0.6 (95% CI 0.553 to 0.648), indicating moderate agreement.(11) To deal with this moderate level of agreement, and to ensure that our review was as sensitive as possible, where reviewers differed in their inclusion criterion, we discussed the relevant title, abstract or full text article, and unless there was an explicit missed exclusion criterion we always erred on the side of inclusion.

Inclusion/exclusion criteria

Eligible participants were adults and children admitted to, and treated within, an acute hospital. The intervention of interest was the change in clinical or antibiotic prescribing outcomes that could plausibly be associated with an introduction of RDTs into the hospital. The comparator(s)/control was current hospital practice without RDT, defined as use of either a manual or automated culture system (Table 1). The primary clinical outcomes were length of stay (LOS) and mortality, and the primary antibiotic outcome was duration of antibiotic therapy. We allowed for the collection of any type of mortality outcome but made provision for separate (30-day and all-cause in-hospital) mortality meta-analyses. Secondary outcomes were: reported changes in antibiotic plan, time to treatment, and turnaround time. (Table 1) We extracted aggregate data from each included study on all outcomes of interest. We included both experimental and observational study designs, synthesised separately. Observational studies comprised prospective and retrospective cohort studies, quasi-experimental studies and interrupted time series analyses. Risk of bias was assessed using the Effective Public Health Practice Project (EPHPP) toolkit for quantitative studies and is included in Table 1.(12)

All statistical analyses were conducted in Stata 15.1.(13) When medians and interquartile ranges were reported as effect estimates, we transposed these into means and standard deviations using the methods of Luo et al, and then conducted subgroup analyses to validate the methodology.(14) We grouped those RDTs that were intended to replace either manual or automated culture, thereby reducing analysis time in the laboratory.

The principal summary effect estimates (summary measures) that were calculated were length of stay (mean difference), in-hospital mortality (risk ratio) and 30-day mortality (risk ratio). Random effects meta-analysis was used due to the heterogeneous interventions and

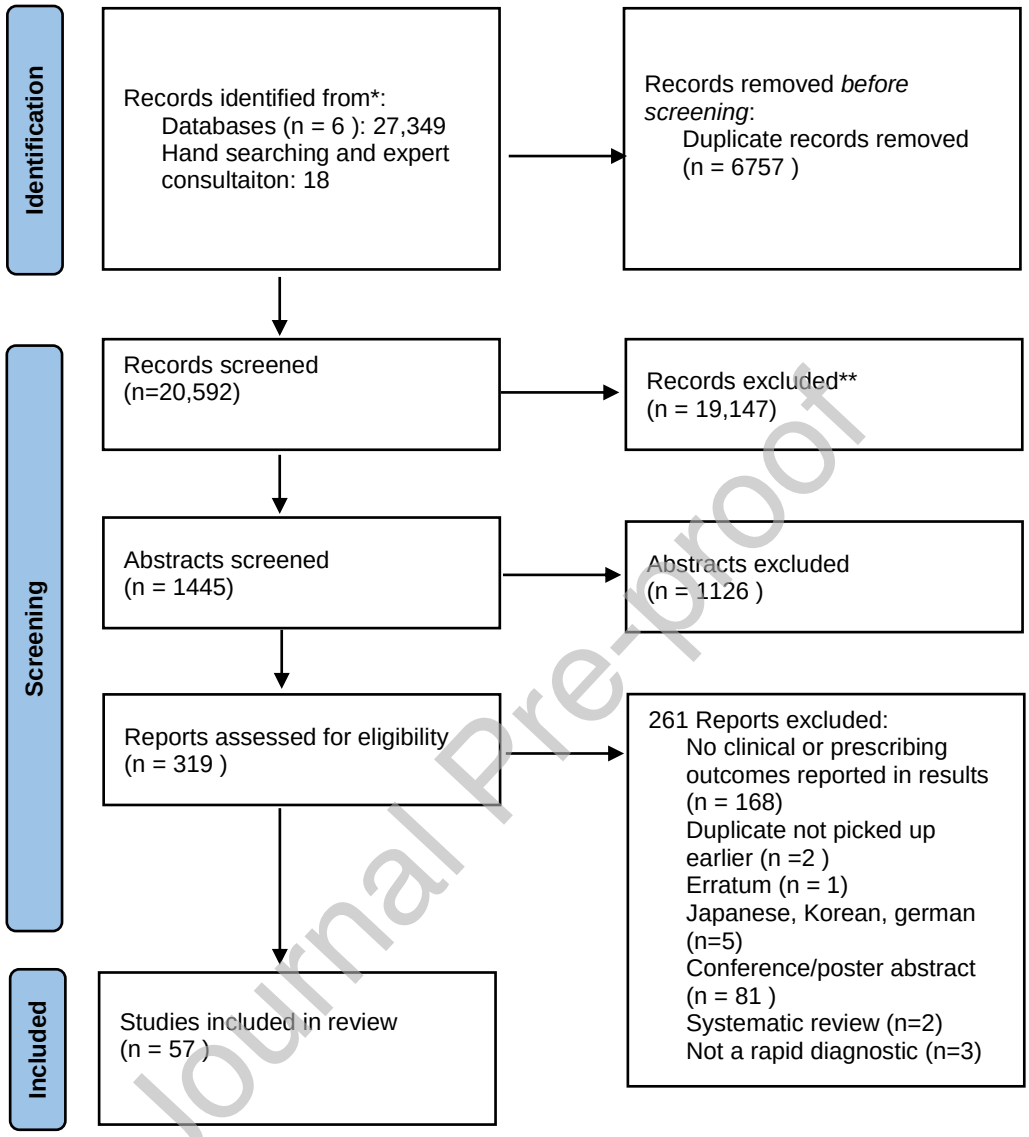
settings of each included study.(3,15) Not all studies that were included in the narrative synthesis were included in the meta-analysis (See Table 1). Higgins' I^2 was used to assess heterogeneity among outcomes in the meta-analyses.(15) Egger's test was not appropriate to conduct since there were small (n<10) numbers of studies in each subgroup analysis.(3,16)

(ii) *The Sankey diagram:* As there were many antibiotic stewardship outcomes of interest reported, but few studies reported the same outcomes of interest, we used a Sankey Diagram to show the outcomes of interest (number of papers included in the narrative synthesis), how those studies can be categorised into subgroups, the attrition on the review pathway from narrative synthesis to potential meta-analysis, and to provide methodological justification for the proportion of the overall evidence that is included in the final meta-analysis. Our Sankey diagram was constructed in the free, open source, online tool SankeyMATIC (BETA) (sankeymatic.com). The code for this tool is available on Github and builds on the open-source infographic design language D3. The tool allows users to: specify the number of flows (where flows are primary studies) in and out between nodes (which are stages or points in the synthesis process); and specify the number of nodes. Flows can transfer between nodes, as they have done in our Sankey diagram. In our Sankey diagram, the width of the arrows is proportionate to the number of outcomes of interest. On the left, separate arrows connote the types of outcomes, and on the right is a list of reasons for evidence attrition or small numbers of studies in the meta-analysis.

Results

There were 57 studies included in the final review. The study selection process is summarised in Figure 1. The included studies are summarised in Table 1 and fully described in Appendix A, online.

Figure 1. PRISMA diagram



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

Table 1. Sample of included studies and extracted characteristics in the narrative systematic review and meta-analysis: full table online

Author (year)	Study Design	Test	Comparator	Patients tested using RDTs	Patients tested using conventional treatment	LOS	Mortality	Reason for exclusion from MA	EPHPP rating (weak/moderate/strong evidence)
Allaouchiche *et al.* (1999) *France*	Randomised Controlled Trial	Multiplex PCR assay	Conventional lab procedures	72	72	✓		Patients in LOS analysis were subdivided by specific genes (oxa-S positive)	moderate
Banerjee *et al.* (2015) *USA*	Three arm-randomised controlled trial	FilmArray Blood Culture ID Panel (rapid multiplex PCR)	Control group: Standard BCB processing	198	207	✓	✓	*NA*	moderate
Bouadma *et al.* (2010) *France*	Multicentre Randomised Controlled trial	Procalcitonin	International and local guidelines for AB treatment	307	314		✓	28-day and 60-day mortality reported	moderate
Cambau *et al.* (2017) *France*	Cluster-randomised crossover trial	LightCycler® SeptiFast	Conventional (standard) work-up	731	685		✓	Patients with "severe sepsis", febrile neutropenia, or suspicion of F11IE; 7-day mortality reported.	moderate
Creamer *et al.* (2010) *Ireland*	Non-randomised clinical trial	Xpert® MRSA assay	direct culture on chromogenic agar plates	349	60			Isolation and turnaround time reported as outcomes	moderate
Cattoir *et al.* (2011) *France*	Controlled trial (non-randomised)	LightCycler® System	Standard phenotypic method	122	128			Favourable and unfavourable outcomes at 12-weeks follow-up reported.	moderate
de Jong *et al.* (2016) *Netherlands*	Randomised Controlled Trial	Procalcitonin-guided antibiotic treatment	Standard of care group	761	785		✓	28-day and 1-year mortality reported.	Moderate
Idelevich *et al.*	Randomised	LightCycler®	VITEK 2	74	76	✓	✓	Febrile neutropoe	moderate

(2015) *Germany*	Controlled Trial	SeptiFast Test MGrade assay						nic patients.	

Of the 57 included studies, 13 met the criteria for inclusion in a meta-analysis of length of stay, eight for meta-analysis of 30-day mortality, and seven for meta-analysis of in-hospital all-cause mortality. There were 30 antibiotic stewardship outcomes reported in 17 studies, but the lack of overlap of reported outcomes among studies made meta-analysis of these outcomes impossible.

Patients whose tests were undertaken using RDTs stayed in hospital an average of 0.36 (95% CI -1.67, 0.96, n.s.) days less than patients whose samples were processed using conventional methods in experimental studies, and 2.52 fewer days than patients whose samples were processed using conventional methods in the observational studies (95% CI -3.88 to -1.17). This can be seen in Figure 2. We conducted separate meta-analyses for experimental and observational studies. There was no significant heterogeneity among the RCTs (I^2= 0%, p=0.532) and moderate heterogeneity among the observational studies (I^2= 37.9%, p=0.106). (17)

Figure 2 Meta-analysis of studies reporting length of stay

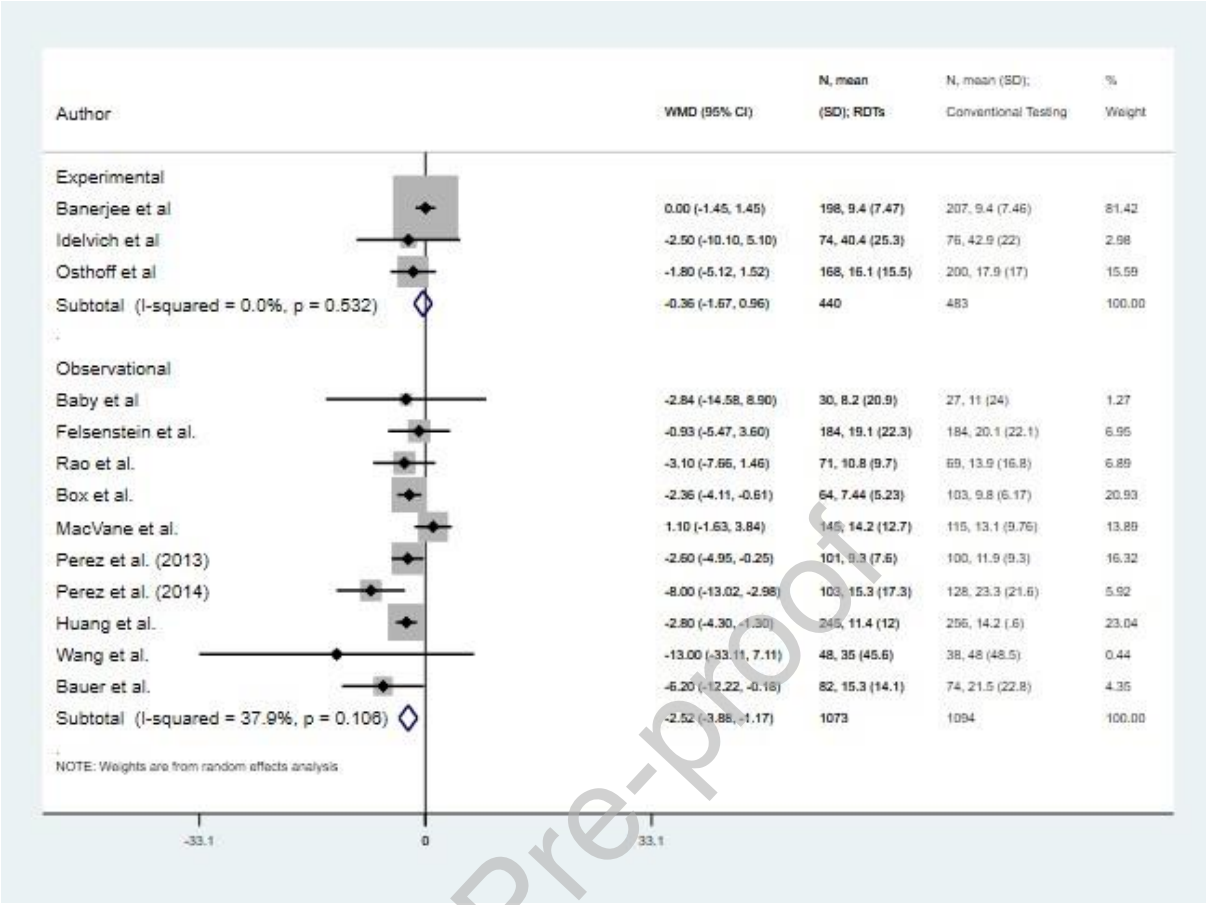


Figure 3 Meta-analysis of studies reporting 30-day mortality

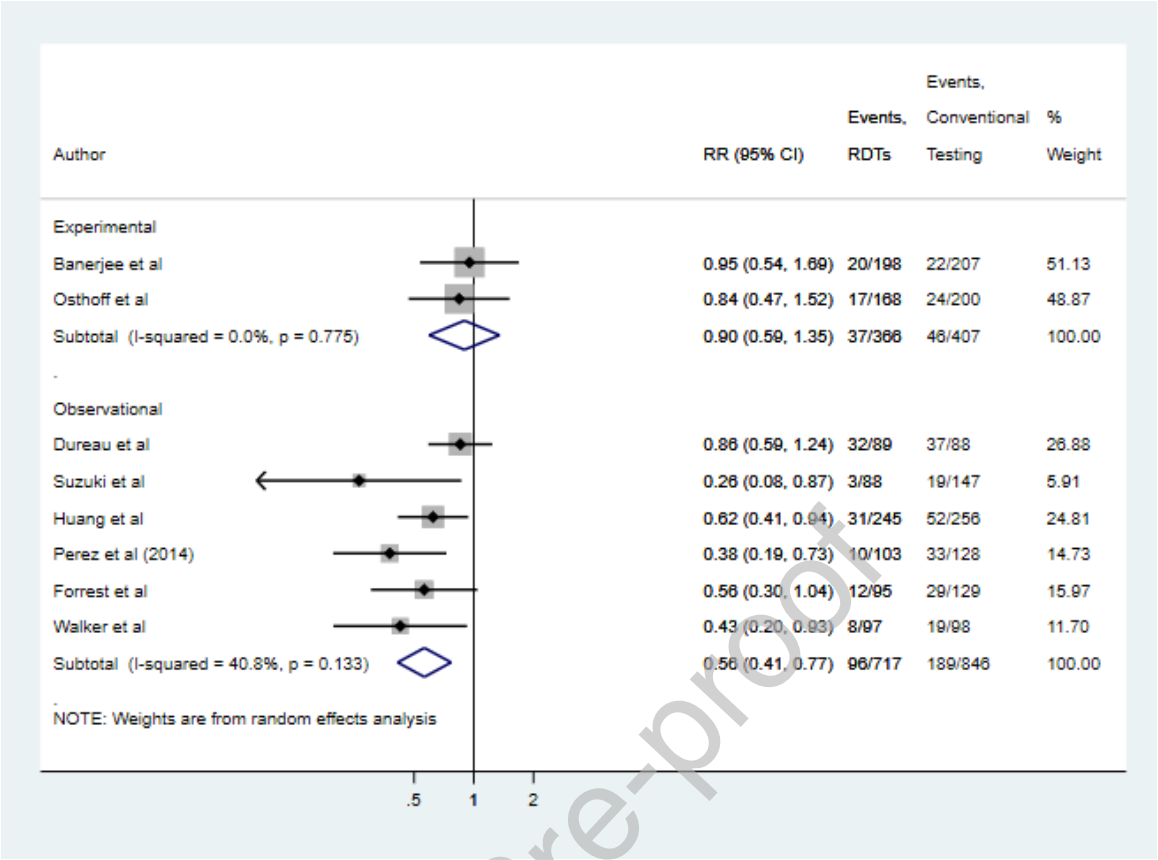
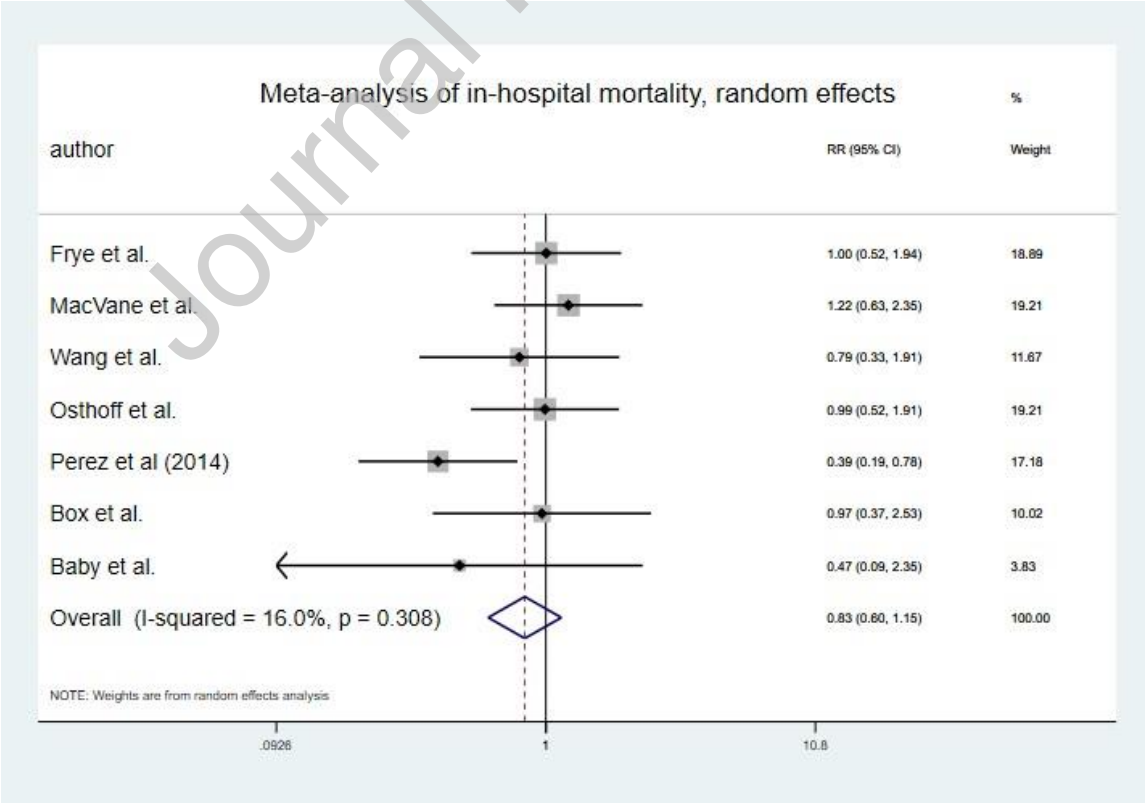


Figure 4 Meta-analysis of studies reporting in-hospital mortality



While 18 studies reported mortality measures, only eight reported 30-day mortality (Figure 3)

and seven reported all-cause in-hospital mortality (Figure 4). The overall risk ratio for 30-day mortality was 0.90 (95% CI 0.59-1.35) for experimental studies, and 0.59 (95% CI 0.41-0.77) for the observational studies. Among the experimental studies, there was no significant difference in 30-day mortality between RDTs and conventional methods. By contrast, there was a strong reduction in mortality in the observational studies, although, as with the length of stay analysis, many observational studies included ASPs in their post-test timeframes, something that the RCTs controlled for by either not including them or by including a third-arm in the trial.

The random effects summary estimate of the effect of RDTs on in-hospital mortality was 0.83 (95% confidence interval 0.60 to 1.15; n.s.). When these seven studies were combined for random effects meta-analysis, heterogeneity was low (X^2=7.14) and the variation in the risk ratio attributable to heterogeneity was also low (I^2=16.0%, p=0.308).

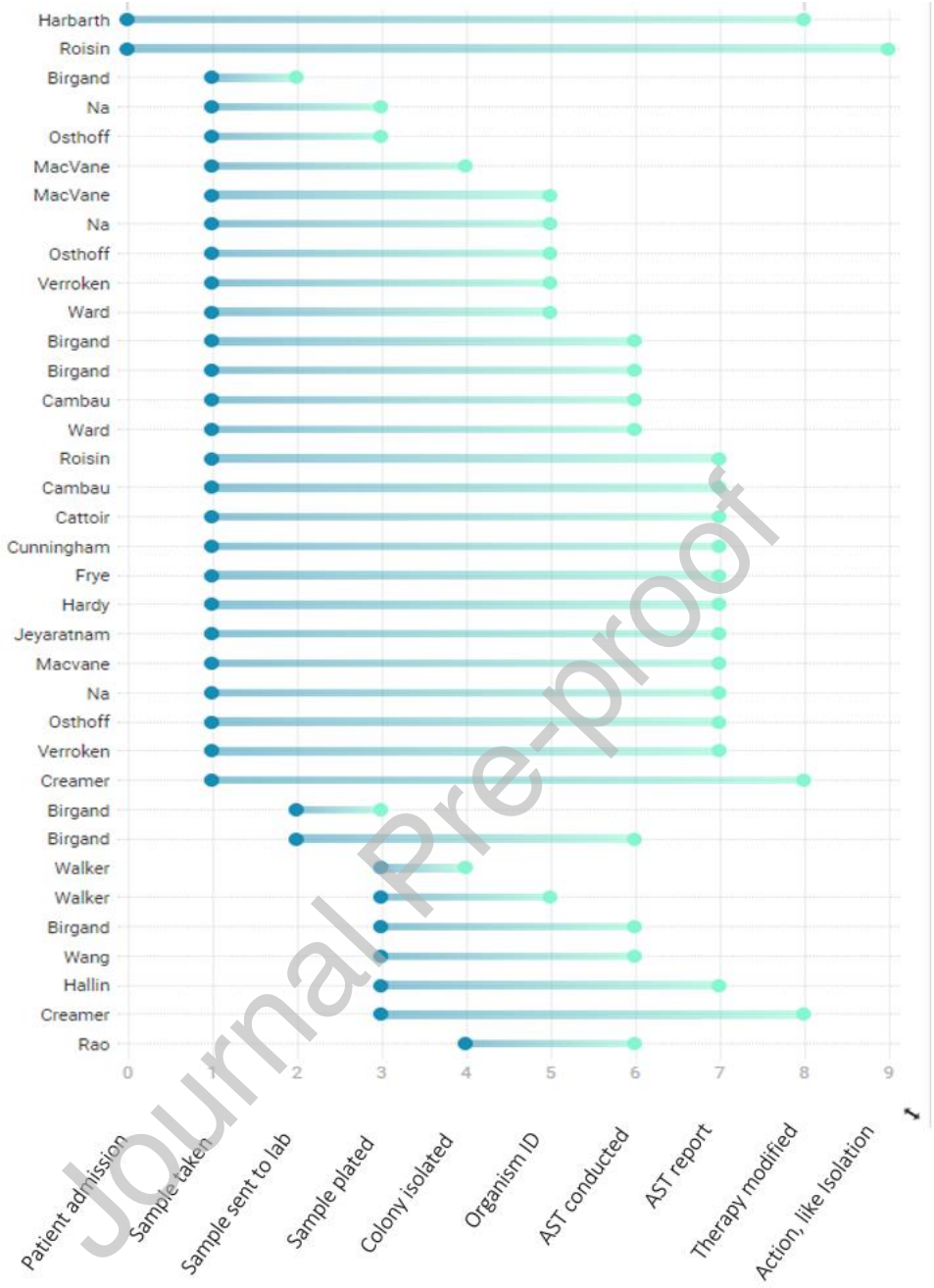
In 17 studies, there were 30 different antibiotic stewardship outcomes included, such as ‘time to first appropriate (de)escalation’, ‘prevention of unnecessary vancomycin’, ‘time from positive result to isolation precautions’, ‘appropriate antibiotic therapy for bacteraemic patients’. Many differences were reported as being statistically significant but no meta-analysis was possible due to the high degree of heterogeneity. A summary of these outcomes is included in supplementary file 1.

Given the small numbers of included studies, there were few opportunities for subgroup analysis. However, we were able to assess the impact of study characteristics on the length of stay summary effect estimates in two ways: by comparing summary effect sizes in moderate and lower quality studies; and by assessing the impact of the statistical transformation of the

reported length of stay from median and range, to mean and standard deviation. In neither case did the subgroup effect estimates differ statistically from the aggregate effect estimates.

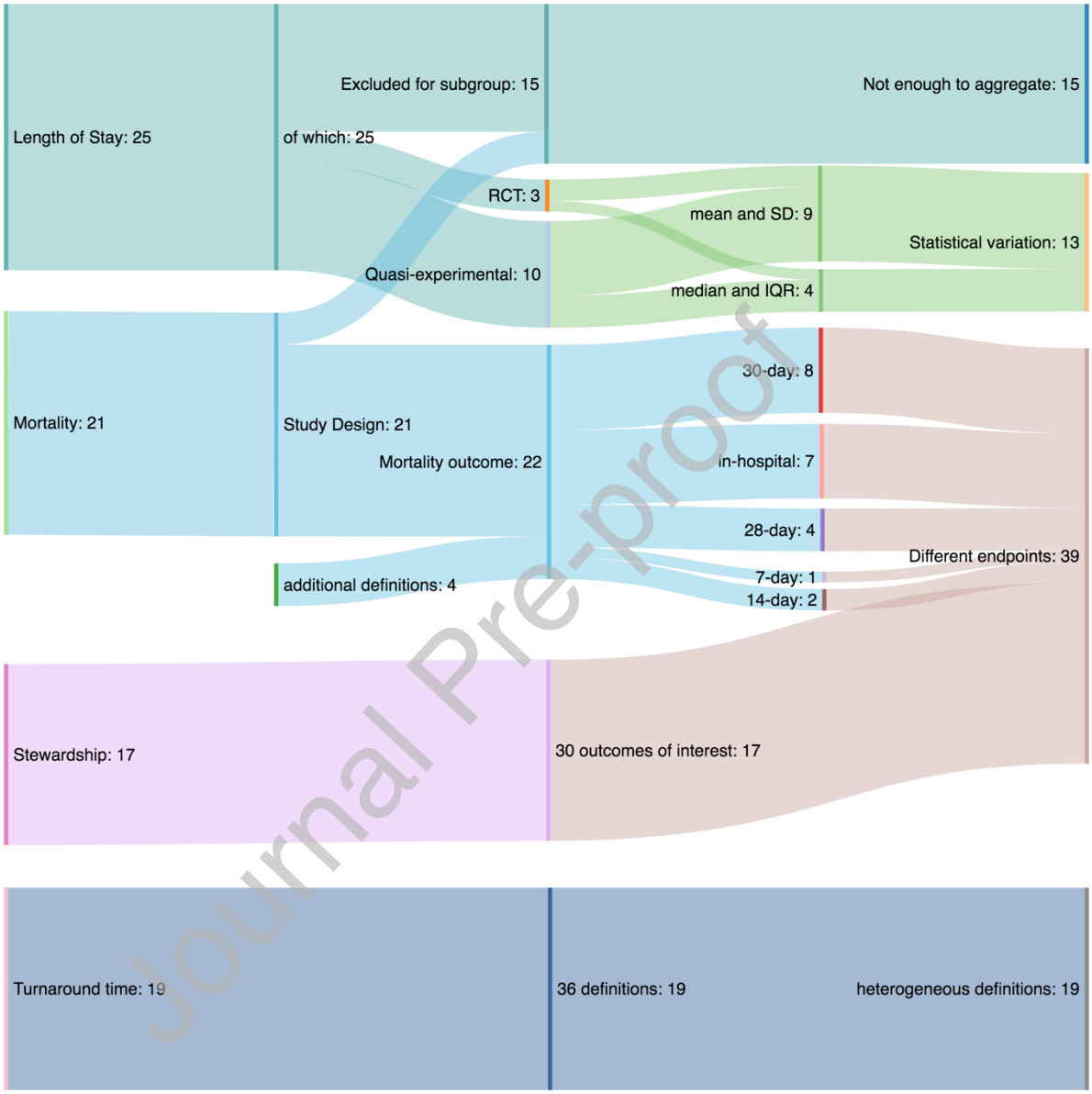
The definitions of ‘turnaround times’, ‘reporting times’ and ‘time to result’, which are the most frequently cited improvements attributed to RDTs, overlapped and varied enormously (See Figure 5). While the stylised pathway in Figure 5 neither captures the nuances of the entire care pathway, nor indicates that some activities can be undertaken concurrently, we validated it with a consultant clinical microbiologist, who judged it to be an appropriate general description of the key steps in the process. The most commonly reported (11/36) timed pathway segment was from “sample-to-report”. Many studies reported on multiple slices of time in the care pathway, however, only one study reported on the effect of RDT use from patient admission through to isolation (see Figure 5).(18) One further study reported on the effect of RDT use from patient admission through to the clinician’s receipt of an antibiotic susceptibility test report AST (and consequent ability to modify therapy, if appropriate).(19)

Figure 5 Bacteriological care pathway mapped to definitions of turnaround time and time-to-result. Where 0-9 represent a simplified bacteriological care pathway, as annotated below



(ii) The use of the Sankey diagram to synthesise the findings

Figure 6 Sankey Diagram with outcomes of interest arranged down the left-hand side followed by the number of studies included in narrative synthesis (outcome: studies). Down the right-hand side of the diagram, four explanations for evidence attrition or small meta-analyses, and how many studies' data fell into this category (explanation: studies). Flow from left to right: follow how the outcomes of interest are narrowed into smaller and smaller groups until they can be described by one of the four reasons. Read together with the description of the results, this diagram visually demonstrates why certain meta-analyses were small (such as mortality), why certain meta-analyses could not be undertaken (such as for stewardship; and why Egger's test was not appropriate on any subgroup. Table 2 explains the data further.



The Sankey diagram (Figure 6) helps the reader to interrogate the body of evidence in the review at a glance. For example, we have not meta-analysed antibiotic stewardship outcomes. The reader would know this by reading the entire paper, but the Sankey diagram summarises the point; there are 30 antibiotic stewardship outcomes ("outcomes of interest" in the

diagram) reported in 17 studies. The third flow on the diagram also shows the number of studies (17), the number of definitions (30), and the reason why they could not be combined quantitatively (different endpoints) within one flow. The data behind the Sankey, and more information about how to read it, is represented in Table 2. This table could be adapted for any review that aims to provide more detail about the exclusion of reported quantitative data from subgroup analyses.

Table 2 Table explaining the Sankey Diagram. Explanation of columns from left to right: the outcome of interest, the number of studies reporting the outcome, whether studies were included or excluded in meta-analysis and why, the number of those studies, whether subgroup or statistical variation further divided the studies, the number of studies in each subgroup, and the consequences for meta-analysis.

Outcome of interest	Number of studies)	Included or excluded from meta-analysis	Number (include or exclude)	Subgroup or statistical variation	Number (subgroup)	Consequence for meta-analysis
Length of stay	25	Excluded	12	n/a	n/a	Not enough to aggregate
		Included RCTs	3	Mean/SD	2	Statistical variation*
				Median/IQR	1	Statistical variation*
		Included quasi-experimental studies	10	Mean/SD	7	Statistical variation*
				Median/IQR	3	Statistical variation*
Mortality	21	Excluded subgroup	3	n/a	n/a	Not enough to aggregate
		Included mortality outcomes	22	30-day	8	Different endpoints*
				In-hospital	7	Different endpoints*
				28-day	4	Different endpoints**
				7-day	1	Different endpoints**
				14-day	2	Different endpoints**
Stewardship	17	Excluded	17	Prescribing outcomes	30	Different endpoints**
Turnaround time	19	Excluded	19	Definitions	36	Heterogeneous definitions***

*leading to small meta-analyses and large confidence intervals
**Not enough of the same outcome to aggregate
***Not enough of the same concept to aggregate

Discussion

Overview of diagnostic testing

Appropriate antibiotic therapy is one of the most important aspects of the successful treatment of bacterial infections. RDTs for bacterial identification and antibiotic susceptibility have been developed to try to reduce the time to appropriate antibiotic therapy, shorten length of stay and improve patient outcomes such as mortality. However, our synthesis suggests that the introduction of RDTs for bacterial identification and antibiotic susceptibility testing is unlikely to lead to lower in-hospital mortality or reductions in length of stay. Moreover, while the available observational studies do suggest a significant reduction in 30-day mortality and length of stay, these studies are heterogeneous, have methodological flaws.

The Sankey diagram revealed that there is great heterogeneity even in the mortality outcomes reported, in spite of the recent emphasis on the need for appropriate outcome selection for evaluation of antibiotic therapies (highly relevant to the evaluation of RDTs for antibiotic susceptibility and resistance) and the emerging consensus towards greater use of core outcomes, in particular 28-day, 30-day, or in-hospital mortality.(20) We suggest that Sankey diagrams can be a valuable aid to transparency in systematic reviews, particularly as a way of showing why particular studies and study outcomes become excluded from the final set of analyses. They can also allow for comparisons to be made across review topics; though the diagram does not show this directly, certain issues such as whether there are structural biases leading to more evidence attrition in certain fields of research, or with industry-funded research for example, could begin to be answered with diagrams such as these. It can also address issues of 'research waste'. The diagram shows the amount of evidence that is wasted

because it cannot easily be included in synthesis processes and consequently is left out, or only included in narrative syntheses, which are traditionally given less evidentiary 'weight' than even a small meta-analysis. There are resource and ethical issues associated with this, and the variability between studies may encourage the production of misleading summaries of the evidence, prevent the production of regular systematic reviews of that evidence, and and encourage and cherry-picking of positive outcomes.

The review itself highlighted major problems in the RDT evidence base. One is that the primary studies are often underpowered. Neither bloodstream infections nor resistant bacterial infections are particularly rare, yet sample sizes are surprisingly small throughout all included studies. Egger's test for small study effect is only recommended with 20 or more studies, for example, but the largest subgroup in this review was 10 studies. A further problem is the lack of consistency and clarity in definitions of outcomes - it is often unclear which parts of the care pathway are being reported when the term 'turnaround time' is used in primary studies, and frequently there is no explanation as to why a particular part of the pathway has been chosen, and whether it was chosen *a priori*. This lack of standard definition, measurement and reporting of these outcomes makes it difficult for service providers and policy makers to use evidence to decide whether to invest in RDTs in general and, in turn, which to purchase. It also makes it impossible to synthesise the evidence comprehensively, as shown graphically in the Sankey diagram. Standardising these definitions would help. For example, 'turnaround time' is most useful to clinical commissioners if defined as the time from patient sampling to results being acted upon by clinicians, as this represents the full care pathway likely to be modified by RDTs. To this end Table 3 proposes some definitions to help standardise and clarify these outcomes for future studies (Table 3).

Table 3: Suggested definitions for diagnostic pathway outcomes in RDT evaluations

Turnaround time	The time from collecting a sample from a patient to a laboratory result being actioned by a clinical decision-maker
Time to result	The time from collecting a sample from a patient to the result being released by the laboratory
Running time	The active time of a technology from sample being inserted/inputted into a technology until when the test is complete and an output has been generated.

Limitations

There are several limitations to this study. First, it proved impossible to synthesise the evidence of the effects of RDTs on turnaround time or other antibiotic stewardship outcomes because of the lack of standard definitions of reported outcomes across studies, as shown in the Sankey diagram. Antibiotic stewardship outcomes represent the main positive impact of RDTs according to some commentators, but this remains a controversial assertion given the limitations in the evidence. Also, while experimental studies sometimes incorporated antibiotic stewardship as a discrete third arm in trials so as to disaggregate the effect of the rapid diagnostic test from the effect of the stewardship intervention, many of the pre-post quasi-experimental studies bundled antibiotic stewardship programmes with the addition of a novel diagnostic test. It remains possible that bundling stewardship measures with the diagnostic test may be confounding the impact of the diagnostic intervention. This would reflect previous research in this area.(21–26) We therefore suggest that care should be taken in future studies not to attribute an impact to diagnostics where the impact could have come from improved stewardship measures.

Given the small number of studies in the area as a whole, there is a need for better evidence on the in-hospital impact of RDTs. Some mathematical modelling studies have endorsed intra-hospital infections averted as a useful metric, but the advent of whole genome

sequencing could be employed alongside RDTs to validate attempts to capture this outcome in real-world evaluations. If rapid diagnostics are to demonstrate clinical value, it is likely to be in terms of their effects on such indirect outcomes.

Conclusion

We recommend that future systematic reviews of similar diagnostic technologies consider adopting a health services research perspective, in line with the current review, which takes account not just of final outcomes (mortality; length of stay) but also intermediate outcomes (appropriate antibiotic therapy). Such an approach allows a wider range of the available evidence to be synthesised to help understand the clinical and health services effects of new technologies destined for the hospital laboratory This review shows that there is insufficient high-quality evidence to conclude that these diagnostic technologies reduce length of stay or mortality. This is likely to be because of presumptive treatments and the complexity of the care pathway. Sankey diagrams can help to show how the range of evidence is able to contribute, or not, to a review's conclusions. They may be of particular value in improving the transparency of systematic reviews of complex interventions where the evidence is disparate and assessing its adequacy. Sankey diagrams may also be of use when a review covers many different outcomes and outcome definitions, and is only partially amenable to meta-analysis.

Contributorship statement
NM secured funding for the study
RG, NM, MPP, SJP, and EE developed the research question and fed into the protocol development for the study.
RG and MAH screened all the articles for inclusion, and extracted all the data
RG wrote the draft manuscript and analysed the data
RG, NM, EE, SJP, MAH, and MPP edited the manuscript and approved its submission.
MPP and SJP supervised RG on this work

Strengths and limitations of this study:

Strengths

This is the first systematic review and MA of the effectiveness of RDTs for bacterial identification and antibiotic susceptibility testing which shows that, despite their widespread use and claims about their value, they do not appear to be effective

We developed a novel method to identify, group, and analyse included studies in a systematic review using a Sankey diagram

Sankey diagrams can help compare patterns of methodological quality and variation in outcomes within primary studies across sectors and topics within a review. They provide a visual way of identifying methodological concerns in the evidence included in systematic reviews.

We demonstrated this technique in an area where systematic review and meta-analysis is underused, namely the clinical effectiveness of rapid diagnostic tests for bacterial identification and antibiotic susceptibility testing.

Limitations

While there appears to be evidence of reporting bias (publication bias, small study effects), the paucity of studies included in our systematic review means that Egger's test is underpowered so the influence of publication bias on the summary estimates is difficult to determine.

There is a lack of standard terminology used to report 'turnaround time' and standard antibiotic escalation and de-escalation outcomes of interest; in addition to the greater use of Sankey diagrams we also recommend standardised definitions of, and greater care in, selecting endpoints.

Data sharing
Reasonable requests for data can be requested by email from the corresponding author.

Patient and public involvement
Patients and the public were not involved in conducting the systematic review and meta-analysis, though two PPI representatives reviewed the research question at the beginning of the research process, and also aided us in the development of plain English summaries for public engagement work related to this research.

Transparency Declaration

All authors have submitted an ICMJE COI form. SJP reports personal fees from Specific, and stock options from Next Gen Diagnostics, outside the submitted work. REG, MAH, NM, EE, and MPP have nothing to disclose.

Role of funding source

This work is supported by the NIHR. This report is based on independent research commissioned and funded by the DHSC Policy Research Programme [102/0001] through its core support to the Policy Innovation Research Unit. The views expressed in the publication are those of the authors and are not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, its arm's length bodies or other Government Departments.

Ethics approval

All the evidence were published, peer-reviewed journal articles and ethics approval is therefore not required to conduct a systematic review or meta-analysis.

Acknowledgments

We acknowledge the advice and support of Jane Falconer (librarian at LSHTM), Mike Sharland (Professor of paediatric infectious diseases, St. George's University of London) Mark Wilcox (Professor of medical microbiology, University of Leeds), and Richard Stabler (associate professor of medical microbiology, LSHTM).

Declaration of Competing Interest

All authors have submitted an ICMJE COI form. SJP reports personal fees from Specific, and stock options from Next Gen Diagnostics, outside the submitted work. REG, MAH, NM, EE, and MPP have nothing to disclose.

References

1. Tugwell P, Knottnerus JA. Whither core outcome sets? J Clin Epidemiol. 2014 Jul 1;67(7):731–3.

2. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: An international initiative to improve outcome measurement in rheumatology. Trials. 2007 Nov 26;8(1):38.

3. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0: updated March 2011. 2011.

4. Ruimy R, Dos-Santos M, Raskine L, Bert F, Masson R, Elbaz S, et al. Accuracy and potential usefulness of triplex real-time PCR for improving antibiotic treatment of patients with blood cultures showing clustered gram-positive cocci on direct smears. J Clin Microbiol. 2008 Jun;46(6):2045–51.

5. Holzknecht BJ, Hansen DS, Nielsen L, Kailow A, Jarlov JO. Screening for vancomycin-resistant enterococci with Xpert vanA/vanB: diagnostic accuracy and impact on infection control decision making. New Microbes New Infect. 2017 Mar 1;16:54–9.

6. Jones RN, College of American Pathologists Microbiology Resource C. Method preferences and test accuracy of antimicrobial susceptibility testing: updates from the College of Amercian Pathologists Microbiology Surveys Program. Arch Pathol Lab Med. 2001 Oct;125(10):1285–9.

7. Warhurst G, Dunn G, Chadwick P, Blackwood B, McAuley D, Perkins GD, et al. Rapid detection of health-care-associated bloodstream infection in critical care using multipathogen real-time polymerase chain reaction technology: a diagnostic accuracy study and systematic review. Health Technol Assess. 2015 May;19(35):167.

8. Smidt N, Rutjes AWS, Van der Windt D, Ostelo R, Bossuyt PM, Reitsma JB, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? Neurology [Internet]. 2006;67. Available from: https://doi.org/10.1212/01.wnl.0000238386.41398.30

9. Hayward AC, Goldsmith K, Johnson AM, Surveillance Subgroup of S. Report of the Specialist Advisory Committee on Antimicrobial Resistance (SACAR) Surveillance Subgroup. J Antimicrob Chemother. 2007 Aug;60 Suppl 1:i33-42.

10. Schmidt M. The Sankey Diagram in Energy and Material Flow Management. J Ind Ecol. 2008 Feb 1;12(1):82–94.

11. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960 Apr 1;20(1):37–46.

12. AHRQ Publication No. 10(14)-EHC063-EF In Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville: Agency for Healthcare Research and Quality; 2014. Chapters available at: http://www.effectivehealthcare.ahrq.gov. Available from: http://www.effectivehealthcare.ahrq.gov

13. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.

14. Luo D, Smith JA, Meadows NA, Schuh A, Manescu KE, Bure K, et al. A Quantitative Assessment of Factors Affecting the Technological Development and Adoption of Companion Diagnostics. Front Genet [Internet]. 2016 Jan 28 [cited 2017 Apr 20];6. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4730156/

15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003/09/06 ed. 2003 Sep 6;327(7414):557–60.

16. Egger M, Smith GD, Altman PG. Systematic reviews in health care : meta analysis in context. BMJ Publishers; 2001.

17. Sedgwick P. Meta-analyses: what is heterogeneity? BMJ. 2015 Mar 16;350:h1435.

18. Roisin S, Laurent C, Denis O, Dramaix M, Nonhoff C, Hallin M, et al. Impact of rapid molecular screening at hospital admission on nosocomial transmission of methicillin-resistant Staphylococcus aureus: cluster randomised trial. PLoS ONE Electron Resour. 2014;9(5):e96310.

19. Harbarth S, Masuet-Aumatell C, Schrenzel J, Francois P, Akakpo C, Renzi G, et al. Evaluation of rapid screening and pre-emptive contact isolation for detecting and controlling methicillin-resistant Staphylococcus aureus in critical care: an interventional cohort study. Crit Care Lond Engl. 2006 Feb;10(1):R25.

20. Timsit J-F, de Kraker MEA, Sommer H, Weiss E, Bettiol E, Wolkewitz M, et al. Appropriate endpoints for evaluation of new antibiotic therapies for severe infections: a perspective from COMBACTE's STAT-Net. Intensive Care Med. 2017/05/02 ed. 2017 Jul;43(7):1002–12.

21. MacVane SH, Nolte FS. Benefits of adding a rapid PCR-based blood culture identification panel to an established antimicrobial stewardship program. J Clin Microbiol. 2016 Oct;54(10):2455–63.

22. Bauer KA, Perez KK, Forrest GN, Goff DA. Review of Rapid Diagnostic Tests Used by Antimicrobial Stewardship Programs. Clin Infect Dis. 2014 Oct 15;59(suppl 3):S134–45.

23. Anderson DJ, Jenkins TC, Evans SR, Harris AD, Weinstein RA, Tamma PD, et al. The Role of Stewardship in Addressing Antibacterial Resistance: Stewardship and Infection Control Committee of the Antibacterial Resistance Leadership Group. Clin Infect Dis. 2017/03/30 ed. 2017 Mar 15;64(suppl_1):S36-s40.

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

25. Brumley PE, Collins CD, Malani AN, Kabara JJ, Pisani J. Effect of an antimicrobial stewardship bundle for patients with Clostridium difficile infection. J Antimicrob Chemother. 2015;71(3):836–40.

26. 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 Dec;72(12):3223–31.