Barriers to Evidence Synthesis

2

Kamran Ahmed, Nick Sevdalis, Ara Darzi, and Thanos Athanasiou

Abstract In the hierarchy of research designs, randomised controlled trials and meta-analyses of randomised controlled trials are considered to be the highest level of evidence. They have been established as essential areas of research since their introduction into clinical sciences. Research in the interventional disciplines such as surgery, rely mostly on observational studies. Therefore, the quality and quantity of randomised trials with regards to interventions remain limited. Researchers in these disciplines face various obstacles during building, assessment or implementation of evidence. This chapter aims to provide a critical overview of the obstacles to randomised trials and meta-analyses. It also proposes solution to these problems.

2.1 Introduction

In the hierarchy of research designs, randomised controlled trials (RCTs) are considered to be the highest level of evidence.[1] RCTs were introduced into clinical epidemiology after evaluation of streptomycin for management of tuberculosis.[2] Since then, RCTs have become the gold standard for assessing the effectiveness of therapeutic agents.

When various randomised studies are available on a same topic, a well-conducted meta-analysis of these randomised trials is regarded as the best level of evidence within evidence-based medicine (EBM). Meta-analyses statistically integrate the results of several independent studies considered to be combinable, thus allowing evaluation of the evidence within traditional studies that is at risk of being overlooked, and provide more precise estimates of treatment effects.[3,4]

K. Ahmed (✉)
Department of Surgery and Cancer, Imperial College London, St Mary's Hospital Campus, London, UK
e-mail: k.ahmed@imperial.ac.uk

T. Athanasiou and A. Darzi (eds.), *Evidence Synthesis in Healthcare*,
DOI: 10.1007/978-0-85729-206-3_2, © Springer-Verlag London Limited 2011

Meta-analyses offer an opportunity to test implicit assumptions about the hierarchy of research designs. Ideally, if associations between exposure and outcome were studied in both randomised controlled trials and cohort or case–control studies, and if these studies were then included in meta-analyses, the results could be compared according to study design.[1] However, RCTs may overlook clinically essential benefits because of poorly constructed design – for instance, inadequate attention to sample size.[5] Therefore, an amalgamation of studies using robust statistical methodologies can overcome some of the deficiencies within the primary studies.

Randomised trials and meta-analyses have been established as essential areas of research since their introduction into clinical sciences. Certain medical disciplines, such as surgical specialities primarily, use observational studies for identification of risk factors and prognostic indicators. In these disciplines, ethical issues related to type and timing of intervention may prevent clinicians from regularly conducing RCTs. Therefore, the quality and quantity of randomised trials with regards to interventions in specialties such as surgery remains limited.[6] Moreover, a number of other factors may limit conduct of good quality trials or meta-analyses. These factors are related to barriers due to building, assessment or implementation of evidence (Fig. 2.1).

This chapter aims to provide a critical overview of the obstacles to randomised trials and meta-analyses. It also proposes solution to these problems.

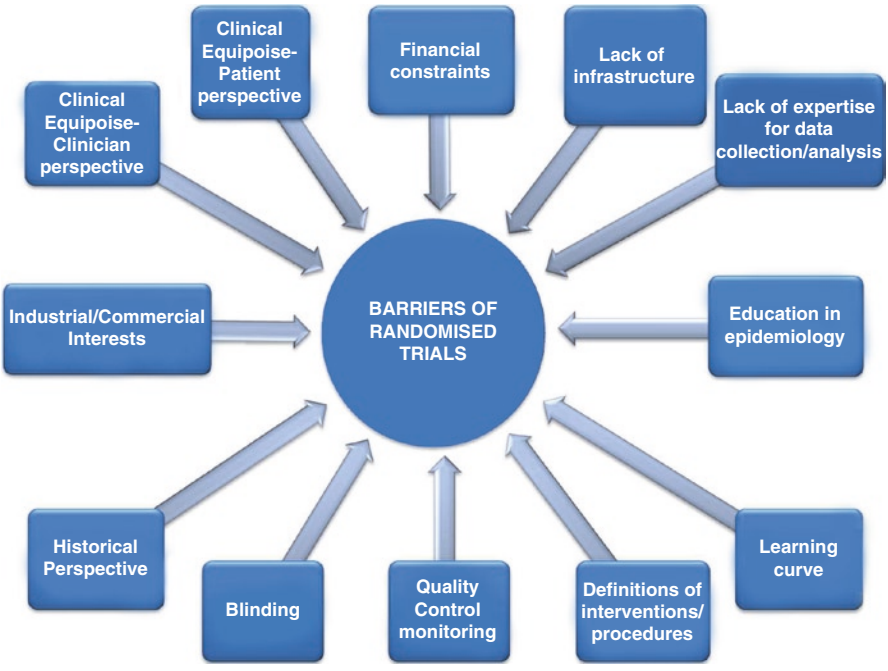

Fig. 2.1 Key determinants of successful building, assessment and implementation of evidence

2.2 Barriers to Randomised Trials

The purpose of a randomised trial is to provide the means by which the highest level of evidence from research can be judiciously and vigilantly applied to the prevention, detection, treatment and follow-up of health conditions. However, there are several obstacles to the successful conduct and application of RCTs. This section covers obstacles to the conduct of such trials (Fig. 2.2).

2.2.1 Historical Perspective

Validation of interventional procedures is generally not based on randomised trials. Conventionally, the steps in a procedure evolve with subtle changes over the passage of time. Once they are in practice, assessment of effectiveness against a placebo becomes difficult. The treatment benefit becomes so obvious that randomisation can be argued to be unethical.[7]

Fig. 2.2 Factors determining conduct and quality of randomised trials

Occasionally, developing therapeutic and diagnostic modalities may jeopardise the conduct of a clinical trial before its conclusion. For instance, an RCT needs to be stopped if novel surgical or technical developments render the results of the trial outdated before its completion.[8]

Moreover, an RCT cannot be conducted if a new technology or intervention is likely to undergo modifications in the near future or if this technology or technique is complex and has been developed only recently.

2.2.2 Clinical Equipoise: Clinician's Perspective

Clinical equipoise refers to uncertainty over whether a treatment will be beneficial or not. Shaw et al. argued that if a clinician has good reasons to believe that a new therapy is better than an existing therapy, he or she cannot take part in a comparative trial of the new versus old therapy. Under such circumstances, the clinician is ethically obligated to offer the new (and believed better) therapy to each new patient with a need for one of these therapies.[9] For this reason, clinicians who believe that they already practise the best option cannot participate into a trial.

On the other hand, uncertainty with regards to the best treatment option is beneficial for a patient. In this case, offering patients randomisation to equally preferred treatments is acceptable and does not violate ethical principles. This 'uncertainty principle' has been successfully used as a main eligibility criterion for large clinical trials.[10] Ambiguity on the part of all participants remains the moral and practical code of conduct that requires ethical justification of randomised trials.

2.2.3 Clinical Equipoise: Patient's Perspective

Equipoise is also important from the perspective of the patient. This is particularly true in RCTs of surgical interventions, where both trial and control arms are likely to have associated risks. Ethical principles dictate that patients should not consent to randomisation unless there is true uncertainty about the superiority of a treatment option.

Type III trials (comparing surgical and nonsurgical treatments) may pose some difficulties with the equipoise of patients.[11] Patients often refuse to take part in such RCTs as they prefer a firm decision on whether they will receive surgical or nonsurgical treatment, and not one left to be determined by chance.

2.2.4 Industrial and Personal Influences

Commercial and personal interests can interfere with the process of building or evaluating evidence, if this process is perceived to be potentially damaging to such interests.[11] For commonly available products, randomised studies can be threatening for the commercial organisations in terms of financial gains. Similarly, for widely practised procedures randomisation can be challenging to the objectivity and the practice of a surgeon.

Competition in a private sector may influence the clinicians.[11] For instance, in the initial 5 years after the introduction of cholecystectomy, only two randomised clinical trials were published.[12,13] This issue came to light after a number of reports were quoted with regards to increasing incidence of bile duct injuries.[14] Similarly, robotic-assisted procedures and single-incision laparoscopic approaches lack randomised studies to support their efficacy despite increasing clinical use of these approaches.[15]

In recent years, however, there is a progressing drift towards conduct of RCTs for newly developed treatment modalities. Different institutions are conducting RCTs worldwide for a variety of interventions.[16,17]

2.2.5 Lack of Funding and Infrastructure

Although randomised studies in interventional disciplines such as surgery are commonly performed across various regions, they are relatively few in number and lack standardised protocols compared to similar RCTs for non-interventional treatments. This may reflect a lack of expertise by participating clinicians in trials and shortage of funding for interventional trials.[18] It may also be due to the fact that funding bodies are reluctant to finance research in disciplines associated with previously poor research results.[19] Due to the lack of well-developed procedures, it has been shown that only a quarter of surgical trials report the randomisation process.[20]

2.2.6 Lack of Expertise for Appropriate Design, Data Collection and Analysis

Kelly et al. reported that many clinicians, especially surgeons, have an overambitious confidence in the ability of randomised trials to determine the practical value of interventions.[21] Consequently, energy is expended on data collection which can be fundamentally flawed due to inadequate power of the study.

Hall et al. explained why lack of focus on appropriate methodology and data analysis remain barriers to a good randomised trial.[20] It can be argued that there are primarily two types of clinical studies: explanatory and pragmatic.[22] Explanatory studies aim to assess whether a treatment has any efficacy in comparison with placebo under ideal, experimental conditions. Pragmatic studies aim to assess the effectiveness of a treatment in everyday clinical practice. Most trials attempt to address both explanatory and pragmatic types simultaneously. This practice, however, may result in findings that are not valid and that ultimately cannot provide a robust answer to either type of question.[20,22]

2.2.7 Life-Threatening and Emergency Situations

Emergency management occurs both during and out of the normal working hours. It makes consent and randomisation of the treatment or intervention difficult. Randomised trials, if

conducted for emergency conditions, may miss clinically important benefits because of insufficient attention to patient selection and sample size. In order to conduct an RCT focused on an emergency condition, very selective inclusion criteria need to be followed.[23,24]

2.2.8 Dramatic Discoveries or Rare Clinical Conditions

In an incurable condition when survival is unlikely with or without treatment, an RCT is unnecessary if even one patient survives when a new treatment is administered. Similarly, if a new treatment modality, for instance appendicectomy for appendicitis, produces a rapid improvement in outcome in uncontrolled or nonrandomised controlled trials an RCT may be unnecessary or even unethical.[8] An RCT should be discontinued if a new intervention shows more than 70% absolute improvement in results compared with an established therapy.[8,11]

2.2.9 Education in Epidemiology

Detailed knowledge of epidemiology principles that are necessary for the competent conduct of an RCT remains rather poor in some groups of clinicians (e.g. surgeons).[11,25] However, there is no objective evidence that clinicians in surgical disciplines lack training of clinical epidemiology. Rather, surgical specialties tend to lack dedicated clinical teams with relevant epidemiological expertise who should be responsible for identification, design and conduct of randomised trials.

2.2.10 Learning Curve

Effective interventional techniques come into practice rapidly. Learning curve that represents average rate of learning is achieved after repeated practice. Various authors argue that RCTs should begin from the first report of a new procedure.[26] However, this approach is not ethically acceptable as the clinicians are in still at the beginning of their learning curves during the introduction phase of a new treatment.[27]

Surgical procedures are complex, and proficiency is achieved after frequent repetition. At early stages of the learning curve, errors and adverse events are likely to occur.[28,29] Randomising between a new and an established operation may introduce bias against the new methods.[30] Moreover, patient randomisation to surgeons, although an option, remains untested.[11]

2.2.11 Definitions of Interventions/Procedures

Procedural learning curves cause difficulty in timing and performing randomised trials. When comparing treatment modalities, clear definition of each treatment step is needed.

This leads to acceptance of limitations of variations for a certain procedure. Variations on an operation, however, are not uncommon and may influence outcomes. Moreover, it can be argued that definitions continue to change during the introductory phase of new interventions. Because of these issues with surgical treatments, non-standardised procedural definitions may lead to controversy whether a trial has truly investigated the intended treatment.[27] Definitions of treatment are overall more challenging for interventions than for drug trials, in which a 'treatment' is simpler to define.

2.2.12 Quality Control Monitoring

The methodology of a clinical trial and the technical quality of clinical interventions may have an impact on outcomes. The expertise of the clinicians who are carrying out the intervention is one of the determining factors of the quality of outcomes. Poor results fail to deliver intended treatment; therefore, effectiveness of the trial remains doubtful. Failure to maintain consistently high quality of procedures may narrow important technical differences and may have impact on outcomes.[31,32]

2.2.13 Blinding

Blinding is important to protect internal validity, and significant bias may result from unsuccessful blinding. Blinding, however, is particularly challenging in trials involving interventions. Examples where placebos are not possible or unethical include surgical interventions, as well as treatments where active participation of the patient is necessary (e.g. physical therapy). In fact, only about a third of published interventional trials have been reported to adequately address principles of blinding.[11,18]

2.3 Proposed Solutions to Overcome the Barriers of RCTs

To improve the quality of randomised trials, the barriers discussed above need to be addressed meticulously (Fig. 2.3). RCTs offer the highest level of evidence for assessing efficacy of treatments and direct relevant evidence from high-quality RCTs should be used wherever possible. Key problems with regard to conduct and quality of trials mainly affect trials in interventional disciplines, such as surgery and interventional radiology. For craft specialities (i.e. specialities that are dependent heavily on minor or major interventions), existing frameworks do not effectively address the range of potential problems, either conceptual or methodological.[8,11]

Figure 2.4 summarises a comprehensive framework that addresses a number of issues identified here through phased introduction of a trial, regular audit of data collection, and continuous evaluation of the quality of the trial. The framework identifies issues around learning curves, variations in technique or type of interventions, which need to be addressed

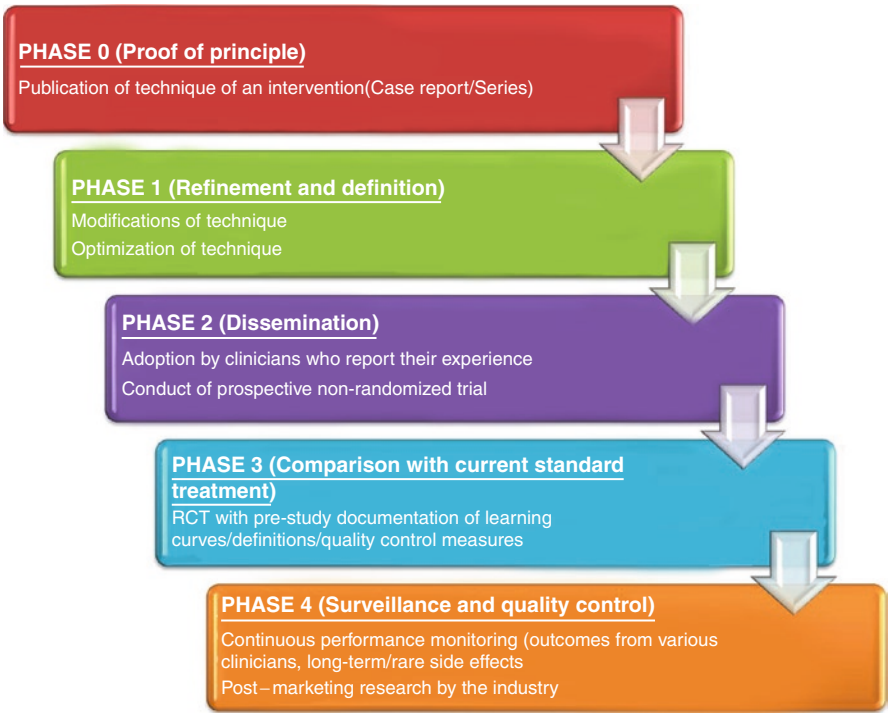

Fig. 2.3 Barriers related to intervention, researcher and methodology of a meta-analysis

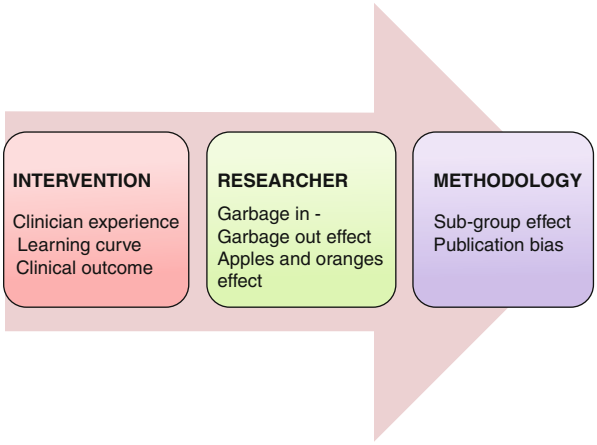

Fig. 2.4 Proposed framework for planning and conduct of RCTs in interventional specialities[27]

and documented appropriately to ensure adequate methodological quality. The framework also proposes an initial phase of non-randomised trials to be incorporated into RCTs in order to determine suitable end points.[11]

2.4 Barriers Specific to Meta-Analyses

The conduct of meta-analysis in case of rare conditions and interventions is particularly difficult and needs timely investigation, standardised definitions, availability of high-quality data (from RCTs and meticulously done observational studies) and statistical expertise.

This section delineates obstacles to meta-analyses and considers solutions to these problems (Fig. 2.5). The problems and their solutions fall into three distinct categories associated with: (1) the intervention, (2) the researcher and (3) the methodology.

2.4.1 Intervention-Related Barriers

The primary differences between meta-analyses in craft disciplines and those in other fields originate from the reproducibility of treatments and variations in practice that are difficult to compare. The outcomes of a surgical procedure principally depend on the level of experience of a clinician. This is not the case in other areas of research such as drug trials, where interventions tend to be significantly more consistent and drugs act in a uniform manner. Early meta-analytic assessment of a new procedure or technique may give a misleading picture of its efficacy due to lack of competence of the surgeons who are carrying out a new procedure. Factors determining whether an interventional procedure will

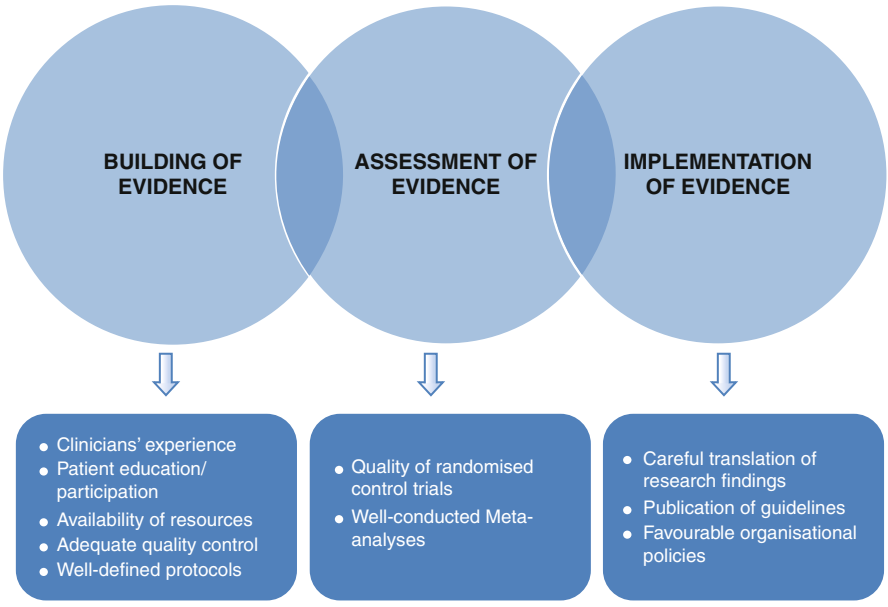

Fig. 2.5 Proposed solutions to address issues related to the conduct and control of RCTs[11]

be carried out competently include the clinician's experience, available equipment and time. Performance continues to improve until a plateau phase is reached as manifested by the 'learning curve'. The experience of a clinician is a key confounder during comparative trials involving interventions. Less experienced clinicians have relatively poorer outcomes,[33] which are less likely to be reported, thus adding further to publication bias.[3,34] These issues compromise the validity of a meta-analysis. For instance, with the advent of new interventions such as Natural orifice transluminal endoscopic surgery (NOTES) which are supported by lower levels of evidence,[35] caution must be exerted when the first meta-analysis of this procedure takes place. Small sample size in individual trials, year of publication and poor definitions need to be anticipated ahead of the analysis otherwise meta-analysis can be fraught with these issues.[36]

Meta-analyses may produce conflicting evidence when results are pooled from small trials with disparate outcomes. If the results are in conflict with large RCTs, the reliability of the evidence becomes debatable.[37,38] In a study comparing 19 meta-analyses and 12 large RCTs on the same topic, LeLorier et al. found that the results in 5 out of 40 outcomes were significantly different between the RCTs and those of the meta-analyses.[39] Meta-analysis cannot be a substitute for large clinical trials.[40] However, it may be a useful guide to clinical decision makers until explicit experimental evidence becomes available.

The year of publication of a study is also a strong confounder to the overall results revealed by a meta-analysis as population characteristics and outcome data may change over time. Furthermore, new developments in technology and changes in clinicians' technical expertise may translate into variable outcomes over time. All these factors need to be considered especially in surgical disciplines where new technologies and techniques are continuously developed and the learning curve is progressively overcome.[41] Increasing accumulation of evidence with time improves the robustness of results reported by a meta-analysis.[42]

2.4.2 Researcher-Related Barriers

A researcher may face several challenges whilst conducting meta-analysis. One of these is the *Garbage in, garbage out* effect: If a meta-analysis includes several low-quality studies, then basic errors in the primary studies will be translated across to the meta-analysis, where the errors may be difficult to identify. The quality of meta-analysis is determined by the quality and type of studies included. Because the nature of procedural interventions often makes it difficult to perform well-conducted RCTs, inclusion in meta-analyses of observational studies (cross-sectional, case series, case-control or cohort design) remains common yet controversial as they are vulnerable to bias by confounding factors.[43]

Another common problem is *the apple and oranges* effect, which results from combining different study designs in an analysis. This may lead to an erroneous result being produced (relative risk, odds ratio or weighted mean difference).[44] This apparent difference in effect across pooled studies is termed as 'heterogeneity'. In a meta-analysis, three principal sources of heterogeneity are clinical (e.g. baseline difference between patients from different studies), statistical (e.g. effects being estimated by individual studies in a meta-analysis

are not identical) and methodological (e.g. design-related heterogeneity).[45] The ultimate aim of pooling data from different studies is to provide a single best estimate of treatment effect between two treatment groups. It, therefore, is vitally important to combine 'apples' with 'apples'.[46]

2.4.3 Methodology-Related Barriers

Several challenging aspects such as the subgroup effect and publication bias can compromise the methodology of a meta-analysis.

A meta-analysis aims to produce an estimate of the average effect seen in trials of a particular treatment termed as *subgroup effect*.[47] It is necessary to determine whether the overall effect applies to all participating individuals, or whether some subgroups have different effect than others.

Publication bias refers to the greater likelihood of studies with positive (i.e. statistically significant) results being published.[22] Exclusion of studies from the meta-analysis because they are small in size, found negative results or for other reasons can bias the results. This is termed as a 'file drawer problem'. It may be intentional or due to the results of a flawed and incomplete literature search.[48] This publication bias may render meta-analysis of published literature misleading, thus compromising patient safety.[49] Another problem is that of 'grey literature' which refers to the studies not published as formal, peer-reviewed journal articles (e.g. those found in conference abstracts, books, theses, government and company reports and other unpublished material). These can also potentially include studies that report negative results and are not published or lie dormant in a researcher's filing cabinet.

2.5 Proposed Solutions to Overcome the Barriers of Meta-Analysis

2.5.1 Solutions to Intervention-Related Problems

It is imperative to account for and wait for the learning curve to be achieved and differentiate between high-volume and low-volume centres. The individual trials should be conducted once the learning curve has been achieved, thereby having experts performing the procedures. Moreover, subgroup analyses between high-volume and low-volume centres should be performed to account for effect of case load, if any, on the results.

Quality control for the included studies should be assessed meticulously. A commonly used scoring method is that developed by Jadad et al. which assigns points based on the presence of randomisation, double blinding, and adequate description of withdrawals and dropouts.[50]

In case of a lack of adequate evidence, a meta-analysis is not an appropriate method of clinical appraisal, as the analysis will suffer from insufficient data and heterogeneous

outcomes being reported. Such a scenario has been reported by Slim et al. in 2008 who tried to clarify the controversy surrounding the timing of elective surgery of colonic diverticulitis in young patients.[35] Out of 15 articles selected for inclusion in this study, only 3 papers reported information regarding the timing of surgery. The authors concluded that the researchers should no longer attempt to address this question by a meta-analysis.[35]

2.5.2 Solutions to Researcher-Related Problems

Several tools have been developed to assess the quality of individual meta-analyses.[51] Guidelines exist to assess the quality of both randomised (QUOROM statement) and observational studies (MOOSE statement).[52,53] A recent update to the QUOROM statement is PRISMA, which focuses on recognition of bias through meticulous quality assessment.[54] These tools can be an invaluable source to understand and quantify sources of variability across studies and should be encouraged.[53] Although several quality assessment tools (checklists) exist, there are discrepancies amongst them. The variability across different checklists suggests that each individual bias-reducing measure such as randomisation, concealment of allocation and blinding should be documented across studies.[55]

Identification of heterogeneity can highlight factors that influence outcomes that are not observable in individual trials. If performed before a new study, it may help the investigator improve the design by incorporating an understanding of the factors that contribute to heterogeneity. There are three ways to assess heterogeneity. First, through assessment of 'between-studies variance – τ^2'. This primarily depends on the particular effect size metric used. Second, 'Cochrane's Q test', which follows a chi-square distribution to make inferences about the null hypothesis of homogeneity. The problem with this test is that it has poor power to detect true heterogeneity when the number of studies is small. None of the above-mentioned methods have a standardised scale. Therefore, they are poorly equipped to make comparisons of the degree of homogeneity across meta-analyses.[56] Third method for quantifying inconsistency is '$I^2 = [(Q - df)/Q] \times 100\%$', where Q is the chi-squared statistic and d*f* represents degrees of freedom.[57] This method is easier to utilise because it defines variability along a scale-free range as a percentage from 0% to 100%. This describes the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error (chance). Heterogeneity could be considered substantial when this value is greater than 50%.[57] It is worth noting that tests for assessment of heterogeneity lack power to reject the null hypothesis of homogeneous results and should be used even if substantial differences between the studies exist as they provide an opportunity for examining why treatment effects differ in different circumstances.[58]

Another way to account for heterogeneity is to make use of 'Random effects model' and 'fixed effects model'. If a test for homogeneity shows homogeneous results then the differences between studies are assumed to be a consequence of sampling variation, and a fixed effects model is appropriate. If, however, the test shows that significant heterogeneity exists between study results then a random effects model is advocated. If the heterogeneity is very high and not compensated by the random effects model, the viability of the meta-analysis becomes questionable.

Sensitivity analysis can also contribute to enhance the quality of the results by considering the extent of reporting of input parameters. It involves: (1) re-analysing the results by using all studies and then excluding poorer quality studies, (2) using both fixed and random effects meta-analyses to assess the robustness of the results to the method used and (3) repeating the meta-analysis by excluding any study that is an outlier to assess its influence.

2.5.3 Solutions to Methodology-Related Problems

Subgroup analysis delineates heterogeneity. However, if trials are split into too many groups, the probability of false-positive results increases (Type I error). Splitting a meta-analysis into subgroups should be subjected to a power analysis. There should also be a strong biological rationale for performing a subgroup analysis and care should be taken in the interpretation of any effects, which are likely to be composite.[59] Sub-group analyses may generate hypotheses which can assist decision-making between different treatment options.[60]

The presence of publication biases can be identified through stratifying the analysis by study size. Smaller effects can be significant in larger studies. If publication bias is present, larger studies are likely to report smaller effects. However, exclusion of the smallest studies has little effect on the overall estimate. Thus, sensitivity analysis is useful in that it assess whether the results from a meta-analysis are valid and not affected by the exclusion of trials of poorer quality or of studies stopped early. It also takes into account publication bias.[34] Because of the potential impact of publication bias, many meta-analyses now include a 'failsafe N' statistic that calculates the number of studies with null results that would need to be added to the meta-analysis in order for an effect to no longer be reliable.

Formal tests for publication bias exist, but in practice few meta-analyses have assessed or adjusted for the presence of this bias. Examination of a 'funnel plot' is one of the simplest and most commonly used methods to detect publication bias.[61] However, visual inspection of funnel plots might be subjective and so statistical tests for detecting plot asymmetry can also be used. Regression tests and rank correlation tests are some of the options available. In addition, various methods such as the 'trim and fill method' and 'weight modelling' could be undertaken to compensate for publication bias.[62] Other biases associated with time lag, English language, citation, duplication and outcome reporting should also be considered equally important when analyzing data.

2.6 Challenges to Implementation of Evidence from Randomised Trials and Meta-Analyses

In the previous sections, we detailed a range of potential barriers to evidence synthesis, either at the level of the randomised trial or at the level of the meta-analysis of such trials, and we also discussed solutions that could help overcome such barriers. Importantly,

however, even well-conducted trails and meta-analysis that should, in theory, inform clinical practice are not taken up as much or as quickly as proponents of evidence-based medicine would expect.[63-66] In fact, lack of adequate uptake of the outcomes of such research work renders the previous discussion irrelevant: why care about the quality of evidence synthesis if the target audience (i.e. clinicians on the ground) will not implement the new findings? Reasons why high-level clinical evidence often does not change practice as much as should be expected are complex, and involve a range of factors. In what follows, three key categories of factors are discussed (the topic has been explored in great detail in numerous publications, to which we point the interested reader for further resources[6,67,68]).

A first category of reasons why novel evidence-based treatments are not implemented revolves around the *nature of the evidence* itself, as well as the processes of collecting and synthesising it. First of all, the rate of novel publications of RCTs and meta-analyses (as well as systematic reviews) exceeds the knowledge absorption rate of any clinician – even those working in academic environments. In 2006, the number of surgery-related RCTs in the Cochrane library was estimated near 35,000. This 'evidence flood' comes through an ever increasing number of journals, both specialty-specific and general medical journals.[69] No single person can possibly cope with this volume of new knowledge. To tackle this problem, it has been suggested that 'evidence filters' should be designed and applied, so that the relevant evidence does reach the intended audiences.[69] Information technology has also been implicated in attempts to make evidence more easily available.[68-70]

A second category of reasons underpinning lack of integration of evidence into practice relate to *clinicians* themselves. As discussed earlier, many clinicians are not familiar with the methodologies and techniques used in evidence synthesis and therefore they are not able to appraise the quality and importance of the work. In addition, searching through multiple databases using 'clever' keywords that are sensitive enough to select relevant papers but also not too inclusive so that hundreds or thousands of entries are retrieved is a skill.[71] Once a paper has been retrieved, appraising the quality of the reported study is an additional skill, necessary to be able to evaluate the robustness of the design and strength of the conclusions.[71] Current clinical training does not routinely equip clinicians with such skills – or at least it does not do so at the level required to routinely browse through databases to find reviews or RCTs that have the potential to change clinical practice. Keeping in touch with the ever-evolving evidence base and becoming informed of new techniques and approaches that can potentially benefit patients is part of clinicians' duty to robust continuing professional development.[72-75] As such, it ought to be recognised as a component of revalidation and recertification and actively encouraged within professional and regulatory networks (e.g. Royal Colleges).

Lack of knowledge or skills is often compounded by a disapproving attitude towards the methodology of RCTs, which reflects *philosophical/epistemological reasons* why evidence is not 'automatically' translated into practice. Clinicians often complain that RCTs report evidence derived from very carefully selected patients, who do not mirror those in their direct care.[76] A similar complaint arises in relation to the external validity of the procedure of the RCT, which is perceived to be carried out 'by ultraspecialists in quaternary care centres' – again not reflecting the 'average' clinician in an 'average' generalist hospital.[77] These questions are valid, and reflect the difficulty of safely extrapolating from a specific

study population to a more general patient population. To some degree, these issues could be addressed at the design stages of the RCT. They should also be extensively addressed in the context of meta-analyses, where patient inclusion criteria should be scrutinised. Importantly, since patient populations will always differ, it should be remembered that direct replication of treatment benefits from a published RCT may not occur due to either random error (inevitable sample variations) or systematic error (which should trigger a new study), or specific subgroup analyses to establish whether treatment effects are uniform across different patient groups.[78]

Finally, it is important to note that *practicalities of the healthcare environment* as well as the prevailing *culture* in an organisation also affect evidence uptake. No matter how motivated to seek the newest, innovative treatment options for their patients, clinicians will not be able to do so if they are constantly working under time pressure to deliver service.[79] High pressure to increase patient throughput favours tried and tested approaches and also does not allow room for exploration of the evidence base. A key reason why junior trainees learn through observation, in addition to this being the traditional apprenticeship model of learning in medicine, is that this is the fastest way to learn how to treat a patient. Access to the evidence base that involves multiple trips to the medical library, or constant use of IT facilities is nearly impossible in an environment where consultations can only last a few minutes. Apart from practicalities, however, evidence-based medicine experts also discuss the issue of an 'EBM culture', which may or not be prevailing within a unit or organisation and which affects the willingness of clinicians to engage with evidence and make the most of it.[70] Taken together with the philosophical/epistemological issues mentioned above, although practicalities are often very demanding, care should be taken in the design of a clinician's job to allow time and 'mental space' for evidence review.

2.7 Conclusions

Randomised controlled trials and meta-analyses are valuable tools for effective evidence synthesis. If used judiciously and conducted with scientific rigor, they can guide clinical decisions and health policy towards improved patient outcomes. Overcoming barriers to robustly synthesising evidence and implementing it to everyday clinical practice can enhance the strength of evidence derived from research studies, and ultimately improve safety and quality of care. Future research should focus on developing refined protocols for the undertaking and reporting of randomised trials and meta-analyses, as well as on better understanding and sustainably overcoming barriers to implementing evidence.

References

1. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887-1892.

2. MRC_Investigation. Streptomycin treatment of pulmonary tuberculosis: a Medical Research Council investigation. *BMJ*. 1948;2:769-782.
3. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ*. 1997; 315:1533-1537.
4. Huque MF. Experiences with meta-analysis in NDA submissions. *Proc Biopharm Sect Am Stat Assoc*. 1988;2:28-33.
5. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med*. 1982;72:233-240.
6. Farrokhyar F, Karanicolas PJ, Thoma A, et al. Randomized controlled trials of surgical interventions. *Ann Surg*. 2010;251:409-416.
7. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ*. 1996;312:1215-1218.
8. Solomon MJ, McLeod RS. Should we be performing more randomized controlled trials evaluating surgical operations? *Surgery*. 1995;118:459-467.
9. Shaw LW, Chalmers TC. Ethics in cooperative clinical trials. *Ann NY Acad Sci*. 1970;169: 487-495.
10. Weijer C, Shapiro SH, Cranley Glass K. For and against: clinical equipoise and not the uncertainty principle is the moral underpinning of the randomised controlled trial. *BMJ*. 2000; 321:756-758.
11. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ*. 2002;324:1448-1451.
12. Barkun JS, Barkun AN, Sampalis JS, et al. Randomised controlled trial of laparoscopic versus mini cholecystectomy. The McGill Gallstone Treatment Group. *Lancet*. 1992;340:1116-1119.
13. McMahon AJ, Russell IT, Baxter JN, et al. Laparoscopic versus minilaparotomy cholecystectomy: a randomised trial. *Lancet*. 1994;343:135-138.
14. Neugebauer E, Troidl H, Kum CK, Eypasch E, Miserez M, Paul A. The E.A.E.S. Consensus development conferences on laparoscopic cholecystectomy, appendectomy, and hernia repair. Consensus statements – September 1994. The Educational Committee of the European Association for Endoscopic Surgery. *Surg Endosc*. 1995;9:550-563.
15. Ahmed K, Khan MS, Vats A, et al. Current status of robotic assisted pelvic surgery and future developments. *Int J Surg*. 2009;7:431-440.
16. Nix J, Smith A, Kurpad R, Nielsen ME, Wallen EM, Pruthi RS. Prospective randomized controlled trial of robotic versus open radical cystectomy for bladder cancer: perioperative and pathologic results. *Eur Urol*. 2010;57:196-201.
17. Ouriel K. The PIVOTAL study: a randomized comparison of endovascular repair versus surveillance in patients with smaller abdominal aortic aneurysms. *J Vasc Surg*. 2009;49:266-269.
18. Solomon MJ, Laxamana A, Devore L, McLeod RS. Randomized controlled trials in surgery. *Surgery*. 1994;115:707-712.
19. Solomon MJ, McLeod RS. Surgery and the randomised controlled trial: past, present and future. *Med J Aust*. 1998;169:380-383.
20. Hall JC, Mills B, Nguyen H, Hall JL. Methodologic standards in surgical trials. *Surgery*. 1996;119:466-472.
21. Diamond GA, Forrester JS. Clinical trials and statistical verdicts: probable grounds for appeal. *Ann Intern Med*. 1983;98:385-394.
22. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867-872.
23. Hochman JS, Sleeper LA, Godfrey E, et al. Should we emergently revascularize Occluded Coronaries for cardiogenic shock: an international randomized trial of emergency PTCA/CABG-trial design. The SHOCK Trial Study Group. *Am Heart J*. 1999;137:313-321.
24. Hochman JS, Sleeper LA, White HD, et al. One-year survival following early revascularization for cardiogenic shock. *JAMA*. 2001;285:190-192.

25. van der Linden W. Pitfalls in randomized surgical trials. *Surgery*. 1980;87:258-262.
26. Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med*. 1983;309:1358-1361.
27. McCulloch P. Developing appropriate methodology for the study of surgical techniques. *J R Soc Med*. 2009;102:51-55.
28. Parikh D, Johnson M, Chagla L, Lowe D, McCulloch P. D2 gastrectomy: lessons from a prospective audit of the learning curve. *Br J Surg*. 1996;83:1595-1599.
29. Testori A, Bartolomei M, Grana C, et al. Sentinel node localization in primary melanoma: learning curve and results. *Melanoma Res*. 1999;9:587-593.
30. Bonenkamp JJ, Songun I, Hermans J, et al. Randomised comparison of morbidity after D1 and D2 dissection for gastric cancer in 996 Dutch patients. *Lancet*. 1995;345:745-748.
31. Bonenkamp JJ, Hermans J, Sasako M, et al. Extended lymph-node dissection for gastric cancer. *N Engl J Med*. 1999;340:908-914.
32. McCulloch P. D1 versus D2 dissection for gastric cancer. *Lancet*. 1995;345:1516-1517. author reply 1517–1518.
33. Sauerland S, Seiler CM. Role of systematic reviews and meta-analysis in evidence-based medicine. *World J Surg*. 2005;29:582-587.
34. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315:629-634.
35. Slim K, Raspado O, Brugere C, Lanay-Savary MV, Chipponi J. Failure of a meta-analysis on the role of elective surgery for left colonic diverticulitis in young patients. *Int J Colorectal Dis*. 2008;23:665-667.
36. Hernandez AV, Walker E, Ioannidis JP, Kattan MW. Challenges in meta-analysis of randomized clinical trials for rare harmful cardiovascular events: the case of rosiglitazone. *Am Heart J*. 2008;156:23-30.
37. Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials. Examples from the management of acute myocardial infarction. *Ann Intern Med*. 1995;123:873-877.
38. Cappelleri JC, Ioannidis JP, Schmid CH, et al. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA*. 1996;276:1332-1338.
39. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med*. 1997;337:536-542.
40. Bailar JC 3rd. The promise and problems of meta-analysis. *N Engl J Med*. 1997;337:559-561.
41. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet*. 1998;351:123-127.
42. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med*. 1992;327:248-254.
43. Huston P, Naylor CD. Health services research: reporting on studies using secondary data sources. *CMAJ*. 1996;155:1697-1709.
44. Wachter KW. Disturbed by meta-analysis? *Science*. 1988;241:1407-1408.
45. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. *Stat Med*. 2005;24:3823-3844.
46. Ioannidis JP, Patsopoulos NA, Evangelou E. Uncertainty in heterogeneity estimates in meta-analyses. *BMJ*. 2007;335:914-916.
47. Davey Smith G, Egger M, Phillips AN. Meta-analysis. Beyond the grand mean? *BMJ*. 1997;315:1610-1614.
48. Rosenberg MS. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*. 2005;59:464-468.
49. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA*. 2006;295:676-680.

50. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials*. 1996;17:1-12.
51. Oxman AD. Checklists for review articles. *BMJ*. 1994;309:648-651.
52. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet*. 1999;354:1896-1900.
53. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283:2008-2012.
54. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535.
55. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999;282:1054-1060.
56. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods*. 2006;11:193-206.
57. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002; 21:1539-1558.
58. Bailey KR. Inter-study differences: how should they influence the interpretation and analysis of results? *Stat Med*. 1987;6:351-360.
59. Sevdalis N, Jacklin R. Interaction effects and subgroup analyses in clinical trials: more than meets the eye? *J Eval Clin Pract*. 2008;14:919-922.
60. Cook DI, Gebski VJ, Keech AC. Subgroup analysis in clinical trials. *Med J Aust*. 2004;180: 289-291.
61. Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. *Ann Surg*. 2009;249:376-383.
62. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*. 2000;56:455-463.
63. Gray SM. Knowledge management: a core skill for surgeons who manage. *Surg Clin North Am*. 2006;86:17-39. vii-viii.
64. Howes N, Chagla L, Thorpe M, McCulloch P. Surgical practice is evidence based. *Br J Surg*. 1997;84:1220-1223.
65. Kingston R, Barry M, Tierney S, Drumm J, Grace P. Treatment of surgical patients is evidence-based. *Eur J Surg*. 2001;167:324-330.
66. Meakins JL. Evidence-based surgery. *Surg Clin North Am*. 2006;86:1-16. vii.
67. Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet*. 2009;374:1097-1104.
68. Knight T, Brice A. Librarians, surgeons, and knowledge. *Surg Clin North Am*. 2006;86:71-90. viii–ix.
69. Glasziou P. Managing the evidence flood. *Surg Clin North Am*. 2006;86:193-199. xi.
70. Lee MJ. Evidence-based surgery: creating the culture. *Surg Clin North Am*. 2006;86: 91-100. ix.
71. McCulloch P, Badenoch D. Finding and appraising evidence. *Surg Clin North Am*. 2006;86: 41-57. viii.
72. Ahmed K, Ashrafian H. Life-long learning for physicians. *Science*. 2009;326:227.
73. Ahmed K, Jawad M, Dasgupta P, Darzi A, Athanasiou T, Khan MS. Assessment and maintenance of competence in urology. *Nat Rev Urol*. 2010;7:403-413.
74. Arora S, Sevdalis N, Suliman I, Athanasiou T, Kneebone R, Darzi A. What makes a competent surgeon?: experts' and trainees' perceptions of the roles of a surgeon. *Am J Surg*. 2009;198: 726-732.
75. Satava RM, Gallagher AG, Pellegrini CA. Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. *J Am Coll Surg*. 2003;196:933-937.

76. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. *PLoS Clin Trials*. 2006;1:e9.
77. Marshall JC. Surgical decision-making: integrating evidence, inference, and experience. *Surg Clin North Am*. 2006;86:201-215. xii.
78. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry*. 2002;59:877-883.
79. Sevdalis N, McCulloch P. Teaching evidence-based decision-making. *Surg Clin North Am*. 2006;86:59-70. viii.

http://www.springer.com/978-0-85729-175-2

Evidence Synthesis in Healthcare
A Practical Handbook for Clinicians
Athanasiou, T.; Darzi, A. (Eds.)
2011, XVIII, 326 p., Hardcover
ISBN: 978-0-85729-175-2