Randomised controlled trials: important but overrated?

[1]JF Boylan, [2]BP Kavanagh, [3]J Armitage
[1]*Consultant Anaesthetist, St. Vincent's University Hospital, Dublin and Honorary Lecturer in Clinical Physiology, University College Dublin, Dublin, Ireland;* [2]*Professor and Chair, Department of Anesthesia, Hospital for Sick Children, University of Toronto, Toronto, Canada;* [3]*Professor of Clinical Trials and Epidemiology and Honorary Consultant in Public Health Medicine, Clinical Trial Service Unit & Epidemiological Studies Unit, University of Oxford, Oxford, UK*

Practising physicians individualise treatments, hoping to achieve optimal outcomes by tackling relevant patient variables. The randomised controlled trial (RCT) is universally accepted as the best means of comparison. Yet doctors sometimes wonder if particular patients might benefit more from treatments that fared worse in the RCT comparisons. Such clinicians may even feel ostracised by their peers for stepping outside treatments based on RCTs and guidelines. Are RCTs the only acceptable evaluations of how patient care can be assessed and delivered?

In this controversy we explore the interpretation of RCT data for practising clinicians facing individualised patient choices. First, critical care anaesthetists John Boylan and Brian Kavanagh emphasise the dangers of bias and show how Bayesian approaches utilise prior probabilities to improve posterior (combined) probability estimates. Secondly, Jane Armitage, of the Clinical Trial Service Unit in Oxford, argues why RCTs remain essential and explores how the quality of randomisation can be improved through systematic reviews and by avoiding selective reporting.

KEYWORDS Randomised controlled trials, Bayesian analysis, subgroup analysis, clinical trials, bias, confounders

DECLARATION OF INTERESTS No conflict of interests declared.

Correspondence to **JF Boylan,
St Vincent's University Hospital,
Elm Park, Dublin 4,
Ireland**

tel. **+353 (0)1221 4262**
e-mail **jboylan@iol.ie**

**J Armitage
Clinical Trial Service Unit,
Richard Doll Building (RDB)
University of Oxford,
Oxford, OX3 7LF
UK**

tel. **+44 (0)1865 743 743**
e-mail
jane.armitage@ctsu.ox.ac.uk

Randomised controlled trials in critical care: caveat emptor

JF Boylan, BP Kavanagh

Several decades ago Yusuf, Collins and Peto made a strong case for large, simple randomised controlled trials (RCTs)[1], and since then, several RCTs (e.g., thrombolysis in acute myocardial infarction) have produced results of great importance. However, there are concerns about RCTs, including high negativity rates, uncertainties in interpretation, and an increasing recognition that treatments, naturally, have varying impacts depending on individual patients and the severity of the disease. These concerns may be especially important in critical care, an area with high mortality, significant physiological variation, and a low anticipated yield in survival benefit. At the same time, some have suggested that trial results are best interpreted by 'methodologists' – as the best-trained and most objective assessors of evidence;[2] this is, possibly, a paradox, given that the 'large simple trial' is supposed – by use of massive numbers – to keep interpretation simple. Here we focus on four elements: prior probability, trial bias, Bayesian analysis; and the application of trial results to diverse patients. We believe that appreciation of these issues could enhance the utility of RCTs in critical care.

PRIOR PROBABILITY

Browner and Newman have compared RCTs in populations with diagnostic tests in individuals.[3] In both cases, we note the pretest probability, and, after the data are in, we then decide about our degree of belief. However, while we usually know the pretest probability (i.e. the prevalence) of disease before a diagnostic test, we almost never know the prior probability (π) of scientific hypotheses being true in an impending RCT. We are therefore, unable to place the results of the RCT in the context of our population of interest. If the lessons from diagnostics hold, usual RCT interpretation is likely to be flawed.

Attempts have been made to determine prior probability (in retrospect): the overall survival for experimental groups in a series of oncology trials was 0.96 (0.89–1.03; mean, 99% confidence intervals), suggesting that a prior probability of the tested hypotheses (that is, that the experimental intervention was beneficial) being true was almost zero.[4] A consistent negative trend in critical care trials suggests that it is also low there.[5]

What is the impact of π on interpretation of an RCT? Very low values mean that the chance of any given 'significant' result being a 'true positive' becomes very low, even with flawless study methodology. For 100,000 RCTs with $\pi=0.01$ (so 1,000 true hypotheses out of 100,000 tested) and with the usual $\alpha=0.05$ and $\beta=0.20$, we could expect 5,000 false positives, 800 true positives, 200 false negatives and 94,000 true negatives. So where RCTs are 'pragmatic' (and especially where this equates

to minimal reliance on physiological mechanisms) most positive results will be false positives.

Recognition of low π in RCTs may suggest that 'most published research findings are false'.[6] If a given π were high, a 'positive' trial would likely be a 'true positive'; however, apart from trials designed to test a (truly) promising hypothesis, most π values would seem to be low.

QUALITY AND BIAS

Delivery of objective results hinges on eliminating experimental bias. The most widely accepted approach to confirm that bias is minimised is to grade the RCT methodology; quality scoring systems or methodological ratings might help the reader decide what to believe. There is some support for quality scoring: higher quality studies have lower effect-size estimates (perhaps suggesting less unrealistic findings), and non-randomised studies are more likely to be contradicted or have their effect estimate downgraded compared with randomised trials. But high 'grades' do not reduce false positive rates, and randomised designs are far from resistant to refutation. Thus no single aspect of study design allows us to draw robust conclusions.

Bias may be of most concern in 'positive' studies. There are several causes. Allocation bias (more 'less sick' patients in the treatment group improve outcome, independently of the treatment) involves manoeuvres, not all of which are detectable, which subvert randomisation and allocation concealment. Superior care in the treatment group could involve post-treatment co-interventions because of altruism in the context of a 'risky' therapy (e.g. increased vigilance); the corollary, inferior care in the control group, is also a possibility. Finally, biased analysis or presentation of data may be subtle, potentially reflecting prior beliefs or commercial pressure.

Among the most controversial questions in critical care in the last few years has been the role of intensive insulin therapy in reducing intensive care mortality. A single centre study suggested a major reduction in mortality,[7] and after several failures to replicate this finding, a multicentre study suggested a significant worsening in mortality;[8] the 'quality' scores[9] in both studies were identical,[10] suggesting that 'quality' does not help us decide which trial, if either, is 'true'.

Randomised, concealed allocation, the cornerstone of trial validity, can be interfered with for benign motives. For example, in a trial investigating a therapy that cannot be blinded, sicker patients may be diverted into the treatment arm if caregivers 'believe' in the intervention. Of course, diverting sicker patients toward the control group would be much more harmful, since this results in a bias towards treatment even if the therapy is not efficacious. Thus even partial knowledge of a random-

isation schedule could influence a patient's group entry, and the traditional display of P values provides no protection against this.[11] No one has modelled biased allocation of patients in trials, but simulation data suggest that even small-scale (2%) transfer of sicker patients out of an intensive care unit causes marked improvements in survival standardised mortality ratio.[12]

A BAYESIAN VIEW

All clinicians are Bayesians in everyday practice, interpreting data in the light of what they already know (or believe). Reviewing test and RCT results, they think 'how likely is the diagnosis/hypothesis to be correct, given the data?'. Trialists use P values – 'how likely are these data if the null hypothesis is true?' (not the same thing) – to persuade us of their conclusions. Many RCT results are not definitive, but it is hard to quantify this. Attempts to estimate 'strength of evidence' based on quality ratings and hierarchical scoring reflect this problem.

To estimate the posterior probability that a hypothesis is true, trialists need to combine a prior probability with experimental data. The prior probability may be based on data already available, or may reflect a belief (expressed as a probability that π lies within a 'null range'). Because both the prior probability and the data can be expressed as distributions with means and standard deviations, we can generate a posterior distribution from them,[13] giving us probabilities for any given treatment effect,[14] including probabilities for a minimum treatment effect.

Bayesian interpretation of RCTs has not gained full acceptance, because of 'subjectivity' concerns, but looking at trial data with prior probability in mind gives us useful insights. The sensitivity of a trial result to the impact of different π values allows us to interpret it cautiously. Assuming no other pitfalls, how likely the hypothesis is to be true after the data are in depends on both the robustness of the data (sample size, event rates) and the prior probability of the hypothesis. Knowing that π in a research field is low may lead clinicians in that field to adopt at least a mildly sceptical stance when assessing a trial. In a Bayesian analysis of the most contentious recent critical care trials, Kalil showed that their interpretation was highly sensitive to the level of scepticism; a reluctance to implement RCT results may be driven by a (very reasonable) belief that the knowledge is provisional.[15]

An important unresolved issue with RCTs is that negative results do not rule out clinically relevant benefit, and positive results do not mean that the reported benefit, an aggregate value, is the most likely benefit. Current trial analysis and reporting usually ignores this. In internal medicine, Wijesundera et al. have pointed out that the efficacy estimates in high-profile positive RCTs were very sensitive to the

EDUCATION

assumed π, and many negative studies were compatible with moderate probability of at least some benefit.[14] If we want to know the state of the investigators' knowledge at the end of a positive RCT, they should be able to tell us, for example, 'this treatment has an 80% chance of producing a relative risk reduction of 0.2 or better', and to give us an idea of how this estimate varies, depending on assumptions about π.

HOW MUCH BENEFIT IN THIS PATIENT?

There is a widespread belief among methodologists that subgroups within RCTs are problematic, and that subdividing data for analysis does more harm than good, a stance backed up by statements, unsupported by data, that quantitative differences (different degrees of benefit) are rare and qualitative differences (benefit in some groups, harm in others) are very rare.[16] Some authorities actively discourage collecting such data: '…less information may mean bigger numbers and hence better science: many trials still collect ten or a hundred times too much information per patient…(we) need to find ways of making trials much simpler and much larger…'.[17]

If the correct approach to subgroup differences is to assume by default that they do not exist, or do not matter, then we can justify subgroup detection tests with a very 'high bar'; such an approach was used in CRASH-2, an RCT suggesting that tranexamic acid reduces all-cause mortality in major trauma involving 'significant haemorrhage'.[18] The trial reported a net improvement in survival, but no difference in blood product usage (half of the patients were not transfused), a finding attributed – without supporting data – to non-survivors being transfused less than survivors. Univariate subgroup analyses (time since injury, blunt vs penetrating trauma, Glasgow coma score, admission systolic blood pressure) using a very stringent threshold ($p<0.001$) found no differences. This led the investigators to conclude that tranexamic acid conferred a general survival benefit; their analysis did not exclude benefit to patients with trivial bleeding, even though most patients (in the control group) did not die, a minority (12.5%) had large-volume transfusion and the relationship between bleeding, transfusion and mortality risk was not presented. It has been shown that risk stratification using simplistic, one-factor-at-a-time analyses limits power to detect subgroup differences, while looking at baseline risk using multivariate stratification provides a much wider variation in baseline risk (10–20 fold, similar to real life illnesses), and more statistical power as a result.[19] Such a multivariate analysis of the data from CRASH-2, and other trials, might provide useful insights into which patients benefit most, and, possibly, which ones do not benefit at all. The current analysis does not tell us if there is a threshold level of injury severity above which patients benefit; in fact it suggests (implausibly) that there is not.

All benefits involve trade-offs, and when the focus is the population, rather than the individual, the 'net' conclusion may be what matters most. But for such 'net' benefits to apply to individuals, the risk/benefit ratio has to be constant across the population. If baseline risk varies widely, the absolute risk reduction will vary widely also,[20] the results may be determined to a disproportionate extent by the outcome in one group (often the highest risk patients) and will be less relevant to 'typical' patients. In the most extreme cases benefits in one subgroup may 'drive' the result for the trial.[21] Most current trials do not explore this optimally,[20] even though applicability of a trial to individual patients is the main reason for carrying trials out. The subgroup effect is not limited to large trials: a re-analysis of a phase II trial of eritoran (a candidate anti-sepsis therapy) that initially reported no net benefit,[22] demonstrated that the efficacy of eritoran was strongly and non-linearly related to baseline risk of death (odds of survival [treatment group] = 1.84 x odds of mortality [control group] [0.77]) when admission severity of illness scores were used.[23] The message is clear: differential drug effects may be detectable if looked for, and may range from benefit in some to harm in others.

By avoiding data on effect modifiers, or using analytical techniques that are underpowered, harm could result from generalising the conclusions to all patients. If experimental therapies are given to patients not 'at risk' in the first place, such trial participants will, at best, have been just making up the numbers. There is a strong case, articulated by Kent et al., that trialists should assume that summary estimates do not apply to individual patients, and should seek out heterogeneity of treatment effect using prospectively defined methods.[24] In the real world, this concern is reinforced by a recent paper suggesting that compliance with a treatment guideline for healthcare associated pneumonia worsened survival.[25] This finding surprised the authors, but not a commentator, who suggested that empirical therapy and failure to understand patient heterogeneity were the main reasons why compliance was harmful.[26]

GOING FORWARD

Critical illness is a collective term for our lack of understanding of many acute disease states. It could be argued that some of the obvious, pragmatic (that is, non-mechanistic) questions in critical care (such as positive end-expiratory pressure, dialysis dosage, global oxygen delivery, gastrointestinal prophylaxis) have now been answered in the negative, suggesting a very low π for the field. But this reflects a low π for the hypotheses underlying 'one size fits all' therapies. Unless we accept the challenges inherent in studying seriously ill patients, with highly variable physiology and the likelihood of variable treatment-induced effects (ranging from benefit to harm), we can expect continuing difficulty. A trial – even a 'simple' one – of any intervention in complex

biological processes will involve both titration (such as: how much should we modulate a pathway that exists for a reason?) and stratification, which involves working out a priori who is unlikely to benefit, rather than studying everyone and deciding in hindsight who is harmed. We should not regard such an approach as a 'simple trial of … a widely practicable treatment'; nor should we leave its results to methodologists. We know enough about trials, and their limitations, to know better.

REFERENCES

1 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; 3:409–20. doi:10.1002/sim.4780030421

2 Hirsh J, Guyatt G. Clinical experts or methodologists to write clinical guidelines? *Lancet* 2009; 374:273–5. doi:10.1016/S0140-6736(09)60787-X

3 Browner WS, Newman TB. Are all significant P values created equal? The analogy between diagnostic tests and clinical research. *JAMA* 1987; 257:2459–63. doi:10.1001/jama.257.18.2459

4 Kumar A, Soares H, Wells R et al. Are experimental treatments for cancer in children superior to established treatments? Observational study of randomized controlled trials by the Children's Oncology Group. *BMJ* 2005; 331:1295. doi:10.1136/bmj.38628.561123.7C

5 Ospina-Tascon GA, Büchele GL, Vincent J-L. Multicenter, randomized, controlled trials evaluating mortality in intensive care: Doomed to fail? *Crit Care Med* 2008; 36:1311–22. doi:10.1097/CCM.0b013e318168ea3e

6 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; 2:e124. doi:10.1371/journal.pmed.0020124

7 Van den Berghe G, Wouters P, Weekers F et al. Intensive insulin therapy in the critically ill patients. *N Engl J Med* 2001; 345:1359–67. doi:10.1056/NEJMoa011300

8 The NICE-SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med* 2009; 360:1283–97. doi:10.1056/NEJMoa0810625

9 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. doi:10.1016/0197-2456(95)00134-4

10 Griesdale DEG, de Souza RJ, van Dam RM et al. Intensive insulin therapy and mortality among critically ill patients: a meta-analysis including NICE-SUGAR study data. *CMAJ* 2009; 180:821–7. doi:10.1503/cmaj.090206

11 Berger VW, Exner DE. Detecting selection bias in randomized clinical trials. *Control Clin Trials* 1999; 20:319–27. doi:10.1016/S0197-2456(99)00014-8

12 Kahn JM, Kramer AA, Rubenfeld GD. Transferring critically ill patients out of hospital improves the standardized mortality ratio. *Chest* 2007; 131:68–75. doi:10.1378/chest.06-0741

13 Spiegelhalter DJ, Myles JP, Jones DR et al. An introduction to bayesian methods in health technology assessment. *BMJ* 1999; 319:508–12.

14 Wijeysundera DN, Austin PC, Hux JE et al. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *J Clin Epidemiol* 2009; 62:13–21.e5

15 Kalil AC, Sun J. Why are clinicians not embracing the results from pivotal clinical trials in severe sepsis? A bayesian analysis. *PLoS One* 2008; 3:e2291. doi:10.1371/journal.pone.0002291

16 Peto R. Statistical aspects of cancer trials. In: Halnan KE, editor. *Treatment of cancer*. London: Chapman and Hall; 1982. p.867–71.

17 Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 1998; 317:1170-1.

18 CRASH-2 trial collaborators. Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with significant haemorrhage (CRASH-2): a randomised, placebo-controlled trial. *Lancet* 2010; 376:23–32. doi:10.1016/S0140-6736(10)60835-5

19 Hayward RA, Kent DM, Vijan S et al. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol* 2006; 6:18. doi:10.1186/1471-2288-6-18

20 Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA* 2007; 298:1209–12. doi:10.1001/jama.298.10.1209

21 Ioannidis JPA, Lau J. The impact of high-risk patients on the results of clinical trials. *J Clin Epidemiol* 1997; 50:1089–98. doi:10.1016/S0895-4356(97)00149-2

22 Tidswell M, Tillis W, LaRosa SP et al: Phase 2 trial of eritoran tetrasodium (E5564), a Toll-like receptor 4 antagonist, in patients with severe sepsis. *Crit Care Med* 2010; 38:72–83. doi:10.1097/CCM.0b013e3181b07b78

23 Barochia A, Cui X, Natanson C et al. Risk of death and the efficacy of eritoran tetrasodium (E5564): design considerations for clinical trials of anti-inflammatory agents in sepsis. *Crit Care Med* 2010; 38: 306–8. doi:10.1097/CCM.0b013e3181b77fe3

24 Kent DM, Rothwell PM, Ioannidis JPA et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials* 2010; 11:85. doi:10.1186/1745-6215-11-85

25 Kett DH, Cano E, Quartin AA et al. Implementation of guidelines for management of possible multidrug-resistant pneumonia in intensive care: an observational, multicentre cohort study. *Lancet Infect Dis* 2011; 11:181–9. doi:10.1016/S1473-3099(10)70314-5

26 Yu J. Guidelines for hospital-acquired pneumonia and healthcare-associated pneumonia: a vulnerability, a pitfall, and a fatal flaw. *Lancet Infect Dis* 2011; 11:248–52. doi:10.1016/S1473-3099(11)70005-6

Randomised trials remain essential and are not overrated

J Armitage

INTRODUCTION

Randomised controlled trials (RCTs) lie at the heart of the practice of evidence-based clinical medicine and are accepted as the most reliable form of evidence. Results from RCTs have led to both the widespread use of effective treatments and a reduction in the use of worthless or harmful treatments. Examples include trials such as ISIS-2, which showed the benefits of aspirin and streptokinase in acute myocardial infarction,[1] and more recently the large randomised statin trials, which first individually[2,3] and then collectively,[4] showed reductions in coronary events and strokes in a wide range of people.

These trials led to rapid changes in practice, widespread use of these effective treatments and undoubtedly avoided large numbers of premature deaths and morbidity and have produced significant health improvements.

However, despite a very large number of trials having been undertaken over the last 50 years, the trials that have delivered major health improvements represent only a minority and are generally large trials counting clinically relevant outcomes. Unfortunately, many trials are poorly designed and reported, often with too few patients to exclude chance effects or inadequately randomised and followed-up to be sure of minimising other biases. The

challenge over the coming years is not to dismiss RCTs as overrated, but to improve the quality and particularly the size of randomised trials so that they continue to deliver valuable information to improve patient care. But how can we do this and are there alternative strategies?

IMPROVING QUALITY

Efforts are ongoing on many fronts to try to improve the quality of randomised evidence. Although randomisation in trials reduces systematic biases, only trials in populations which include large numbers of clinically relevant outcomes can produce reliable results.[5] Systematic reviews of all the randomised evidence in particular disease or therapeutic areas is one way of increasing the number of endpoints available to be analysed. Provided there are a number of good quality trials in a particular area, such reviews (particularly if done in collaboration with the trialists) allow clear treatment effects to be demonstrated in a range of different types of patient. Examples include the collaborations reviewing anti-thrombotic treatments, blood pressure lowering therapy and the early treatment of breast cancer.[6–8] Systematic reviews also illuminate areas where there is little evidence of value and so stimulate the initiation of new trials. But, frequently, such reviews reveal large numbers of inadequate trials sometimes also with clear evidence that not all the trials in that area have been published. In an effort to minimise this latter problem of selective reporting of trials with interesting or positive results and failure to report negative trials, major efforts have been made over many years to ensure that all RCTs are registered and the results of trials made public.[9] This should allow future reviews to be able to encompass all relevant randomised evidence and at least minimise publication bias.

Systematic reviews also highlight deficiencies in the way that trials are reported. The CONSORT group have been trying over many years to improve reporting standards[10] but there is still room for improvement. It was hoped that improving reporting would feed back to improve trial design but there is little evidence to support this, particularly in terms of the size of trials. A recent review of 600 randomised trials reported in PubMed journals in December 2006,[11] ten years after the first CONSORT paper, showed that the median number of patients randomised in these trials was still only 62 (10th–90th centile: 19–392) illustrating that the majority of trials are small and unlikely, therefore, to be having any material impact on clinical practice. So, despite some recent progress, there is still a way to go to make most trials informative.

Randomised trials of medicinal agents are currently subject to what many consider to be excessive regulation,[12–14] partly as a result of the implementation of the European Union Directive on Clinical Trials and partly, in the UK, because of NHS Research Governance requirements. Although well intentioned, these regulations mean that significant amounts of time and effort are spent by those running trials ensuring compliance with the various regulations, thereby detracting from the efficient running the trial. This has led to higher costs and complexity of running trials which typically results in fewer patients being recruited and therefore less reliable studies. Again, active efforts are being made on several fronts to improve the situation.[13]

ARE THERE RELIABLE ALTERNATIVES TO RCTS?

The aim of a randomised comparison is to achieve an unbiased assessment of the intervention or drug treatment being evaluated. Randomisation alone doesn't guarantee this and, to minimise bias, blinding of outcome assessment, complete follow-up in each randomised group and intention-to-treat analyses are required.

Observational studies, either prospective or case-control can, in some circumstances, provide reliable results about effects of treatments, particularly if effect sizes are large, but confounding remains an issue if effect sizes are moderate.[15] Routinely recorded information gathered into databases is increasingly providing an opportunity to readily investigate the quantitative effects of treatments. Such databases can be useful for looking at treatment patterns, however, as it is impossible to fully eliminate confounding by unmeasured (and unmeasurable) characteristics of both patients and prescribing doctors, such non-randomised analyses risk serious bias if used to try to measure treatment effects. Examples include claims based on routinely collected GP data that statins reduce the risk of dementia[16] or increase the risk of cataracts[17] neither of which are borne out by data from large randomised studies and probably represent problems of residual confounding. So, observational studies generally do not allow for reliable assessments of treatment effects.

CAN WE USE SURROGATE MARKERS RATHER THAN CLINICAL OUTCOMES?

Epidemiologists spend much effort searching for risk factors for disease and in many well-established circumstances (such as blood pressure and risk of stroke) interventions which reduce the risk factor reduce the related outcome in line with expectations. However, there are also numerous examples in the literature of strong, consistent associations in observational studies between risk factors or biomarkers and disease outcomes which have not been shown to be causal in randomised trials. A recent example was the promising positive association between blood homocysteine levels and risk of cardiovascular disease which in observational studies remained clearly significant even after controlling for other cardiovascular risk factors.[18] However, several large randomised trials of lowering blood homocysteine levels with B vitamins have failed to demonstrate any effect on cardiovascular events despite significant reductions in homocysteine levels.[19]

In the past, blood pressure-lowering, lipid-lowering and many other drugs have been licensed based on small randomised trials showing reductions in risk factor or biomarker levels. But recent experience with the cholesteryl ester transfer protein inhibitor, torcetrapib, which, despite reducing low-density lipoprotein cholesterol and increasing high-density lipoprotein cholesterol, caused harm not benefit, suggests that this strategy is not always a safe one.[20] Torcetrapib was not licensed and appropriately was only being used in the context of randomised trials, but there are several past examples of treatments in common use which were later shown to be harmful in large randomised trials. To name a few: anti-arrhythmic drugs to suppress ventricular tachycardias,[21] hormone replacement therapy[22] and steroids in head injury.[23] One wonders how many treatments are currently in widespread use without randomised evidence of effects on longer-term clinical outcomes which might be causing harm – perhaps non-steroidal anti-inflammatory drugs.[24]

CONCLUSION AND FUTURE PERSPECTIVES

Randomised trials are the cornerstone of the current practice of medicine and are a powerful lever for change. As illustrated above, experience indicates that there are no shortcuts and that large randomised trials are essential for reliably assessing the clinical effects of treatments. However, these large trials are increasingly expensive and complex to undertake and innovative methods, perhaps better exploiting electronic media, need to be found to allow the development of large cost-effective randomised trials of many more interventions. Steps are also urgently needed to simplify the regulation of clinical trials to ensure that they remain viable. Randomisation is now the accepted norm in healthcare evaluation and mandatory before most new interventions are introduced. This acceptance is a major advance which needs to be built on and strengthened rather than dismissed. Perhaps, wider use of randomisation in other areas of life might provide a more robust evidence base to inform policy in other areas.[25]

REFERENCES

1 ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet* 1988; 2:349–60.

2 Randomised trial of cholesterol lowering in 4,444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* 1994; 344:1383–9.

3 Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; 360:7–22. doi:10.1016/S0140-6736(02)09327-3

4 Cholesterol Treatment Trialists' Collaboration. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* 2005;366:1267–78. doi:10.1016/S0140-6736(05)67394-1

5 Peto R, Baigent C. Trials: the next 50 years. Large scale randomised evidence of moderate benefits. *BMJ* 1998; 317:1170–1.

6 Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 2002;324:71–86. doi:10.1136/bmj.324.7329.71

7 Collins R, Peto R, MacMahon S et al. Blood pressure, stroke, and coronary heart disease. Part 2, Short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990;335:827–38. doi:10.1016/0140-6736(90)90944-Z

8 Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. *Lancet.* 1992; 339:71–85.

9 DeAngelis CD, Drazen JM, Frizelle FA et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *JAMA* 2004; 292:1363–4. doi:10.1001/jama.292.11.1363

10 Begg C, Cho M, Eastwood S et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; 276:637–9. doi:10.1001/jama.276.8.637

11 Hopewell S, Dutton S, Yu LM et al. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ* 2010; 340:c723. doi:10.1136/bmj.c723

12 Stewart PM, Stears A, Tomlinson JW et al. Regulation – the real threat to clinical research. *BMJ* 2008; 337:a1732. doi:10.1136/bmj.a1732

13 Haynes R, Bowman L, Rahimi K et al. How the NHS research governance procedures could be modified to greatly strengthen clinical research. *Clin Med* 2010; 10:127–9.

14 Academy of Medical Sciences (AMS). A new pathway for the regulation and governance of health research. London: AMS; 2011. Available from: http://www.acmedsci.ac.uk/p47prid88.html

15 Brown ML, Gersh BJ, Holmes DR et al. From randomized trials to registry studies: translating data into clinical information. *Nat Clin Pract Cardiovasc Med* 2008; 5:613–20. doi:10.1038/ncpcardio1307

16 Jick H, Zornberg GL, Jick SS et al. Statins and the risk of dementia. *Lancet.* 2000; 356:1627-31. doi:10.1016/S0140-6736(00)03155-X

17 Hippisley-Cox J, Coupland C. Unintended effects of statins in men and women in England and Wales: population based cohort study using the QResearch database. *BMJ* 340:c2197. doi:10.1136/bmj.c2197

18 Homocysteine Studies Collaboration. Homocysteine and risk of ischemic heart disease and stroke: a meta-analysis. *JAMA* 2002; 288:2015–22. doi:10.1001/jama.288.16.2015

19 Clarke R, Halsey J, Lewington S et al. Effects of lowering homocysteine levels with B vitamins on cardiovascular disease, cancer, and cause-specific mortality: Meta-analysis of 8 randomized trials involving 37 485 individuals. *Arch Intern Med* 2010; 170:1622–31. doi:10.1001/archinternmed.2010.348

20 Barter PJ, Caulfield M, Eriksson M et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med* 2007; 357:2109–22. doi:10.1056/NEJMoa0706628

21 The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* 1989; 321:406–12. doi:10.1056/NEJM198908103210629

22 Beral V, Banks E, Reeves G. Evidence from randomised trials on the long-term effects of hormone replacement therapy. *Lancet* 2002; 360:942–4. doi:10.1016/S0140-6736(02)11032-4

23 Edwards P, Arango M, Balica L et al. Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous corticosteroid in adults with head injury-outcomes at 6 months. *Lancet* 2005; 365:1957–9.

24 Hermann M, Ruschitzka F. Cardiovascular risk of cyclooxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs. *Ann Med* 2007; 39:18–27. doi:10.1080/07853890601073445

25 Macintyre S. Good intentions and received wisdom are not good enough: the need for controlled trials in public health. *J Epidemiol Community Health* 2010; epub ahead of print 8 Dec. doi:10.1136/jech.2010.124198

EDUCATION