The evolution of the randomized controlled trial and its role in evidence-based decision making

P. J. DEVEREAUX & S. YUSUF
From the Department of Medicine and Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada


The randomized controlled trial has been used in medical research for a little over half a century. This manuscript provides an overview of some of the history and evolution of the randomized controlled trial during this period. There exists hierarchies of evidence for therapeutic, diagnostic and prognostic questions, and the randomized controlled trial is at the top of the therapeutic hierarchy. Despite being at the top of the therapeutic hierarchy randomization in itself does not guarantee the trial results approximate the true effect. Issues that result in systematic and nonsystematic deviations from the truth in randomized controlled trials must also be considered. We present a model for evidence-based decision making that includes the following components: the clinical state, patient preferences, research evidence from a range of studies and clinical expertise. We discuss the role of the randomized controlled trial within evidence-based decision making.

Keywords: evidence-based medicine, hierarchies of evidence, randomized controlled trial.

Evolution of randomized controlled trials

Fifty-five years ago, the publication of the British Medical Research Council trial of streptomycin in patients with pulmonary tuberculosis formally heralded the arrival of the randomized controlled trial (RCT) into medical research [1]. Despite this historical study, medical researchers were initially slow to undertake RCTs. However, the last two decades have witnessed exponential growth in the use of RCTs in clinical research. In the late 1980s, there were approximately 5000 RCTs published each year, and by 1998 the annual publication rate had more than doubled surpassing 12 000 publications of RCTs per year [2]. This growth relates to the acceptance that properly conducted RCTs provide our best means to assess the efficacy of an intervention [3].

The first RCTs were primarily in the area of infectious disease and these trials sought large treatment effects in patients at high risk of clinically important outcomes. Given these conditions, small trials were adequate (e.g. RCT of streptomycin in tuberculosis). Later as investigators evaluated preventive interventions, even in the area of infectious diseases, because of the low rates of events,
large trials were needed (e.g. the polio vaccine trial randomized more than 400,000 children to active vaccine or placebo) [4]. The application of the experiences of infectious disease RCTs that targeted large treatment effects to chronic diseases (where the pathophysiology is usually multifactorial hence any intervention targeting a single risk factor is unlikely to have more than a moderate treatment effect) were disappointing because whilst they excluded large treatment effects, they could not reliably detect or exclude moderate but important effects (e.g. beta-blockers in acute myocardial infarction) [5]. Throughout the last 20 years, fundamental changes in the design and outcomes of RCTs have occurred. There has been a shift away from small RCTs that evaluate surrogate outcomes (e.g. physiological outcomes) towards large simple RCTs that evaluate major clinical outcomes (e.g. death, stroke). This change was partly fuelled by greater understanding of key criteria that were critical to the conduct of a reliable trial; namely addressing important questions on reducing mortality or morbidity by the conduct of large reliable trials (Table 1) [6]. These criteria suggested the desirability and possibility of large, simple RCTs that evaluated major clinical outcomes [6].

The shift away from small inconclusive RCTs that evaluated surrogate outcomes towards large simple RCTs that evaluated major clinical outcomes was also supported by several examples where there have been contradictory results across small and large RCTs and studies assessing surrogate outcomes and major clinical outcomes (Table 2).

Hierarchies of research evidence

Clinical decisions involve issues of diagnosis, prognosis and/or therapy. The research evidence required to address these issues can be observational or experimental, based on whether the researchers assign the exposure or not [26]. The research evidence can also result from systematic reviews or meta-analyses.

There are two types of observational studies: descriptive studies [there is no comparison group (e.g. case series)] and analytic studies [there is a comparison group (e.g. cohort study, case–control study, cross-sectional study)] [26]. There are two types of experimental studies: RCTs [patients are assigned to the interventions based on a truly random process (e.g. flipping of a coin, a computer generated random allocation sequence)]; and non-randomized controlled trials [patients are assigned to the interventions being evaluated based on a non-random process (e.g. admission date, hospital number, alternate assignments)] [27].

A systematic review is a study that addresses a focused research question, has explicit eligibility criteria, and undertakes an extensive search of the literature to identify studies that fulfill the eligibility criteria [28]. A meta-analysis is the pooling of data across two or more studies [29]. A meta-analysis can pool results from any studies, however high-quality meta-analyses only pool data from systematic reviews.

There exists an underlying truth to all of the clinical questions that are considered. For example, there is a true underlying impact of clopidogrel on mortality in patients with acute myocardial infarction showing ST segmental elevation. In order to make informed evidence-based decisions, the results of the research that are relied upon should provide the closest approximation to the truth possible. Within evidence-based decision making, there has been the realization that not all research evidence is equal. As a result of this hierarchies of evidence

Table 1 Arguments for large simple randomized controlled trials that evaluate major clinical outcomes

<table>
<thead>
<tr>
<th>Criteria impacted</th>
<th>Argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>From a population perspective it is more ‘important’ to establish an effective treatment for a common disease, as opposed to a rare disease, and studies of common diseases can be large in scale.</td>
</tr>
<tr>
<td>Importance</td>
<td>From a population and patient perspective it is more ‘important’ to establish an effective widely practicable therapy, as opposed to a nonpracticable therapy (i.e. interventions that are only possible in very restricted settings), and studies of practicable therapies can be simple.</td>
</tr>
<tr>
<td>Importance</td>
<td>From a patient perspective it is more ‘important’ to evaluate major clinical outcomes, as opposed to surrogate outcomes, and the follow-up protocols in studies evaluating major clinical outcomes can be simple.</td>
</tr>
<tr>
<td>Reliability</td>
<td>Most treatment effects are moderate in size and for ‘reliable’ detection of such effects studies must be large in scale.</td>
</tr>
</tbody>
</table>
have been established, based on study design. Table 3 provides an example of the hierarchy of evidence for therapeutic research questions. The hierarchy of evidence will vary depending on whether one is addressing a diagnostic, prognostic, or therapeutic question.

Although there is some debate about the order within the hierarchies there has been a profound shift to move RCTs and systematic reviews of RCTs to the top of the therapeutic hierarchy. The placement of RCTs at the top of the therapeutic research hierarchy has occurred due to the realization that RCTs are superior to observational studies in evaluating treatments because RCTs eliminate bias in the choice of treatment assignments and provide the only means to control for unknown prognostic factors [30]. Further there have been several examples where there have been contradictory findings between observational studies and RCTs (Table 4).

Table 2 Examples of contradictory results across small and large randomized controlled trials (RCTs) and trials assessing surrogate versus major clinical outcomes

<table>
<thead>
<tr>
<th>Results from lower level evidence (i.e. small RCTs or studies evaluating surrogate outcomes)</th>
<th>Results from higher level evidence (i.e. large RCTs or meta-analysis of large RCTs evaluating major clinical outcomes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A meta-analysis of several small RCTs of magnesium in acute myocardial infarction demonstrated a statistically significant 50% reduction in death with approximately 1000 patients randomized and a moderate number of deaths (i.e. 128) ($P &lt; 0.001$) [7].</td>
<td>A large RCT of 58 050 patients demonstrated no benefit, in fact there was a trend towards excess mortality with magnesium ($P = 0.07$) compared to placebo in patients with acute myocardial infarction [8]. Most recently a second large RCT of 6213 patients demonstrated no mortality benefit with magnesium [9].</td>
</tr>
<tr>
<td>An RCT that compared the efficacy of an angiotensin II receptor blocker (ARB) to an angitensin-converting enzyme (ACE) inhibitor, in patients with NYHA II–IV heart failure, demonstrated a statistically significant 46% reduction in death with ARB treatment ($P = 0.03$) [10]. This trial randomized 722 patients but had a very small number of deaths (i.e. 49).</td>
<td>A large RCT of 3152 patients demonstrated no mortality benefit for ARB therapy compared with ACE inhibitor therapy, in fact there was a trend in the opposite direction compared with the earlier study ($P = 0.16$) [11]. Further, a large RCT of 5477 patients in postmyocardial infarction patients who had congestive heart failure also demonstrated a trend towards higher mortality with an ARB compared with an ACE inhibitor ($P = 0.07$) [12].</td>
</tr>
<tr>
<td>In a study evaluating milrinone, this drug demonstrated improvement in left ventricular function during exercise [13].</td>
<td>An RCT [14] of 1088 patients and a meta-analysis of several RCTs [15] demonstrated a 28% relative increase in mortality with milrinone compared with placebo. RCT was stopped early after 1906 patients were enrolled, because of a 28% relative increase in mortality with ibopamine compared with placebo [17].</td>
</tr>
<tr>
<td>A cross-over RCT of patients with NYHA II–III heart failure demonstrated that ibopamine improved cardiac output, stroke volume and ejection fraction [16]. A physiological study demonstrated beta-blockers result in a decline in ejection fraction and increases in end-diastolic volume in patients with prior myocardial infarction [18].</td>
<td>A meta-analysis of 18 RCTs [19] and three large RCTs (CIBIS –2 [20], MERIT-HF [21] and COPERNICUS [22]) in patients with heart failure found a 32% relative risk reduction in death in patients receiving beta-blockers. An RCT was stopped early after 1498 patients were enrolled because of a demonstrated higher mortality rate with encainide and flecainide compared with placebo in patients postmyocardial infarction with frequent PVCs [25]. The increased risk of death was such that these drugs resulted in an extra death for every 20 patients treated with encainide or flecainide. It is estimate that more Americans were killed by these drugs than died in the Vietnam War [24].</td>
</tr>
<tr>
<td>An observational study demonstrated that premature ventricular complexes (PVCs) after a myocardial infarction increase the risk of sudden death [23]. As a result of this such patients were frequently treated with antiarrhythmic drugs because they suppressed PVCs [24].</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 Hierarchy of evidence for evaluating a therapy (most reliable on top; least reliable at the bottom)

| Systematic review of several large randomized controlled trials |
| Single large randomized controlled trial |
| Systematic review of several small randomized controlled trials |
| Single small randomized controlled trial |
| Systematic review of several cohort studies |
| Single cohort study |
| Systematic review of several case–control studies |
| Single case–control study |
| Systematic review of several cross-sectional studies |
| Single cross-sectional study |
| Case series |

© 2003 Blackwell Publishing Ltd Journal of Internal Medicine 254: 105–113
Systematic reviews of RCTs are thought to belong towards the top of all the research hierarchies because of the belief that the clearest answer to any research question comes from bringing together all of the high-quality studies that have addressed the question of interest. This enhances power and precision factors [40] important in reducing non-systematic deviations from the truth (i.e. random error, a factor which will be discussed below). Also it was the realization that systematic reviews or much larger trials would have established the efficacy of life-saving drugs (e.g. thrombolytic drugs) much earlier (i.e. 15 years earlier for thrombolytic drugs), thereby saving several hundreds of thousands of lives [41].

It is important to realize that the best evidence is the evidence that exists at the highest position within a given hierarchy. Therefore, one should not discard analysis of observational data that has assessed a therapeutic intervention, when no RCTs have evaluated the therapeutic intervention of interest. Observational studies are likely to reliably detect a treatment effect when the true treatment effect is very large (e.g. penicillin for the treatment of streptococcal pneumonia). Although the results of observational studies may have moderate biases (because of selection bias, uneven distribution of unknown covariates, measurement error in covariate values, and inappropriate adjustment in analyses because of mismodelling) [42] as long as these biases and the play of chance (because of small sample sizes) result in errors significantly less than the demonstrated very large treatment effect, an appropriate conclusion about a treatment benefit will be obtained. However, it is important to remember that most interventions and procedures have moderate treatment effects, and although moderate biases may not prevent detection of a treatment effect, when the true effect is extremely large, they are likely to interfere with the detection of a treatment effect when the true treatment effect is moderate in size.

**Pitfalls in the design and conduct of randomized controlled trials**

At any level within a hierarchy, the likelihood that the evidence approximates the truth will vary based on issues of study methodology and size that can result in systematic and non-systematic deviations from the truth [43]. Although the RCT is at the top of the therapeutic hierarchy, randomization in itself does not guarantee the results approximate the truth. Understanding the potential impact of the size

---

**Table 4 Some Examples of contradictory results between observational studies and randomized controlled trials**

<table>
<thead>
<tr>
<th>Results from observational studies</th>
<th>Results from randomized controlled trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>An observational study of extracranial to intracranial bypass surgery suggested a ‘dramatic improvement in the symptomatology of virtually all patients’ undergoing the procedure [31].</td>
<td>An RCT of 1377 patients demonstrated a 14% relative increase in the risk of fatal and nonfatal stroke in patients undergoing this procedure compared with medical management [32].</td>
</tr>
<tr>
<td>A meta-analysis of 16 cohort studies and three cross-sectional angiographic studies (including studies of women with known coronary artery disease) demonstrated a relative risk of 0.5 (95% CI 0.44–0.57) for coronary artery disease amongst women taking estrogen [33].</td>
<td>A secondary prevention RCT of 2763 women did not demonstrate any reduction in coronary heart disease events but did demonstrate an increase in thromboembolic events in patients receiving estrogen (relative hazard 2.89) [34]. A primary prevention RCT (Women’s Health Initiative) of 16 608 women has demonstrated that hormone replacement therapy increases the risk of coronary artery disease [hazard ratio (HR) 1.29], stroke (HR 1.41), pulmonary emboli (HR 2.13), and breast cancer (HR 1.26) [35].</td>
</tr>
<tr>
<td>An observational cohort study of 5133 adults showed a statistically significant inverse association between vitamin E intake and coronary mortality [36].</td>
<td>An RCT of 9541 adults with or at high-risk of coronary artery disease found no significant difference between the patients randomized to vitamin E or placebo [37].</td>
</tr>
<tr>
<td>An observational retrospective cohort study of 972 postmyocardial infarction patients showed patients with electrical and mechanical dysfunction treated with digoxin had a statistically significant higher risk adjusted mortality rate compared with patients who did not receive digoxin [38].</td>
<td>An RCT of 6800 patients with mild to moderate heart failure, of which 65% had a prior myocardial infarction, found no significant difference in mortality rates between patients randomized to digoxin or placebo [39]. This study did demonstrate that patients treated with digoxin had a statistically significant lower likelihood of being hospitalized with heart failure.</td>
</tr>
</tbody>
</table>
and various methodological features of an RCT allows a clinician to determine how likely the results approximate the truth (Table 5).

**Systematic deviations from the truth**

Systematic deviations from the truth result from biases introduced by various methodological features of an RCT. Importantly research has demonstrated that some RCT methodological features can result in moderate biases that may alter the treatment effect by up to 40% [30]. The impact of such bias is sometimes profound and larger than the moderate treatment effects that investigators seek to detect. The methodological features of an RCT that are important to evaluate include: concealment of randomization, avoidance of ascertainment bias, completeness of patient follow-up, and intention-to-treat analysis.

Concealment of randomization means that the individuals enrolling patients into the RCT are unaware of the upcoming treatment assignments [44]. Without this safeguard, investigators can systematically decide in advance which patients receive the experimental and control interventions. Concealment of randomization may be compromised when nonrandom methods of treatment allocation (e.g. admission date, hospital number, alternate assignments) are utilized.

‘Ascertainment bias occurs when the results or conclusions of a trial are systematically distorted by knowledge of which intervention each participant is receiving’ [45]. Blinding or masking of the participants, healthcare providers, data collectors and outcome assessors is the best means to avoid ascertainment bias by avoiding systematic imbalances in endpoint evaluations and effective concomitant cointerventions [46]. If the outcome assessed is total mortality or a clear indisputable clinical endpoint, then systematic imbalances in the endpoint evaluations are unlikely even when the groups involved in a trial are not blinded. However, even this type of outcome does not eliminate the potential systematic imbalances in effective concomitant cointerventions that can occur when the participants and healthcare providers are not blinded.

RCTs that evaluate management strategies (e.g. medical versus surgical management of coronary artery disease) do not restrict effective cointerventions to a single group. In this type of study, the blinding of individuals who can administer effective cointerventions is not a concern, however, for certain types of outcomes (e.g. symptom relief, transient ischaemic attacks), the effect of not blinding the groups can systematically bias endpoint evaluations. Fortunately, in most RCTs, blinding is possible and even in the circumstances when blinding of patients, healthcare providers and data collectors is impossible (e.g. some surgical trials) it is generally possible to blind the outcome assessors.

Completeness of patient follow-up is important because the patients who were not accounted for at the end of the trial may have experienced the outcome of interest. Consideration of the overall number of events in both treatment groups and the corresponding number of patients lost to follow-up in these groups is necessary to determine if it is possible that a meaningful systematic deviation from the truth has occurred.

An intention-to-treat analysis should always be the primary analysis to be undertaken and considered. This type of analysis means that all patients are analysed in the groups to which they were randomized regardless of what treatment they received. Failure to analyse by intention to treat could likely damage the balance in prognosis between the two groups which is the underlying purpose of randomization [47].

Systematic deviations from the truth can also result from emphasizing treatment effects in different patient subgroups. In general, the true effects that are likely to occur in a subgroup are best represented by the overall treatment effect rather than what is actually observed in the subgroup. Therefore, emphasis on subgroup results, especially

---

**Table 5 Minimizing systematic and nonsystematic errors in randomized controlled trials**

Methods to minimize systematic deviations from the truth (i.e. biases)
- Concealment of randomization
- Avoidance of ascertainment bias
- Complete patient follow-up
- Intention to treat analysis
- Main emphasis on primary results (i.e. avoid emphasizing data-dependent results from subgroup analyses)

Methods to minimize nonsystematic deviations from the truth (i.e. random error)
- Large number of outcomes (most commonly achieved by randomizing a large number of patients)
data-derived emphasis is generally misleading [48]. Undue emphasis on a subgroup analysis may result in many individuals being denied a life-saving drug. For example, a subgroup analysis in the Canadian aspirin trial in patients with a transient ischaemic attack resulted in the Food and Drug Administration (FDA) only approving aspirin therapy in male but not female patients [49]. Unfortunately, many women were denied aspirin therapy after a transient ischaemic attack, despite the fact that the drug decreases the risk of death in both men and women [50].

**Nonsystematic deviations from the truth**

Nonsystematic deviations from the truth result from random error (i.e. the play of chance on the outcome amongst individuals in the treatment and control groups) [51]. Despite the use of a random process in a small trial by chance alone, the two treatment groups may have very different prognoses. Therefore, any differences observed may be due to chance, as opposed to a treatment effect. The only way to overcome the problem of random error is to have large trials with a lot of outcome events.

Many individuals fail to appreciate just how large RCTs have to be to reliably detect clinically meaningful moderate risk reductions. A hypothetical RCT, evaluating the effect of a new investigational drug on patient mortality in the setting of acute myocardial infarction, is used to highlight this point. If the true relative risk reduction (RRR) was a clinically meaningful 20%, and 1000 patients were randomized to receive the new investigational drug and 1000 patients to receive a placebo, 80 deaths could be predictably seen (i.e. 8%) in the experimental group and 100 deaths (i.e. 10%) in the placebo group. Despite what some would consider a large study, the true 20% RRR would not be reliably identified in this study because the result would not reach conventional levels of statistical significance ($P = 0.1$). In this situation where the truth was a 20% RRR in mortality with the new investigational drug many individuals because of the trial size may falsely conclude that the differences observed were only due to chance and that the new investigational drug is not efficacious. This example highlights the importance of undertaking truly large RCTs with a large number of events to reliably detect clinically important moderate treatment effects.

**The role of randomized controlled trials in evidence-based decision making**

Recently, a group of authors have developed a model for evidence-based decision making (Fig. 1) [52, 53]. 'Evidence-based decision making depends upon utilizing clinical expertise to integrate information about a patient’s clinical setting and circumstances with the best research evidence whilst incorporating the patient’s preferences and actions’ [53]. The components of this model are briefly described, and the role of the RCT in evidence-based decision making is indicated.

**Research evidence**

Research evidence is central to evidence-based decisions because without valid research evidence, all components of the model may be misinformed. As discussed in detail above, the RCT generally provides our best means to determine the effect of therapeutic interventions. Therefore, when making therapeutic decisions, ideally it is an RCT or a meta-analysis of RCTs that should inform the decision.

**Clinical state and circumstances**

RCTs provide us with results that reflect the average patient in the treatment groups of the trial. In clinical practice, our patients rarely resemble the average patient from the RCT. Patient characteristics will typically put a patient at higher or lower risk of the
outcomes or side-effects than the average patient in the RCT. Thus, consideration of the patient’s clinical state is important to allow individualization of a patient’s underlying risk from the disease state and from the treatment. Furthermore, a certain type of patient may have been excluded from a trial, but there may be no strong biological rationale as to why the proven treatment should not be effective in this type of patient. In such cases, it is reasonable to extrapolate the results to such patients, as long as there is no undue risk of adverse effects.

Patients’ preferences and actions
Evidence should not be used to tell patients what to do, rather evidence should be used to allow patients to make informed decisions. All treatment decisions involve a weighing of the potential benefits, risks, inconveniences and costs. Patients’ preferences/values may weigh these potential outcomes differently allowing individuals to rationally make different decisions despite being presented with the same evidence. Studies have demonstrated patients and physicians vary in their weighing of the risk of stroke and bleeding when considering anticoagulation therapy for atrial fibrillation [54], and patients and physicians vary in their weighing of the potential benefits, risks, inconveniences and costs of antihypertensive therapy [55]. As such, incorporation of patient preferences into evidence-based decision making is desirable.

Clinical expertise
Clinical expertise is the final overarching component of the evidence-based decision making model. Clinical expertise is required to establish, balance and integrate the patient’s clinical state and circumstances, preferences and actions, and the best research evidence. In this regard, clinical expertise requires both clinical and content area knowledge, skills at critical appraisal and clinical diagnosis – all tempered with good judgement based on incorporating the patient’s preferences and circumstances.

Conclusions
Although the RCT is a relatively new tool in the overall history of medical research, it has been established as our current best means to evaluate the efficacy of an intervention. As such, the RCT plays a central role in providing the evidence basis for evidence-based decision making about therapeutic interventions. However, randomization in itself does not ensure that the conclusion reached is the right one. RCTs need to enroll an adequate number of patients and utilize appropriate methodology (e.g. concealment of randomization) to minimize the possibility of nonsystematic and systematic deviations from the truth. In keeping with this there has been an appropriate shift away from undertaking small inconclusive RCTs assessing surrogate outcomes, towards conducting appropriately designed larger (and sometimes simple) RCTs that evaluate the effects of treatments on major clinical outcomes. These trials have and will continue to make a substantial impact on the health of the patients.

Conflict of interest statement
No conflict of interest was declared.

References
10 Pitt B, Segal R, Martinez FA et al. Randomised trial of losartan versus captopril in patients over 65 with heart failure (eval-

Received 16 March 2003; revision received and accepted 23 March 2003.

Correspondence: Dr P. J. Devereaux, McMaster University, Faculty of Health Sciences, Clinical Epidemiology & Biostatistics, Room 2C12, 1200 Main Street West, Hamilton, ON, L8N 3Z5, Canada (fax: 905-524-3841; e-mail: philipj@mcmaster.ca).