

Trials and fast changing technologies: the case for tracker studies

Richard J Lilford, David A Brauholtz, Roger Greenhalgh, Sarah J L Edwards

New or variant treatments—and we use the word in a wide sense to include procedures and devices as well as drugs—should be subject to randomised controlled trials.¹ Treatments may also develop, changing in ways that are widely considered to be improvements. For example, a new version of a surgically fitted device supersedes the old. This complicates existing comparisons of the device compared with medical treatment. And it leads to another issue—when should researchers start a randomised controlled trial in a clinical area where there is rapid technological change? Start too early and the resultant comparisons may seem likely to turn out to be irrelevant, but start too late and the chance of collecting much good quality data will have been lost, perhaps forever if clinical opinion has “gelled” despite the absence of randomised controlled trial data. The problem is compounded by the considerable time it takes to design, commission, and establish a full scale clinical trial.

These problems are encountered widely, particularly with devices. These may be licensed even before their health effects have been studied in detail and are subject to frequent modifications in design and use. A good example is endovascular aortic aneurysm repair, in which a Dacron tube is positioned within the abdominal aorta and held in place by an expandable stent. In 1991, Parodi et al showed that aneurysms could be repaired in this way.² Several stent graft systems have emerged since then, with changes occurring almost monthly.

In these circumstances useful evaluation by randomised controlled trial evaluation might be thought impossible, and researchers and commissioners might choose to wait for things to stabilise.³ In this paper we argue against waiting and advocate the use of trials which start early on in periods of rapid technological change and which follow and inform developments. We call these studies “tracker trials” because the content of the trial will track changes in treatments or beliefs of clinicians. These studies are distinct from conventional randomised controlled trials which are one off events, following preset and rigid protocols.

Tracker trials

At the outset, a tracker trial will typically initially consist of a set of randomised comparisons of various examples of a new type of technology, each with stand-

Summary points

Evaluating treatments is difficult when developments or variants arise frequently

In these circumstances randomised controlled trials should not await stability, but should track progress over time, providing unbiased comparisons at each stage

These “tracker trials” should be guided by flexible protocols, without prefixed sample size (or duration), and will require sophisticated interim analyses

Following clinical practice flexibly will enable tracker trials to be comprehensive—collecting maximum amounts of randomised data and ensuring standardised outcome measures across centres

Starting trials while technology is changing will ensure maximum use of information after it has stabilised

Tracker trials would also be able to monitor treatments and centres to detect poor performance quickly and to provide an effective early warning system

Department of Public Health and Epidemiology, University of Birmingham, Birmingham B15 2TT

Richard L Lilford
professor of health services research

David A Brauholtz
senior research fellow

Sarah J L Edwards
research associate

Imperial College of Science Technology and Medicine, Charing Cross Hospital, London W6 8RF

Roger M Greenhalgh
professor of surgery

Correspondence to: D A Brauholtz
D.A.Brauholtz@Bham.ac.uk

BMJ 2000;320:43–6

ard treatment. The key observation is that numbers of completely different new treatments do not usually arise independently at the same time. So, where many different treatments are available and arising, most will be more or less closely related to each other. Before any comparative data are available there may be no reason to prefer any particular treatment, but there may already be good reasons to believe in generic “family resemblances.” Thus, if a variety of new surgical treatments all use the same form of access, comparative data from one of these (against a standard treatment, say) would give some information about the expected comparative performance of all treatments with the same form of access. At the same time, some of these treatments may involve fitting a metal device, others a plastic one. Comparative data from a particular metal device would give some information about all

treatments using metal devices. Maximum possible collection of randomised controlled trial data would result through allowing each clinician to randomise between trial arms they feel are reasonable alternatives, and maximum information relating to each treatment and to each “family characteristic” (for example, use of metal) would arise from combining information using family resemblances.

In short, tracker trials allow different treatments to be compared and the effects of particular components of treatments to be evaluated. Since practitioners may be familiar with (and prepared to use) only a few of the available new treatments, comparisons between new treatments will, more often than not, have to be made on an observational basis. Note that the concept of making observational comparisons between different new treatments within a randomised trial is not new. For example, the MRC European trial of amniocentesis versus chorion villus sampling included non-randomised (observational) comparisons between different techniques and devices for carrying out the chorionic sampling.⁴ What is novel, however, is the potential to modify experimental subgroups as the trial proceeds.

Features of tracker trials

Flexibility

Tracker trials must be flexible and include competing treatments as they arise. A tracker trial adapts to clinical practice by including at any point in time treatments that are considered viable alternatives. For this reason, protocols should be revised frequently—new arms may be required as additional treatments or variants emerge. Conversely, arms may also be removed. If the number of viable alternatives settles down to two or three, then those may be factored simultaneously into the randomisation, provided that the skills to use them are not too dissimilar and that this reflects individual clinical opinion. Of course, the treatment that was previously standard may itself have become obsolete. In other words, a trial that starts out comparing treatments based on an altogether new (generic) technology with the standard treatment may gradually evolve over the years into a trial of different treatments all based on the new approach. It follows that an end date for a tracker trial cannot be set in advance.

Inclusivity

Tracker trials should include all operators or centres irrespective of skill or experience. When new treatments are technically demanding, the operator’s learning curve matters. This presents particular difficulties and complexities where technology is changing. Trials typically try to avoid this problem by recruiting only experienced operators. With a new treatment, most operators are in some sense inexperienced, making it difficult to restrict recruitment in this way. More generally, since learning curves are an integral part of a treatment they should not be ignored.⁵ Thus, a surgical technique that is superior to medical treatment in the hands of an experienced surgeon, but markedly inferior in those of a novice, will seem better in a typical trial restricted to experienced surgeons. But how then are surgeons to acquire the necessary

experience? Tracker trials should collect and analyse data on operator experience because of the implications for service delivery. The methodological aspects are the topic of a current review funded by the NHS methodology research programme.⁶

Complex analysis

Tracker trials will require more complex analysis and more sophisticated use of findings, which will not be clear cut, at least in the early stages. Investigating the effects on outcomes of characteristics of patients or diseases, experience of operators, and treatments and components of treatments used is clearly more complex than in conventional trials.⁷

Sophisticated commissioning

Tracker trials require more sophisticated methods of commissioning and management. Research commissioners need flexible budgets (at least in terms of the duration of the study), and organisations hosting research also need to be able to respond flexibly. Since the trial protocol will evolve in practice and since the duration of the trial cannot be fixed in advance, the trial steering committee will be more intimately involved in vetting the trial protocol than is normally the case. This need to make crucial funding decisions during the course of a study calls for a more flexible approach to research. This has been referred to as the iterative commissioning process.⁸

Advantages of tracker trials

Tracker trials combine the advantages of registers of new technologies (which involve detecting adverse incidents and comparisons across different devices) with those of randomised controlled trials (which yield unbiased data). Early randomisation is the key to many benefits.

Take advantage of equipoise while it exists

Early randomisation may emerge as the only randomised controlled trial option. If and when technologies stabilise, it may be too late to randomise: clinicians may have developed firm if unsubstantiated views, such that they are no longer equipoised.⁹ The longer the wait, the larger the number of prematurely optimistic clinicians, because those already performing a procedure tend to have a rosier view than those basing judgments solely on the published reports.¹⁰ Those who adopt new technologies early may then influence others who do not want to be left behind. Thus, laparoscopic cholecystectomy and coronary artery stenting in patients with mild to moderate angina came into general use before trials showed no benefit over mini-laparotomy and medical treatment respectively.^{11 12} Some surgeons consider, on observational data alone, that the time for a randomised controlled trial of endovascular “coiling” of intracranial aneurysms has passed. As Martin Buxton, professor of health economics at Brunel University, has remarked, “It’s always too early to start a trial, until it is too late.”

Maximise data collection

Only a trial in place before the technologies stabilise can collect data in the early period of stability (usually recognised only in retrospect), given the lead time for

launching a trial. Tracker trials thus maximise collection of randomised data comparing available treatments.

Contribute to development of technology

The early, good quality comparative data that a tracker trial will provide, albeit in small quantities, can help determine which variants of a new technology are further developed and which are not. Even non-randomised comparisons between different treatments are likely to be less biased if each one is separately randomly controlled using a standard treatment as a benchmark.¹³

Monitoring the progress of tracker trials

When a new technology is introduced in the health service, sensitive, short term performance monitoring of new devices and of centres is essential. Conventional trials preclude the routine auditing of outcomes and provide only very delayed feedback. Conventional monitoring by a trial data monitoring and ethics committee may be infrequent, not compare centres, and produce action only on strong evidence of poor performance—at least on the main outcome measure. This is perhaps why the issue of whether a randomised controlled trial of endovascular aortic aneurism repair should start in the midst of so much technological development originally split the clinical community of surgeons and radiologists. Routine outcome monitoring is one change in UK surgery that resulted from the Bristol case.¹⁴ It allows early detection of technologies or centres with very bad outcomes. The data monitoring and ethics committee for a tracker trial would therefore have three responsibilities:

- To ensure that treatments which are clearly superior are quickly adopted, by publishing the results (they would also usually stop the trial)
- To detect at an early stage if particular devices or centres are performing poorly
- To ensure information gathered in the trial is used to guide development of better treatments.

Sample sizes

Statistical modelling has confirmed the intuitively appealing notion that the more rapidly new treatments are arising, the earlier should be the point at which unpromising treatments are rejected.¹⁵⁻¹⁷ In such situations, the use of conventional sample size calculations (with conventional significance levels) seems particularly inappropriate. A more rational approach would take into account additional factors such as the frequency with which new contenders are likely to emerge and would produce correspondingly smaller sample sizes. Unfortunately, such quantification is extremely difficult. Tracker trials must therefore involve regular and flexible assessment of all relevant data (internal and external) without prefixed sample sizes.

Organised approach

Membership of a tracker trial data monitoring and ethics committee, with the dual responsibilities of auditing and evaluating new treatments, would involve frequent meetings and difficult decision making. There would be a potential conflict between rejecting unpromising treatments quickly to benefit patients

generally (moral interest) and avoiding premature abandonment of expensively developed treatments (commercial interest). However, this approach—based on all available data, properly analysed and appraised by specially constituted committees, with members who have the confidence of all sectors involved and the authority to take controversial decisions—seems preferable to allowing technologies to diffuse passively and develop in an ad hoc and possibly idiosyncratic way.

Feedback trials

There is another possibility which will encourage greater openness and avoid forcing data monitoring and evaluation committees to make dichotomous decisions in the face of evidence that may be inconclusive. Instead of sequestering trial analyses, the monitoring committee could routinely and frequently feed them back to clinicians and patients, making them available publicly.¹⁸⁻¹⁹ The effect of the data on specimen prior beliefs could be presented within a (bayesian) decision analytic framework.²⁰⁻²¹ Statistical aspects of bayesian monitoring and analysis of trials are much discussed.²²⁻²⁹ A feedback trial seems more flexible and democratic than forcing clinicians and patients to base decisions only on their prior beliefs, personal experience, and data acquired outside the trial, while keeping trial data for the data monitoring and ethics committee alone.³⁰ It also spreads the burden of decision making by using the collective knowledge of providers of care and allows that information to be combined with patients' values, thus avoiding a stark and possibly erroneous verdict by the monitoring committee. Feedback is currently being used in a trial of early versus delayed delivery for preterm, growth retarded fetuses.³¹ This trial features regular feedback of interim results to participating clinicians, and no adverse recruitment effects have been observed. On the other hand, a matched case-control study with a frequentist statistical perspective found reduced recruitment in open trials.³²

Essence of tracker trials

The essence of a tracker trial is to provide, in the context of increasing numbers of treatments, a combination of methods that will:

- Detect quickly treatments that are performing poorly or are potentially dangerous (and thereby provide an early warning system)
- Reject unpromising new treatments, and otherwise inform the use of available treatments and the development of improved treatments
- Eventually (when stability ensues) provide maximum information as to which treatments are best.

We feel that bayesian or feedback approaches are particularly suitable for the first two tasks. If desired, a hybrid solution could be used, so that once stability had arrived comparative data could be sequestered in the usual way and subject to conventional data monitoring within a hypothesis testing (conventional) paradigm.

At heart, our message is that the methodological tools for tracker trials exist, and that researchers and research commissioners should be more imaginative in making use of the full repertoire available to them. A

Hypothetical example of tracker trial—treating aortic aneurisms

Past

In 1996, 139 endovascular aortic aneurism repair devices, manufactured by a range of commercial and non-commercial organisations, were implanted in many patients. The devices were used as an alternative to the established open repair method and also in patients unfit for open repair. Data are needed on how the available technologies compare in both areas, especially in the medium to long term.

Hypothetical future

Year 1999

- All the major players (NHS research and development, royal colleges, trusts, manufacturers) agree to support a tracker trial in this area
- Under the aegis of the NHS Health Technology Assessment Programme, a major London university is contracted to coordinate the tracker trial
- A steering committee and standing protocol committee is established, which rapidly sets up communication links with clinicians around the country
- The “sets” of treatments that are currently viewed as alternatives by practitioners are established. At this stage, these all comprise a comparison of various treatments with the “standard”
- Comparisons between treatments for endovascular aortic aneurism repair will necessarily be non-randomised (but less biased than simple observational comparisons)
- Appropriate end points and risk factors are identified and agreed, and forms are designed
- A data monitoring and evaluation committee is constituted; it has substantial support from statisticians and is closely in touch with clinicians
- Contacts with and regular searches for other (for example, overseas) research in the area are instituted

Year 2000

- Analyses and protocol revisions are undertaken quarterly, publishing the results

Year 2001

- As endovascular aortic aneurism repair devices become more widely used, monitoring establishes that some “learners” have poor results. Royal colleges institute improved training and supervision
- Three older devices with relatively poor results fall into disuse and the corresponding observational comparisons are removed from the protocol
- A new drug X is launched, which is thought to help repair aneurysms. The protocol committee introduces it as a “factor” in all the existing trial arms (that is, using a factorial design)

Year 2002

- Analyses (prompted by necropsy findings) of accumulated data from several repair devices that use one particular material find a poor medium term outcome. This result causes replacement of the material in all future fittings of devices and recall and checking for patients treated with this material

Year 2003

- Two leading devices emerge. They have equivalent long term results to open surgery but with much less morbidity and similar total costs
- Equipose between open and closed repair is lost but comparisons of the two leading devices continues on a randomised basis

hypothetical example of how a tracker trial might proceed is shown in the box.

We thank Professor Adrian Grant for comments and suggestions, in particular for suggesting that early randomisation may help regularise the currently, sometimes vague, ethics of uncontrolled experimentation with new technologies.

Funding: RJL, DAB, and SJLE acknowledge support from the NHS Executive. Views and opinions are our own and do not necessarily reflect those of the NHS Executive.

Competing interests: None declared.

- 1 Banta HD, Sanes JR. Assessing the social impacts of medical technologies. *J Community Health* 1978;3:245-58.
- 2 Parodi JC, Palmaz JC, Barone HD. Transfemoral intraluminal graft implantation for abdominal aortic aneurysms. *Ann Vasc Surg* 1991;5:491-9.
- 3 Feinstein AR. An additional basic science for clinical medicine. II. The limitations of randomised trials. *Ann Int Med* 1983;99:544-50.
- 4 Russell I. Evaluating new surgical procedures. *BMJ* 1995;311:1243-4.
- 5 MRC Working Party on the Evaluation of Chorion Villus Sampling. Medical Research Council European trial of chorion villus sampling. *Lancet* 1991;337:1491-9.
- 6 Grant A. *Assessing the learning curves of health technologies*. Available at NHS R&D Health Technology Assessment Programme website www.hta.nhsweb.nhs.uk/projects/962502.htm. (accessed 8 September 1999).
- 7 Spodick DH. The randomized controlled clinical trial. Scientific and ethical bases. *Am J Med* 1982;73:420-5.
- 8 Lilford R, Jecock R, Shaw H, Chard J, Morrison B. Commissioning health services research: an iterative method. *J Health Services Res Policy* 1999;4:164-7.
- 9 Chalmers TC. When should randomisation begin? *Lancet* 1968;i:858.
- 10 Lazaro P, Fitch K, Martin Y, Bernstein S. Physician recommendations for coronary revascularisation: variations by clinical specialty. In: *Abstracts from the 14th annual meeting of the International Society for Technology Assessment in Health Care*. Ottawa: International Society for Technology Assessment in Health Care, 1998.
- 11 Majeed AW, Troy G, Nicholl JP, Smythe A, Reed MW, Stoddard CJ, et al. Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy. *Lancet* 1996;347:989-94.
- 12 NHS Centre for Reviews and Dissemination. *Management of stable angina*. 1997. York: NHS Centre for Reviews and Dissemination, 1997.
- 13 Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50:683-91.
- 14 Warden J. NHS hospital doctors face compulsory audit. *BMJ* 1998;316:1851.
- 15 Strauss N, Simon R. Investigating a sequence of randomized phase II trials to discover promising treatments. *Stat Med* 1995;14:1479-89.
- 16 Brunier HC, Whitehead J. Sample sizes for phase II clinical trials derived from bayesian decision theory. *Stat Med* 1994;13:2493-502.
- 17 Yao TJ, Begg CB, Livingston PO. Optimal sample size for a series of pilot trials of new agents. *Biometrics* 1996;52:992-1001.
- 18 Prescott RJ. Feedback of data to participants during clinical trials. In: Tagnon HJ, Staquet MJ, eds. *Controversies in cancer: design of trials and treatment*. New York: Mason Publishing UK, 1979:55-61.
- 19 Edwards SJL, Lilford RJ, Braunholtz DA, Thornton J, Jackson J, Hewison J. *Ethical issues in the design and conduct of randomised clinical trials*. NHS R&D Health Technology Assessment Programme website, www.hta.nhsweb.nhs.uk (accessed 9 September 1999). (NHS HTA monogram 1998(2)15.)
- 20 Lilford RJ, Braunholtz D. The statistical basis of public policy: a paradigm shift is overdue. *BMJ* 1996;313:603-7.
- 21 Lilford RJ, Thornton JD. Decision logic in medical practice. The Milroy lecture 1992. *J R Coll Physicians* 1992;26:400-12.
- 22 Berry DA. Decision analysis and bayesian methods in clinical trials. *Cancer Treat Res* 1995;75:125-54.
- 23 Varlan E, Le Paillier R. Decision analysis: theory and methods in clinical development and monitoring of clinical trials. *Therapie* 1996;51:348-55.
- 24 Carlin BP, Sargent DJ. Robust bayesian approaches for clinical trial monitoring [correction appears in *Stat Med* 1997;16:1300]. *Stat Med* 1996;15:1093-106.
- 25 Freedman LS, Spiegelhalter DJ, Parmar MK. The what, why and how of bayesian clinical trials monitoring. *Stat Med* 1994;13:1371-83.
- 26 George SL, Li C, Berry DA, Green MR. Stopping a clinical trial early: frequentist and bayesian approaches applied to a CALGB trial in non-small-cell lung cancer. *Stat Med* 1994;13:1313-27.
- 27 Gray RJ. A bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics* 1994;50:244-53.
- 28 O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990;46:33-48.
- 29 Parmar MK, Spiegelhalter DJ, Freedman LS. The CHART trials: bayesian design and monitoring in practice. CHART Steering Committee. *Stat Med* 1994;13:1297-312.
- 30 Thornton JG, Lilford RJ. Preterm breech babies and randomised trials of rare conditions. *Br J Obstet Gynaecol* 1996;103:611-3.
- 31 Lilford R. Formal measurement of clinical uncertainty: prelude to a trial in perinatal medicine. The Fetal Compromise Group. *BMJ* 1994;308:111-2.
- 32 Green SJ, Fleming TR, O'Fallon JR. Policies for study monitoring and interim reporting of results. *J Clin Oncol* 1987;5:1477-84.

(Accepted 29 July 1999)