Randomised trials and big data analysis: we need the best of both worlds

Tom Treasure[1†]

Johanna JM Takkenberg[2]

[1]Clinical Operational Research Unit,
University College London
London
UK

[2] Department of Cardio-Thoracic Surgery,
Erasmus University Medical Centre,
Rotterdam,
The Netherlands

[†]Corresponding author
Tom Treasure
tom.treasure@gmail.com
Phone +44(0)7957168754

Text word 2,976

Randomised trials; registry studies; meta-analysis

Randomised trials or sophisticated analyses of 'Big Data'

Seventy years ago establishing the worth of an operation was more straightforward. There was little of any use to be done for structural heart disease. Cyanotic heart disease was particularly lethal - most 'blue babies' died. Some struggled through childhood, burdened by symptoms, only to die young.[1] This bleak outlook was transformed in the 1940s, first by an extracardiac systemic to pulmonary artery shunt devised by Alfred Blalock[2] and then by a direct intracardiac operation on the right ventricular outflow tract devised by Russell Brock.[3] When the mechanistic effect of surgery was clearly seen and could be consistently achieved, and the clinical course of the patient was observed to be substantially improved, successive operations entered practice.

With refinement of surgical techniques, as the new and the old operations were compared by simple observation of outcomes, it was still easy to see whether the new treatment offered better survival and/or relief of symptoms and a better quality of life. We say 'easy to see' with some reservation because it should not be overlooked that many ineffective treatments also became accepted and continued in practice for generations. It was only after thousands of years that bloodletting was abandoned in the treatment of fever and sepsis. It took 90 years for surgeons to turn their backs on radical mastectomy for breast cancer in favour of less mutilating operations which were proven in randomised trial to be no less effective in controlling the primary cancer, and to be greatly superior in terms of complications.[4] There are numerous reversals brought about by controlled trials.[5]

With further progress it has become increasingly difficult to discern 'signal from noise'.[6] In current practice adjunctive systemic and interventional treatments are often applied synchronously or sequentially. Comorbidity in elderly patients fogs the issue further. To which if any of the several components of combined treatments should benefit be ascribed? There has been a tremendous diversification of both treatment options and patient populations, and more marginal differences are being tested, making a straightforward comparison of observational findings incapable of determining the better treatment. At the same time this diversification hampers the generalizability of RCTs. Outcomes are no longer black or white but a full spectrum of colours. And we have arrived, not before time, in an era of patient centred care. Evidence-based, patient-specific, and often value-sensitive decisions have to be made for wide diversity of patients.

The RCT came to be regarded as the gold standard in finding evidence for a treatment in clinical practice. The essential feature is that the treatment is randomly assigned so that all known and unknown factors that might influence the outcome of the treatment under test are similarly present in both groups. Any difference in outcome can then be attributed to the relative effectiveness of the treatments in achieving the pre-specified desirable outcome. A downside of the RCT is that data are acquired specifically to answer one research question and as further questions arise, new data must be acquired starting all over again.

Large databases and registers are now available. In theory, with the use of sophisticated statistical analyses, the difference in outcome attributable to the treatment may be discernible by statistical adjustment for other factors influencing outcome, or by matching patients to exclude effects other than those due to the treatment. The data set continues to accrue patients and can be used repeatedly to answer other questions. As larger and better organised observational data sets are collected, and new meta-analytic techniques are developed, is the RCT's place unassailable? If, as seems reasonable, RCTs and more complex analytical methods are to co-exist, what are their relative merits? To explore the issues we will use the

example of bilateral versus left only internal mammary artery (BIMA versus LIMA) grafts in coronary artery surgery. This was the subject of a much lauded debate at EACTS 2017.[7] We will then consider the choice of surgery versus ablative radiotherapy for lung cancer.

The double or single mammary artery debate

The 31st Annual Meeting of EACTS in October 2017 hosted high level discussions about the evidence that might guide practice. Prominently placed was a session on whether the goal of coronary artery bypass surgery should be BIMA grafts for all coronary operations, or should the standard of care be an operation including a LIMA. Professor Nick Freemantle was quoted in EACTS Daily News applauding our association saying '*It is to the credit of EACTS that they are having a debate on this topic at the meeting'*. The superiority of a LIMA for the left anterior descending was established on the basis of the observation that a left internal mammary artery as a pedicled graft to the left anterior descending coronary artery had better patency rates at 10 years than aorto-coronary saphenous vein grafts, and with commensurately better clinical outcomes.[8] The unresolved question remains: does the addition of a right internal mammary artery graft (RIMA) provide a further useful incremental gain in long term clinical outcomes?

Two major studies provide us with material on which to make comparisons of the relative merits of the two methods of seeking evidence: an RCT versus a sophisticated analysis of observational data.

1. ART, a randomised controlled trial in 3102 patients was published in 2016 in the New England Journal of Medicine.[9]
2. Meta-analysis of 29 observational studies including 89,399 patients including 12 propensity matched studies in 20,525 patients was published in 2017 in Heart.[10]

ART: the Arterial Revascularisation Trial

The means of assigning patients

When the ART trial was mooted in about 2004 it was proposed that expertise based randomisation would be used. This was dubbed expertise based randomisation where two operations were to be compared and the dyad of surgeon and her preferred operation was not disrupted. Although this was proposed as a means of helping surgeons to engage in RCTs[11] in fact the principle is inherent in trials of surgeon versus cardiologist delivered therapies. It also applies in surgical resection versus radiotherapy, or other ablative techniques for cancer.

In the ART trial this would have meant that each patient was randomly assigned either to a surgeon who favours BIMA or to a surgeon who prefers one arterial graft, a LIMA to the LAD. This would makes sense if one can assume that the surgeons are of comparable competence and it is a device which is under evaluation, such as a choice between heart valves. For LIMA/BIMA comparison the problem is obvious. Surgeons who prefer to use two mammary arteries may be the more deft and speedy operators, working with teams more practiced at mammary artery dissection and undertaking the additional surgery more expeditiously. Any result from such a trial would have been confounded by differing expertise. Expertise based randomisation was therefore counselled against. Surgeons had to be competent to do either and then be willing to allow their preference to be overridden by randomisation. Even if they had personal preferences the existence of the two approaches indicated 'group equipoise'. Once randomly assigned, the allocation to LIMA or BIMA must be adhered to; subsequent modification of the operative plan by the surgeon would undermine the trial design. Once the protocol is agreed those carrying out the trial must put

individualised judgement aside; this highlights a problem in encouraging surgeons to accept random allocation of their patients.

Table. Patient characteristics and outcomes for LIMA and BIMA

The outcomes in ART

In the RCT at five years there was no difference in the primary outcome of interest which is survival; ten year survival rates are not yet available. There was no difference in hospital mortality, bleeding, myocardial infarction or stroke. There were more sternal wound complications with BIMA attributable to the added risk of bilateral interference with sternal blood supply.

The meta-analysisof LIMA vs BIMA

Acquiring the data

Systematic reviews are now greatly facilitated by electronic searching and retrieval of a large number of sources. In this case 3678 articles were identified. Adhering to pre-specified inclusion and exclusion criteria, these were narrowed down to 120 potentially relevant articles. Finally 29 studies were pooled for analysis. The large majority (27/29) were retrospective observational studies, and in 12 studies there was propensity matching.

The outcomes in the meta-analysis

Five-year survival was higher with BIMA than LIMA other than in a diabetic subset and was seen throughout the 25 years of follow up in the pooled analysis with the difference widening at 10 and 20 years.(Fig.1) The authors calculated an overall hazard ratio of 0.78 which translates to a pooled cumulative 5 and 10 year mortality of 7.7% and 17.9% respectively for BIMA and 13% and 29.5% respectively for LIMA. The need for subsequent revascularisation after BIMA was half that after LIMA. Stroke, sternal wound infection and revascularisation were all significantly higher with LIMA than BIMA.

Why the difference in conclusions?

For early and late mortality and for important in hospital events there appeared to be a clear answer in favour of BIMA in the meta-analysis, differences not seen in the RCT. Because the RCT was based on random assignment, current received opinion (which we share) is that the RCT provides the more trustworthy answer, with the caveat that the conclusion may only be applicable under the circumstances of the study. As an exercise in weighing the comparative worth of an RCT versus a big data matching study, let us consider how the differences may have come about and the implications for selecting and interpreting the two contrasting research methods.

RCTs struggle to accrue sufficient patients

The collected observational data provided a pool of patients thirty times larger than the RCT. Big data are very attractive and suggest more reliability and generalisability. The important point here however is that the RCT was big enough for us to be confident that we have not missed any possible important difference but it does illustrate the attraction of accessing big data sets.

Sex, age and the diabetic incidence of included patients

Registries are ‘real world’ populations but RCTs are a selected sample so there are inherent limitations in the interpretation and application of evidence from RCTs. The inclusions and exclusions are in the trial protocol to satisfy all the considerations of ethics and equipoise, but

the resulting populations and the ways in which they are treated may have departed from the typical clinical scenario under evaluation. From the Table we can deduce that that, due to the constraints of selection and equipoise, women patients may have been underrepresented in RCT compared to observational 'Big Data'.

In ART 25% of patients were over 70 years of age. In fact the average ages were very similar (64 in ART and 63 in the meta-analysis) but the point still merits consideration. Older patients are more vulnerable to perioperative hazards such as stroke, infarction and death while, as a group, the elderly may gain less benefit from a difference in graft patency beyond 10 years. Factors other than the second IMA graft will exert more weight in determining survival so BIMA vs LIMA advantages, even if confirmed in an RCT, may matter less to 'real world' patients.

Diabetic patients were less likely to be randomised in the trial and are more frequent among non-randomly assigned LIMA patients in 'real world' practice.

Post randomisation differences in treatments

The intended purity of the comparison may be eroded by well-intended adjustments in treatments to redress the perceived imbalance in benefit between LIMA and BIMA. It has been pointed out that about 22% of patients assigned to LIMA the surgeon used a radial artery graft to the right coronary. As explained at EACTS by Mario Gaudino, a radial artery graft has superior characteristics to a vein graft and maybe as good as a RIMA thus reducing any separation in survival attributable to the second mammary artery graft.[12;13] However there was a comparable (20%) radial artery use in the BIMA group, moderating what was otherwise a cogent argument.

Analysis of existing data can answer a question more quickly than an RCT

RCTs take a very long time from conception to publication. It is more than 12 years since the ART trial protocol was agreed and things have changed meanwhile. This makes RCTs irksomely inflexible to the individual surgeon wanting to exercise constantly updated clinical judgement. It also means the clinical research question may have moved on.

Complication rates

Recognised complications should be just as reliably recorded for trial and non-trial patients but recording bias might be *less* in registry data where no particular hypothesis is under test. Significantly higher rates of sternal wound problems were seen in both studies with a comparable magnitude of difference, a finding that has face validity – that is to say that it makes it 'makes sense' to the clinically well informed. However, in the non-RCT data analyses, BIMA was associated with a significantly lower in hospital mortality and stroke rate. These early differences are not likely to be attributable to the addition of a RIMA graft to the heart. Therefore (in our opionion) they lack face validity. It suggests to us that better risk 'real world' patients are given elective BIMA operations, and perhaps marginally more skilful surgeons are doing more BIMA operations. The very reason for counselling against expertise based randomisation in ART was probably evident in the meta-analysis.

Mid and long-term survival

The failure of BIMA to show the anticipated benefit in survival in the ART trial may be in explicable in a philosophical way. Both arms will also have had the opportunity of the best medical advice including optimising their 'life style' with respect to smoking, diet, weight and exercise. Antiplatelet medication, cholesterol lowering and other pharmacological

secondary prevention incrementally reduce the risk of coronary events, the need for further interventions, and death. All patients would have received revascularisation to all affected territories, delivered by trial quality teams. Any theoretical benefit to be gained by the marginal effect of the second mammary artery graft may be just too small to show against the marginal disadvantages of more complex surgery. That is not to say that the better biological characteristics of an arterial artery are negated. Individual patients may have benefited from longer lasting myocardial perfusion, but as a policy the clinical advantage is too small to show by five years.

A big question in the treatment of primary lung cancer

We looked for a similar example in general thoracic surgery and the treatment of lung cancer seemed an obvious candidate. As radiotherapy has become more efficacious with sophisticated stereotactic techniques and on the other hand older and frailer patients are more harmed by surgery, is it time for the less invasive radiotherapy treatment to begin replace the more tried and tested surgical method?

What have RCTs told us?

In contrast to coronary artery surgery, there have been vanishingly few randomised trial of lung cancer surgery and none of any size.[14] Comparing surgery and radiotherapy, two incomplete and very undersized trials were pooled. The analysis suggested that radiotherapy might not be inferior to surgery.[15]

Analysis of observational data

To answer this question on available observational data, use has been made of the very large SEER data base (Surveillance, Epidemiology and End Results). It appears to show a clear advantage for surgery.[16] All the flaws suggested above for the LIMA-BIMA comparison are of course present. Very few patients suitable to have either treatment would have had radiotherapy for primary lung cancer within current practice guidelines. The largest biasing factor is that in current practice any patient suitable and fit for surgery is offered surgery as the ‘gold standard’; the frail, elderly and marginal patients are more likely to have radiotherapy.

Is a fair test by random assignment feasible?

To do an RCT would require surgeons and radiation oncologists, respecting each other’s position and their own inherent beliefs, to seek neutral informed expert help in devising a robust trial. Patients deemed suitable for either treatment would have to be introduced to uncertainty by trained trial staff who present the pros and cons to the patient from a clearly stated standpoint of not knowing which treatment is better under these circumstances. All questions such as ‘what would you have if it were you?’ must be deflected. It is difficult for a clinician, to whom a patient has come because of their expertise in this disease, to baldly reply ‘I don’t know’. It would work better if it is only *after* random assignment does the patient go to the assigned practitioner according to expertise based randomisation. The surgeon is then free to boost the individual patient’s trust and confidence in a clinical consultation and need not appear to dissemble by saying she doesn’t know which is the better treatment.[17]

Where are we now?

The traditional pyramid of evidence may no longer be sustainable in the current era due to the diversification and the increased complexity of clinical decision making. Perhaps it is time to move towards a more integrated approach to advancing knowledge where clinical trials are

embedded in large registries and networks of large datasets, and outcomes are no longer only death and complications but more focused on well-being. Other study design options are cluster-randomized trials, adaptive trials, and trials that are embedded within clinical care data or administrative platforms. For an outstanding analysis of the difficulty we face in obtaining evidence for practice we recommend a Nature review.[18] Cardiac surgery has followed cardiology in performing trials. Because the effects are more obviously mechanistic in the early days, observational studies were deemed sufficient in cardiac surgery, but in the important question of choosing the best combination of vascular conduits, as we have seen, big well done studies resulted in different answers. Getting reliable data is not easy by either route so combining all available methods is the best way to get to trustworthy guidelines for practice. Thoracic oncology has proved to be a much harder field to evaluate in the modern era. Thoracic surgery was established as specialty and the repertoire of anatomical lung resection techniques were already well rehearsed at a time when heart surgery was ruled out of bounds.[1] The place of surgical resection should rightly be evaluated alongside other ablative techniques, established systemic therapies, and if we are fortunate, as yet undreamt of methods of treatments. It may be a fruitful testing ground for the new imaginative methods of seeking evidence for practice.

297 Table of conclusions BIMA & LIMA

	ART RCT[9]			Meta-analysis [10]		
	LIMA	BIMA	P	LIMA	BIMA	Test statistics
Number	1554	1548		66958	19644	
Female Sex	14%	15%		26%	15%	
Age years	64±9	64±8				
Diabetes	23%	24%		39%	25%	
Hospital mortality	1.2%	1.2%		2.1%	1.2%	P=0.04
Major bleeding	2.6%	3.1%	HR 1.18; 0.44	3.2%	2.9%	P=0.51
Myocardial infarction	3.5%	3.4%	HR 0.97; p=0.86			
Stroke	3.2%	2.5%	HR 0.78; p=0.24	2.9%	1.3%	P=0.0003
Sternal wound complication/ infection	1.9%	3.5%	HR 1.87; p=0.005	1.4%	1.8%	P=0.0008
Revascularisation	6.6%	6.5%	HR 0.98; p=0.91	10%	4.8%	P=0.005
5 year mortality	8.4%	8.7%	HR 1.04; p=0.77	13%	7.7%	HR* 0.78; P< 0.00001
Composite of death, MI and stroke at five years	12.7%	12.2%	HR0.96; p=0.69			

298

299 *HR of 0.78 is not specifically for 5 years but an overall hazard for death throughout the study

Reference List

1 Treasure T: The Heart Club. ed 1st, London/New York, Clink Street, 2017.

2 Treasure T: Turning blue babies pink: Alfred Blalock's shunt for Fallot's Tetralogy. J R Soc Med 2017;110:376-379.

3 Treasure T: Early days in congenital heart surgery in the UK: the Peacock Club. Heart 2017.

4 Mukherjee S: A Radical Idea; The Emperor of All Maladies. New York, Scribner, 2010, pp 60-72.

5 Prasad V, Cifu A: Ending Medical Reversal: Improving Outcomes, Saving Lives. Baltimore, Johns Hopkins University Press, 2015.

6 Glasziou P, Chalmers I, Rawlins M, McCulloch P: When are randomised trials unnecessary? Picking signal from noise. BMJ 2007;334:349-351.

7 Freemantle N: ART: a statistician's perspective. EACTS Daily News 2017;Sunday 8 October:30-31.

8 Loop FD, Lytle BW, Cosgrove DM, Stewart RW, Goormastic M, Williams GW, Golding LA, Gill CC, Taylor PC, Sheldon WC, .: Influence of the internal-mammary-artery graft on 10-year survival and other cardiac events. N Engl J Med 1986;314:1-6.

9 Taggart DP, Altman DG, Gray AM, Lees B, Gerry S, Benedetto U, Flather M: Randomized Trial of Bilateral versus Single Internal-Thoracic-Artery Grafts. N Engl J Med 2016;375:2540-2549.

10 Buttar SN, Yan TD, Taggart DP, Tian DH: Long-term and short-term outcomes of using bilateral internal mammary artery grafting versus left internal mammary artery grafting: a meta-analysis. Heart 2017;103:1419-1426.

11 Devereaux PJ, Bhandari M, Clarke M, Montori VM, Cook DJ, Yusuf S, Sackett DL, Cina CS, Walter SD, Haynes B, Schunemann HJ, Norman GR, Guyatt GH: Need for expertise based randomised controlled trials. BMJ 2005;330:88.

12 Gaudino M, Tranbaugh R, Fremes S: Bilateral versus Single Internal-Thoracic-Artery Grafts. N Engl J Med 2017;376:e37.

13 Gaudino M: Not so state of the ART: interim analysis "wake-up call" to surgical community. EACTS Daily News 2017;Sunday 8 October:31-32.

14 Treasure T, Russell C, Morton D, Macbeth F, Utley M: Surgical resection of lung cancer in England: more operations but no trials to test their effectiveness. Thorax 2012;67:759-761.

15 Chang J, Senan S, Smit ERJ: Surgery versus SABR for resectable non-small cell lung cancer. Lancet Oncology 2015;16:e374-e375.

16 Paul S, Lee PC, Mao J, Isaacs AJ, Sedrakyan A: Long term survival with stereotactic ablative radiotherapy (SABR) versus thoracoscopic sublobar lung resection in elderly people: national population based study with propensity matched comparative analysis. BMJ 2016;354:i3570.

17 Treasure T, Baum M: An approach to randomization into surgical clinical trials. Br J Surg 2017;104:11-12.

18 Lauer MS, Gordon D, Wei G, Pearson G: Efficient design of clinical trials and epidemiological research: is it possible? Nat Rev Cardiol 2017;14:493-501.