



## OPEN ACCESS

# IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages

Patrick L Ergina,<sup>1,2</sup> Jeffrey S Barkun,<sup>3</sup> Peter McCulloch,<sup>4</sup> Jonathan A Cook,<sup>5</sup> Douglas G Altman,<sup>6</sup> On behalf of the IDEAL group

<sup>1</sup>Cardiothoracic Surgery Division, McGill University Health Centre, Royal Victoria Hospital, Montreal, Quebec, Canada H3A 1A1

<sup>2</sup>Oxford International Programme in Evidence-Based Health Care, University of Oxford, Oxford, UK

<sup>3</sup>Department of Surgery, McGill University, Montreal, Canada

<sup>4</sup>Nuffield Department of Surgical Science, University of Oxford, UK

<sup>5</sup>Health Services Research Unit, University of Aberdeen, Aberdeen, UK

<sup>6</sup>Centre for Statistics in Medicine, University of Oxford, UK

Correspondence to: P L Ergina  
patrick.ergina@muhc.mcgill.ca  
Accepted: 15 March 2013

Cite this as: *BMJ* 2013;346:f3011  
doi: 10.1136/bmj.f3011

The IDEAL framework describes the stages of evaluation for surgical innovations. This paper considers the role of observational studies in the exploration and assessment stages. At the exploration stage, the surgical intervention is usually more widely used, and observational studies should collect prospective data from multiple surgeons, deal with factors such as case mix and learning, and prepare for a definitive evaluation at the next stage of assessment. Although a randomised controlled trial is preferable, a high quality observational study would be acceptable if a randomised trial is not feasible or, on rare occasions, deemed unnecessary.

## Introduction

The evaluation of new innovations, from idea developed to accepted practice, has been less orderly in surgery and other interventional therapies than in clinical pharmacology. The IDEAL framework for surgical innovations and recommendations has been designed to describe the stages of evaluation for these interventional therapies (idea, development, exploration, assessment, and long term study), and to highlight the study designs and reporting standards that are likely to prove most useful at each stage.<sup>1,2</sup> The first two IDEAL stages are covered in the first paper in this series.<sup>3</sup> This second article focuses on the IDEAL recommendations for the use of observational studies in the exploration and assessment stages, and discusses the options for observational study designs and reporting protocols (box 1), using examples of surgical innovations. The final paper in the series covers the undertaking of a definitive randomised controlled trial, mainly in the assessment stage, as well as the long term stage.<sup>4</sup>

## Reaching the exploration stage (IDEAL stage 2b)

By the exploration stage, the innovation is usually already practiced by many surgeons on an increasing number of less carefully selected patients. Promising evidence of safety and beneficial short term outcomes—without unacceptable complications—will have been generated, or further development would have been halted. Under the IDEAL framework, use of retrospective studies should be limited to hypothesis generation in the earliest stages. Typically, the early evaluation in the development stage (2a) will use small observational studies without contemporaneous comparison groups in highly selected cohorts of patients. The exploration stage (2b) offers the opportunity to obtain higher quality evidence in a more representative patient population and to deal with factors that could hinder the conduct of a proper methodological evaluation. Although developmental refinement of the intervention will probably not cease completely by this point, its adoption by multiple surgeons across different sites will increase variation in the patient case mix, driven by surgeons' practices and centre infrastructure and policies. One focus of studies in this stage should be to capture variation in practice. In addition, careful tabulation of patient characteristics could suggest potential covariates and confounders influencing outcomes.

## Nature and challenges of the exploration stage: preparing for a definitive evaluation

In 1987, Martin Buxton observed that "It's always too early (to do a randomised trial) until, unfortunately, it's suddenly too late."<sup>5</sup> In observing past innovations, the exploration stage is often the "tipping point" of a surgical innovation (for example, laparoscopic procedures)—as described by Everett Rogers, where adopters' characteristics act as drivers or barriers (figs 1 and 2).<sup>6</sup> Factors such as whether the technique is too complex or too onerous to learn, and the strength of physician or patient preferences might critically affect its adoption.<sup>7</sup> This point could also be described as a time of "clinical equipoise," because further exponential adoption of this innovation by "early majority" and "late majority" adopters is consistent with a conviction of likely efficacy (for example, trends in diffusion of laparoscopic surgery<sup>8</sup>). It is at this stage when changes in regulatory structure might have the most profound effects in promoting randomised controlled trials in surgery (for example, approval requirements from the US Food and Drug Administration for drug trial phases for proof of safety and efficacy).

Several factors are needed to facilitate a definitive evaluation (preferably a randomised controlled trial). These include gathering practical information and fully evaluating the effect of the innovation (benefits and harms) that earlier evaluations would be ill equipped to represent. In

### Box 1 | Recommendations for observational studies at stages 2b (exploration) and 3 (assessment)

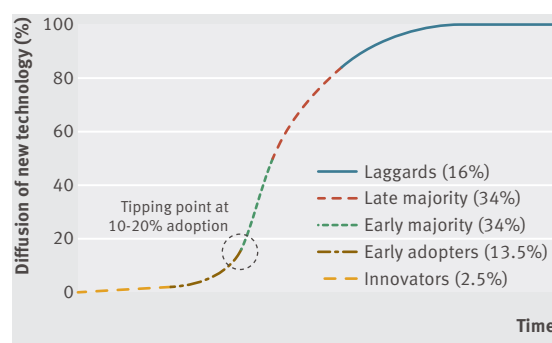
#### Exploration

Observational studies should generally be prospective and have a protocol  
A range of outcomes should be collected using standardised definitions  
Observational studies that are uncontrolled (for example, those based on registry and routine data collection) should be diagnosis based rather than procedure based whenever possible  
Important patient risk factors and variations in the interventions should be explored  
Studies should record and report surgeon experience (including any specific training received).  
Where possible, the effect of skill differences and learning should be assessed using appropriate data analysis  
Prospective, collaborative observational studies should be designed with a definite evaluation in mind (preferably a randomised controlled trial)

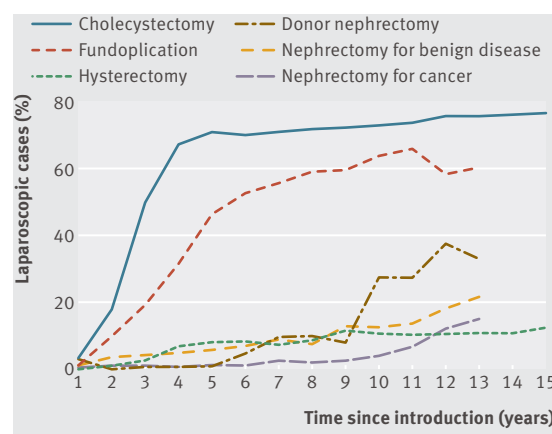
#### Assessment

Definitive observational studies should use a quasi-experimental study design; protocol driven controlled studies with standardised eligibility and prospective data collection  
Possible designs include non-randomised controlled trials and interrupted time series  
Key patient and centre characteristics likely to confound analysis should be considered before conducting study and collecting appropriate data, which would facilitate assessment and adjustment of the case mix, and help matching to control for potential confounding

**Fig 1 | Theoretical adoption curve showing the different stages (according to adopter type) in the diffusion of a surgical innovation<sup>6</sup>**



**Fig 2 | Example of surgical innovation: laparoscopic procedure adoption.<sup>8</sup>** Reproduced from reference 8 with permission. Data are percentage of operations carried out using a laparoscopic approach in 1989-2003, from the Nationwide Inpatient Sample, a nationally representative annual sample of hospital admissions in the United States



the meantime, the new intervention still needs appropriate evaluation, and the highest possible methodological quality of evidence from observational studies should be sought at this stage. Prospective (and possibly controlled) observational studies are the most likely design at stage 2b—their value can be maximised, based on four recommendations.

Firstly, observational studies should collect data for consecutive patients from multiple surgeons (and preferably multiple centres) undertaking the new intervention.<sup>9</sup> Ideally, these studies would also be based on disease or diagnosis rather than solely on a new procedure, which would include patients irrespective of subsequent treat-

ment. Such a prospective design is a substantial advance on the usual single surgeon (or single centre) retrospective case series of selected patients undergoing a novel intervention, which have predominated in the surgical literature. There is evidence that retrospective designs can be more susceptible to bias than prospective designs when comparing randomised studies with non-randomised (including both prospective and retrospective) studies.<sup>10</sup>

A well conducted, large prospective observational study can form the basis for identifying important patient characteristics (the case mix), technical intervention variables (including potential co-interventions), and clinical outcomes of interest. A recent example of this type of collaboration is the International Registry of Acute Aortic Dissection, which uses this design for evidence to guide surgical, endovascular, and medical practice in acute aortic dissection (box 2).<sup>11</sup> Data collection sponsored by professional organisations or the government can also help the conduct of later comparative observational studies (for example, the American College of Surgeons' national programme for surgical quality improvement).<sup>12</sup>

Secondly, studies at this stage should collect data for a range of outcomes using standardised definitions as well as key patient characteristics. Not only benefits but also harms should be assessed. Surgical research has focused considerably on the risks of short term harm (surgical complications), although with varying extensiveness and clarity. Standardised frameworks should be used—for example, the Dindo-Clavien system<sup>14</sup> for postoperative complications.

Thirdly, surgical skill differences and associated learning curves can affect outcomes,<sup>15</sup> and an evaluation of surgical variation and learning should be incorporated into study designs at this stage whenever possible.<sup>16</sup> We recommend identifying relevant variables that can measure the effect of skill and learning (for example, surgeon or centre "volume," operating times, quality measures, and appropriate outcomes), and analysing the data sequentially to assess learning, where possible.<sup>17</sup> In a sequential statistical analysis of cases (using a cumulative sum control chart) early in the use of robotic beating heart surgery, researchers detected several complications needing further investigation.<sup>18</sup>

Finally, studies should be conducted not necessarily to be definitive, but rather to prepare for a definitive evaluation study (preferably a randomised controlled trial). We suggest that professional or government bodies promote collaborative multicentre observational studies to evaluate important new interventions in their specialty, and incorporate their work as a strong foundation towards a definitive randomised controlled trial, as a secondary aim. Collected information can inform the timing of a trial (or another type of high quality, prospective study) with respect to equipoise, the key research question, and the appropriate study population. In addition, standardisation of the intervention, quality assurance techniques, and appropriate validated and measurable outcome measures can be assessed. Several successful examples of this approach to consensus development of a trial have been published.<sup>19</sup> In some circumstances, a feasibility or pilot

#### Box 2 | Example of observational study at exploration stage (2b)

##### International Registry of Acute Aortic Dissection study<sup>13</sup>

###### Clinical background at time of conduct

Aortic dissection is defined as a tear in the aorta

Acute aortic dissection (within 14 days of onset) needs urgent treatment because it is associated with increased mortality and morbidity

There are two types of aortic dissection (A and B), according to location

The effect of developments in surgical and medical management is uncertain

###### Design

Observational study with registry data collection

Eligibility was based on diagnosis—all patients with an acute aortic dissection in 12 large referral centres (six countries)

Study included 464 patients between 1 January 1996 and 31 December 1998

Data were collected at presentation and from routine hospital records until discharge

###### Findings

Physical findings at presentation were diverse, classic findings were often absent

For patients with a type A dissection, medical management was associated with a hospital mortality of 58%, compared with 26% mortality for surgical management

For patients with a type B dissection, medical management was associated with a hospital mortality of 11%, compared with 31% for surgical management

**Box 3 | Example of observational study at assessment stage (3)***Minimally invasive, open radical prostatectomy with and without robotic assistance*<sup>28</sup>

Clinical background at time of conduct

Open retropubic radical prostatectomy (RRP) is commonly used to treat prostate cancer

Use of minimally invasive radical prostatectomy (MIRP) with or without robotic assistance had been proposed as an alternative, and its use is increasing

*Design*

Non-randomised controlled trial nested within data collection from a population based registry

Men diagnosed with prostate cancer as their first and only cancer were eligible

Men who underwent MIRP between 2002 and 2005 (n=1938) were compared with men who underwent RRP (n=6899) using a propensity score adjusted statistical analysis

Registry data were linked with US Medicare administrative data

*Findings*

Compared with RRP, MIRP resulted in a shorter length of stay, fewer strictures, and fewer 30 days respiratory and miscellaneous surgical complications—but a higher occurrence of incontinence, erectile dysfunction, and 30 day genitourinary complications

Use of additional postoperative cancer treatments was similar for both approaches

trial could be a natural intermediate step between a prospective observational study and a definitive randomised controlled trial,<sup>20</sup> which can identify specific enablers and barriers.

**Nature and challenges of assessment (IDEAL stage 3)***Use of observational studies as a definitive evaluation in lieu of a randomised controlled trial*

Assessment is the stage in the IDEAL framework that requires a definitive evaluation, preferably a randomised controlled trial. On rare occasions, a randomised comparison might be considered unnecessary, owing to the magnitude of evidence from early evaluations (for example, the parachute scenario<sup>21</sup>). However, the risks of error due to bias are easily underestimated; therefore, as the magnitude of the treatment effect becomes smaller, one should be cautious about relying on such evidence. Criteria based on the signal to noise ratio suggesting that at least a 5-fold to 10-fold improvement in improvement or cure is needed for a randomised controlled trial to be considered unnecessary, have been proposed.<sup>22</sup> Few new interventions achieve such striking results, and most will need a randomised controlled trial to give confidence of their efficacy. More likely reasons for not using a randomised trial are that it is considered impractical; this can be due to anticipated recruitment difficulties, the low likelihood of a timely completion (for example, key technology becoming outdated by the end of the trial); or the study will be prohibitively expensive. In this scenario, careful consideration of how to obtain observational data of the greatest value and quality is particularly important.

Any observational study conducted as an alternative to a high quality, randomised controlled trial should have as many positive design features of such a trial as possible.<sup>23</sup> The study should have a prospective design with a detailed research protocol (ideally published at the outset) that clearly describes and defines a standardised intervention, the eligibility criteria and characteristics for patients being treated with the novel intervention, and the incorporation of quality control measures regarding delivery of the intervention. A unique circumstance when an observational study might be needed is if there

is no viable alternative therapeutic option (for example, organ transplantation of the heart<sup>24</sup> or liver<sup>25</sup> for severe advanced stage disease). Many examples of prospective, uncontrolled observational studies have successfully provided evidence to guide practice in surgery.<sup>26</sup>

We consider in turn two quasi-experimental designs: non-randomised controlled trials and interrupted time series. These designs are methodologically stronger options than uncontrolled prospective observational studies,<sup>27</sup> and could fulfil the role of a definitive evaluation when a randomised controlled trial is infeasible (box 3 shows an example).

*Non-randomised controlled trials*

The preferred observational design is a non-randomised controlled trial; a study in which a cohort of patients undergoing a novel surgical intervention is compared with a concurrent control group undergoing standard treatment (standard surgical, medical, or no treatment). The study should incorporate the positive design features associated with a randomised controlled trial (for example, a prospective design and standardised data collection), with the exception of randomisation and blinding. Such studies provided the first convincing prospective evidence for benefits in coronary artery bypass surgery<sup>29</sup> and laparoscopic cholecystectomy.<sup>30</sup>

In a randomised controlled trial, random allocation will probably achieve balance for known and unknown risk factors and minimise bias. Selection bias in a non-randomised controlled trial can be addressed by controlling for known risk factors (case mix) in the analysis. Relevant risk factors, how they should be documented, and potential for bias should be considered before starting data collection. Patient characteristics at study entry should be thorough documented. Treatment group assignment can, however, have a different risk pattern at baseline, and this can lead to groups being less comparable after statistical adjustment (for example, regression) owing to the “constant risk fallacy” where the assumption of constant risk across different organisations (for example, hospitals) may be inappropriate.<sup>31</sup>

Nevertheless, adjustment or matching for known prognostic factors should generally be done where possible (for example, using propensity scoring and corresponding analysis, which is an increasingly popular approach<sup>32</sup>), while recognising the limitations of such analyses. The estimated treatment effects can be assumed to be unbiased only if matching stratified analyses or regression techniques are sufficient enough to fully deal with risk imbalance—that is, when treatment allocation is ignorable in terms of baseline risk.<sup>33</sup> A cautionary example of the importance of risk adjustment is the Veterans Affairs National Surgical Quality Improvement programme’s study of long term outcomes after bariatric surgery. The survival advantage observed in the unmatched cohort disappeared when researchers used propensity scoring to analyse a matched cohort.<sup>34</sup> In some instances, results from randomised and observational studies have corresponded with respect to the magnitude of the effect size, although in general, observational studies have a greater risk of bias.<sup>35 36</sup>



## SUMMARY POINTS

Observational studies at IDEAL exploration stage (2b) should collect data for consecutive patients from multiple surgeons, including key case mix characteristics that are likely to influence outcome. Where appropriate, adjustment or matching should be used to control for potential confounding in the statistical analysis

Studies at the exploration stage should investigate the effect of technical parameters as well as skill and learning, assess the full range of outcomes, and prepare for a definitive (preferably randomised) evaluation at the assessment stage (3)

A non-randomised controlled trial or interrupted time series could fulfil the role of a definitive evaluation at the assessment stage if a randomised controlled trial is not feasible, or on rare occasions, considered unnecessarily

*Interrupted time series*

The interrupted time series is an alternative quasi-experimental design for an observational study that could potentially be used at the assessment stage.<sup>37</sup> The design uses a temporal rather than concurrent control group. A key outcome of interest (such as anastomotic leakage, graft failure, or death) is measured sequentially during a time period before the new intervention is introduced (that is, the interruption) and measured again during the same period afterwards.

Interrupted time series may be particularly suited to evaluating interventions that can be implemented at a centre with a long history of treating a particular disease (such as congenital heart disease). Although the design has been used to evaluate the effect of new interventions, it has not typically been used for evaluating clinical intervention efficacy. This design can be more susceptible to bias than non-randomised controlled trials, if not enough patient data are available to investigate and control risk factors. The design is particularly useful to assess secular trends in clinical care—that is, changes with time that could affect outcomes for all patients. The design should, whenever possible, be strengthened by adding a control group (that is, a parallel time series from a group where the new intervention is not used).

An interrupted time series has been used to track the effect of new surgical interventions (such as laparoscopic cholecystectomy on rates of bile duct injury<sup>38</sup>), evaluate quality of care (for example, in relation to rates of cardiac surgery mortality<sup>39</sup>), and estimate associated healthcare costs.<sup>40</sup> Surgical studies are often complicated by the nature of complex interventions and potential co-intervention effects (for example, medical and anaesthesia treatment of surgical patients), and an interrupted time series could isolate these effects by tracking the onset of factors (that is, interruptions) other than the surgical intervention itself.

**Summary**

After the refinement and definition of the innovation in small studies with short endpoints for preliminary investigations at IDEAL development stage, the evaluation of a new surgical intervention enters the exploration stage. At this stage, researchers should obtain the highest possible quality of evidence from prospective observational studies

and prepare for a definitive evaluation (preferably with a randomised trial design). Key factors for the evaluation to address include defining patient prognostic variables, characterising and standardising the surgical intervention, assessing learning, and identifying appropriate outcomes. Studies should use clear standardised definitions of key concepts, and be designed to promote a definitive evaluation at the assessment stage. Observational studies at the exploration stage should be based on a disease or indication, rather than just on the new procedure or technology of interest.

A randomised controlled trial is the preferred study design for definitive evidence and should be used wherever possible. But a high quality observational study may be acceptable if a trial is not feasible or, on rare occasions, deemed unnecessary. Observational studies should be carefully designed and conducted to maximally reduce the risk of bias. In such cases, quasi-experimental study designs should be considered (in particular non-randomised controlled trials or controlled interrupted series).

The Health Services Research Unit is core funded by the Chief Scientist Office of the Scottish Government Health and Social Care Directorates. Views expressed are those of the authors and do not necessarily reflect the view of the Chief Scientist Office or the funders.

Contributors: JAC and PM formulated the IDEAL series to which this paper belongs. PE and JB wrote the first draft of this paper, and JC, DA, and PM all commented on the draft. All authors approved the final version, and PE is guarantor. The papers were informed by the IDEAL workshop in December 2010.

IDEAL workshop participants (December 2010): Doug Altman, Jeff Aronson, David Beard, Jane Blazeby, Bruce Campbell, Andrew Carr, Tammy Clifford, Jonathan Cook, Pierre Dagenais, Philipp Dahm, Peter Davidson, Hugh Davies, Markus Diener, Jonathan Earnshaw, Patrick Ergina, Shamiram Feinglass, Trish Groves, Sion Glyn-Jones, Muir Gray, Alison Halliday, Judith Hargreaves, Carl Heneghan, Jo Carol Hiatt, Sean Kehoe, Nicola Lennard, Georgios Lyratzopoulos, Guy Maddern, Danica Marinac-Dabic, Peter McCulloch, Jon Nicholl, Markus Ott, Art Sedrakyan, Dan Schaber, Frank Schuller, Bill Summerskill.

Funding: The IDEAL group meeting in December 2010 was funded by the National Institute for Health Research's Health Technology Assessment programme, Johnson & Johnson, Medtronic and Zimmer (all unrestricted grants). JAC holds a Medical Research Council Methodology Fellowship (G1002292).

Competing interests: All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: PM received financial support from the National Institute for Health Research's Health Technology Assessment programme, Johnson & Johnson, Medtronic, and Zimmer for the IDEAL collaboration and for a workshop; no other financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

- 1 Barkun JS, Aronson JK, Feldman LS, Maddern GJ, Strasberg SM, et al. Evaluation and stages of surgical innovations. *Lancet* 2009;374:1089-96.
- 2 McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009;374:1105-12.
- 3 McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* 2013;346:f3012.
- 4 Cook JA, McCulloch P, Blazeby JM, Beard DJ, Marinac-Dabic D, Sedrakyan A. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ* 2013;346:f2820.
- 5 Buxton MJ. Problems in the economic appraisal of new health technology: the evaluation of heart transplants in the UK. In: Drummond MF. Economic appraisal of health technology in the European Community. Oxford Medical Publications, 1987:103-18.
- 6 Wilson CB. Adoption of new surgical technology. *BMJ* 2006;332:112-4.
- 7 Ergina PL, Cook JA, Blazeby JM, Bouton I, Clavien PA, Reeves BC, et al. Challenges in evaluating surgical innovation. *Lancet* 2009;374:1097-104.

- 8 Miller DC, Wei JT, Dunn RL, Hollenbeck BK. Trends in the diffusion of laparoscopic nephrectomy. *JAMA* 2006;295:2480-2.
- 9 McCulloch P. Developing appropriate methodology for the study of surgical techniques. *J R Soc Med* 2009;102:51-5.
- 10 Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30.
- 11 Tsai TT, Trimarchi S, Nienaber CA. Acute aortic dissection: perspectives from the International Registry of Acute Aortic Dissection (IRAD). *Eur J Vasc Endovasc Surg* 2009;37:149-59.
- 12 Hall BL, Richards K, Ingraham A, Ko CY. New approaches to the National Surgical Quality Improvement Program: the American College of Surgeons experience. *Am J Surg* 2009;198(5 suppl):S56-62.
- 13 Hagan PG, Nienaber CA, Isselbacher EM, Bruckman D, Karavite DJ, Russman PL, et al. The International Registry of Acute Aortic Dissection (IRAD): new insights into an old disease. *JAMA* 2000;283:897-903.
- 14 Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004;240:205-13.
- 15 Vickers AJ, Savage CJ, Hruza M, Tuerk I, Koenig P, Martínez-Piñero L, et al. The surgical learning curve for laparoscopic radical prostatectomy: a retrospective cohort study. *Lancet Oncol* 2009;10:475-80.
- 16 Lilford RJ, Braunholtz DA, Greenhalgh R, Edwards SJ. Trials and fast changing technologies: the case for tracker studies. *BMJ* 2000;320:43-6.
- 17 Cook JA, Ramsay CR, Fayers P. Statistical evaluation of learning curve effects in surgical trials. *Clin Trials* 2004;1:421-7.
- 18 Novick RJ, Fox SA, Kiaii BB, Stitt LW, Rayman R, Kadera K, et al. Analysis of the learning curve in telerobotic, beating heart coronary artery bypass grafting: a 90 patient experience. *Ann Thorac Surg* 2003;76:749-53.
- 19 Degiuli M, Sasako M, Ponti A, Calvo F. Survival results of a multicentre phase II study to evaluate D2 gastrectomy for gastric cancer. *Br J Cancer* 2004;90:1727-32.
- 20 Arnold DM, Burns KE, Adhikari NK, Kho ME, Meade MO, Cook DJ. The design and interpretation of pilot trials in clinical research in critical care. *Crit Care Med* 2009;37(1 suppl):S69-74.
- 21 Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 2003;327:1459-61.
- 22 Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007;334:349-51.
- 23 Black N. Complementarity comes of age. *Transplantation* 2008;86:28-29.
- 24 Robbins RC, Barlow CW, Oyer PE, Hunt SA, Miller JL, Reitz BA, et al. Thirty years of cardiac transplantation at Stanford University. *J Thorac Cardiovasc Surg* 1999;117:939-51.
- 25 Starzl TE, Klintmalm GB, Porter KA, Iwatsuki S, Schröter GP. Liver transplantation with use of cyclosporin A and prednisone. *N Engl J Med* 1981;305:266-9.
- 26 Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet* 2008;372:2152-61.
- 27 Shadish WR, Cook TD, Campbell DT. Experimental and quasiexperimental designs for generalized causal inference. Houghton Mifflin, 2002.
- 28 Hu JC, Gu X, Lipsitz SR, Barry MJ, D'Amico AV, Weinberg AC, et al. Comparative effectiveness of minimally invasive vs open radical prostatectomy. *JAMA* 2009;302:1557-64.
- 29 Hultgren HN, Pfeifer JF, Angell W, Lipton MJ, Bilisoly J. Unstable angina: comparison of medical and surgical management. *Am J Cardiol* 1977;39:734-40.
- 30 Attwood SE, Hill AD, Mealy K, Stephens RB. Prospective comparison of laparoscopic versus open cholecystectomy. *Ann R Coll Surg Engl* 1992;74:397-400.
- 31 Nicholl J. Case-mix adjustment in nonrandomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* 2007;61:1010-3.
- 32 Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. *Eur Heart J* 2011;32:1704-8.
- 33 Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR good research practices for retrospective database analysis task force report—part III. *Value Health* 2009;12:1062-73.
- 34 Maciejewski ML, Livingston EH, Smith VA, et al. Survival among high-risk patients after bariatric surgery. *JAMA* 2011;305:2419-26.
- 35 Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:iii-x, 1-173.
- 36 Shikata S, Nakayama T, Noguchi Y, et al. Comparison of effects in randomized controlled trials with observational studies in digestive surgery. *Ann Surg* 2006;244:668-76.
- 37 Matowe LK, Leister CA, Crivera C, Korth-Bradley JM. Interrupted time series analysis in clinical research. *Ann Pharmacother* 2003;37:1110-6.
- 38 Rutledge R, Fakhry SM, Baker CC, Meyer AA. The impact of laparoscopic cholecystectomy on the management and outcome of biliary tract disease in North Carolina: a statewide, population-based, time-series analysis. *J Am Coll Surg* 1996;183:31-45.
- 39 Marshall G, Shroyer AL, Grover FL, Hammermeister KE. Time series monitors of outcomes. A new dimension for measuring quality of care. *Med Care* 1998;36:348-56.
- 40 Sun P, Chang J, Zhang J, Kahler KH. Evolutionary cost analysis of valsartan initiation among patients with hypertension: a time series approach. *J Med Econ* 2012;15:8-18.