Assessing validity of observational intervention studies – the Benchmarking Controlled Trials

Antti Malmivaara

Centre for Health and Social Economics, National Institute for Health and Welfare, Helsinki, Finland

ABSTRACT

Background: Benchmarking Controlled Trial (BCT) is a concept which covers all observational studies aiming to assess impact of interventions or health care system features to patients and populations.

Aims: To create and pilot test a checklist for appraising methodological validity of a BCT.

Methods: The checklist was created by extracting the most essential elements from the comprehensive set of criteria in the previous paper on BCTs. Also checklists and scientific papers on observational studies and respective systematic reviews were utilized. Ten BCTs published in the Lancet and in the New England Journal of Medicine were used to assess feasibility of the created checklist.

Results: The appraised studies seem to have several methodological limitations, some of which could be avoided in planning, conducting and reporting phases of the studies.

Conclusions: The checklist can be used for planning, conducting, reporting, reviewing, and critical reading of observational intervention studies. However, the piloted checklist should be validated in further studies.

KEY MESSAGES

- Benchmarking Controlled Trial (BCT) is a concept which covers all observational studies aiming to assess impact of interventions or health care system features to patients and populations.
- This paper presents a checklist for appraising methodological validity of BCTs and pilot-tests the checklist with ten BCTs published in leading medical journals. The appraised studies seem to have several methodological limitations, some of which could be avoided in planning, conducting and reporting phases of the studies.
- The checklist can be used for planning, conducting, reporting, reviewing, and critical reading of observational intervention studies.

The experimental studies, randomized controlled trials (RCTs), provide the least biased information of the efficacy of medical interventions (1). However, RCTs mostly assess effectiveness of interventions in ideal settings and they focus on specific interventions. Their ability to assess effectiveness of clinical pathways or interventions targeting health care system features is limited. Thus there is an obvious need for valid observational data on actual performance in routine settings (2).

A recent paper presents the novel concept of Benchmarking Controlled Trial (BCT) and a comprehensive set of methodological criteria to be considered when appraising evidence from observational intervention studies (3). BCTs can be used to assess impacts of clinical interventions and impacts of features of the health care systems.

The aim of this paper is to create a simple checklist for assessing methodological validity of a BCT and to pilot-test the checklist with recent BCTs published in the Lancet and in the New England Journal of Medicine.

Methods

The original comprehensive checklist for methodological validity issues of BCTs was based on author’s previous work with RCTs, systematic reviews and observational studies (1,4–6). Also checklists and scientific papers for observational studies and respective systematic reviews were utilized (7,8). The current...
checklist was created by extracting the most essential elements from the comprehensive set of criteria in the previous paper on BCTs (open access: http://www.tandfonline.com/doi/full/07853890.2011.586901/07853890.2015.1027255) (3). The ten BCTs analyzed in the original paper on BCT were used to assess feasibility of the checklist created. The appraisal was rechecked and errors were corrected by the author.

Results

The ten main methodological issues and description of how to assess whether they possess a risk for bias are presented in Table 1. The issues 1 and 2 evaluate whether statistical power calculations were made, and whether there is a description of patient selection before patients were eligible to the study. Issues 3 and 4 consider documentation of baseline characteristics and how comparable are the index and reference groups. Issue 5 relates to documentation of processes, and issues 6 and 7 relate to outcomes and proportion of drop-outs. Issue 8 encompasses documentation of outcome relevant health care system features; and issue 9 covers the essential elements for producing high quality services in ordinary health care, particularly staff competence. Issue 10 evaluates whether the statistical analyses are appropriate.

The results of the pilot testing of the checklist show that there is considerable variation between the studies in realization of the methodological issues (Table 2). Of the ten validity criteria, two studies scored 7, one study 6, two studies 5, three studies 4, and two studies...
Table 2. Validity of recent Benchmarking Controlled Trials published in the Lancet and in the New England Journal of Medicine (3). Studies 1–5 assessed impact of clinical interventions, and studies 6–10 impact of health care system features.

Author, year, country

| Aim of the study | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------------|---|---|---|---|---|---|---|---|---|----|
| Coleman et al., Lancet Jan 8, 2011 | To assess between country differences for selected cancer survival | Yes | Yes | Unclear | Unclear | NA | b | Yes | Yes | 4 b |
| Pearse et al., Lancet Sep 22, 2012 | To assess mortality rates and patterns of critical care resource use for non-cardiac surgery patients across countries | Yes | Yes | Unclear | Unclear | NA | b | Yes | Yes | 5 b |
| Birkmeyer et al., NEJM Oct 10, 2013 | To assess the effect of surgical skill as a determinant for complication rates after bariatric surgery | No | No | Yes | Yes | Yes | Yes | Yes | No | Yes | 7 Paper provides no declaration of conflict of interests. All authors' declaration forms are available in the internet. |
| Karthikesalingam et al., Lancet Mar 15, 2014 | To compare in-hospital mortality of patients with rupture of an abdominal aortic aneurysm in two countries | No | Yes | Unclear | Unclear | NA | b | Yes | Yes | 4 b |
| Chung et al., Lancet April 12, 2014 | To assess 30-day mortality for acute myocardial infarction between two countries | No | Yes | Yes | Yes | Yes | Yes | Yes | NA | b | Yes | 7 b |
| Finks et al., NEJM June 2, 2011 | To assess impact of high-volume hospitals for decreased mortality after five major surgical procedures | No | No | Unclear | Unclear | Unclear | Yes | Yes | No | No | Yes | 3 Paper provides no declaration of conflict of interests. All authors' declaration forms are available in the internet. |
| Song et al., NEJM Aug 9, 2011 | To assess the effect of a quality system on health care spending and on quality of ambulatory care | Yes | No | Unclear | Unclear | Yes | Yes | Yes | No | No | Yes | 5 Paper provides no declaration of conflict of interests. All authors' declaration forms are available in the internet. |
| Wallace et al., NEJM May 31, 2012 | To assess the impact of night-time intensivist physician staffing for mortality of intensive care patients | No | No | Yes | Yes | Yes | Yes | Yes | No | No | Yes | 6 Paper provides no declaration of conflict of interests. All authors' declaration forms are available in the internet. |
| Sutton et al., NEJM Nov 8, 2012 | To analyze impact of a hospital pay-for-performance program with patient mortality in three acute diagnoses | No | No | Unclear | Unclear | Unclear | Yes | Yes | No | No | Yes | 3 Paper provides no declaration of conflict of interests. All authors' declaration forms are available in the internet. |
| Aiken et al., Lancet May 24, 2014 | To assess impact of nurse workloads and nurses' educational qualifications in hospital mortality after common surgical procedures in several countries | No | No | Unclear | Unclear | Unclear | Yes | Yes | No | Yes | Yes | 4 Authors declare no conflicts of interests |

aREM = Real Effectiveness Medicine framework, in which competence is considered the sine qua non for effectiveness in health care (2).

bThe study question includes impacts of the whole health care system including the clinical processes; therefore items 5, 8 and 9 are not needed for a valid answer to the study question in these studies. However, lack of information on items 5, 8 and 9 impair possibilities to make inferences of the reasons for between country differences.
3. Four studies made comparisons between countries, and consequently evaluate the impact of the whole health care system including all the clinical processes as determinants of outcomes. Therefore items 5, 8 and 9 are not needed for a valid answer to the study question in these studies. However, lack of information on items 5, 8 and 9 impair possibilities to make inferences of the reasons for between country differences.

One study presented statistical power calculations. Four studies fulfilled the criterion of information on selection of patients, because the whole catchment area (country) was covered. Three studies documented a valid and sufficient description of baseline characteristics in the index and control groups, and baseline comparability was considered adequate in these studies. Four studies showed valid and sufficient documentation of adherence to intervention, and description of other treatment processes. All the studies had sufficient documentation of the outcomes. The drop-out rates were acceptable, and the statistical analyses were appropriate in all ten studies. No study described health care system related features. Staff competence was evaluated only in one study, in which impact of surgical skill was the very study question.

Discussion

This paper presents, for the first time, a checklist for assessing validity of observational intervention studies, the BCTs. The checklist is intended for supporting planning, conducting, and reporting phases of the studies, and others should be acknowledged in the discussion. The piloted checklist is suggested for anyone interested in assessing validity of observational intervention studies. However, the checklist should be validated in further studies.

Acknowledgements

The author has developed the idea for the paper and written the manuscript solely. Riitta Malmivaara, MA, is acknowledged for productive discussions.

Disclosure statement

The author declares no support from any organisation for the submitted work; no financial relationships with any organisation that might have an interest in the submitted work; and no other relationships or activities that could appear to have influenced the submitted work.

Funding information

No outside funding.

References

1. Furlan AD, Pennick V, Bombardier C, van Tulder M, Editorial Board, Cochrane Back Review Group. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine (Phila Pa 1976). 2009;34:1929–41.
2. Malmivaara A. Real-effectiveness medicine—pursuing the best effectiveness in the ordinary care of patients. Ann Med. 2013;45:103–6.
3. Malmivaara A. Benchmarking Controlled Trial—a novel concept covering all observational effectiveness studies. Ann Med. 2015;47:332–40.
4. Hakkinen U, Malmivaara A. [Guest editors]. The PERFECT project: measuring performance of health care episodes. Ann Med. 2011;43. (Suppl 1). doi: 07853890.2011.586901./07853890.2011.586901.
5. Sihvonen R, Paavola M, Malmivaara A, Itala A, Joukainen A, Nurmi H, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. N Engl J Med. 2013;369:2515–24.
6. Croft P, Malmivaara A, van Tulder M. The pros and cons of evidence-based medicine. Spine. 2011;36:E1121–5.
7. Vandenbroucke J. When are observational studies as credible as randomised trials? Lancet. 2004;363:1728–31.
8. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 2007;4(10):e296.