

# BMJ Open Quality of randomised controlled trials in medical education reported between 2012 and 2013: a systematic review protocol

Martin G Tolsgaard,<sup>1</sup> Cheryl Ku,<sup>2</sup> Nicole N Woods,<sup>3</sup> Kulamakan Mahan Kulasegaram,<sup>4</sup> Ryan Brydges,<sup>5</sup> Charlotte Ringsted<sup>6</sup>

**To cite:** Tolsgaard MG, Ku C, Woods NN, *et al.* Quality of randomised controlled trials in medical education reported between 2012 and 2013: a systematic review protocol. *BMJ Open* 2014;**4**:e005155. doi:10.1136/bmjopen-2014-005155

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2014-005155>).

Received 28 February 2014  
Revised 14 July 2014  
Accepted 15 July 2014



CrossMark

For numbered affiliations see end of article.

**Correspondence to**  
Dr Martin G Tolsgaard;  
[martintolsgaard@gmail.com](mailto:martintolsgaard@gmail.com)

## ABSTRACT

**Introduction:** Research in medical education has increased in volume over the past decades but concerns have been raised regarding the quality of trials conducted within this field. Randomised controlled trials (RCTs) involving educational interventions that are reported in biomedical journals have been criticised for their insufficient conceptual, theoretical framework. RCTs published in journals dedicated to medical education, on the other hand, have been questioned regarding their methodological rigour. The aim of this study is therefore to assess the quality of RCTs of educational interventions reported in 2012 and 2013 in journals dedicated to medical education compared to biomedical journals with respect to objective quality criteria.

**Methods and analysis:** RCTs published between 1 January 2012 and 31 December 2013 in English are included. The search strategy is developed with the help of experienced librarians to search online databases for key terms. All of the identified RCTs are screened based on their titles and abstracts individually by the authors and then compared in pairs to assess agreement. Data are extracted from the included RCTs by independently scoring each RCT using a data collection form. The data collection form consists of four steps. Step 1 includes confirmation of RCT eligibility; step 2 consists of the CONSORT checklist; step 3 consists of the Medical Education Research Study Quality Instrument framework; step 4 consists of a Medical Education Extension (MEdEx) to the CONSORT checklist. The MEdEx includes the following elements: Description of scientific background, explanation of rationale, quality of research questions and hypotheses, clarity in the description of the use of the intervention and control as well as interpretation of results.

**Ethics and dissemination:** This review is the first to systematically examine the quality of RCTs conducted in medical education. We plan to disseminate the results through publications and presentation at relevant conferences. Ethical approval is not sought for this review.

## Strengths and limitations of this study

- The first systematic review of the quality of randomised controlled trials in medical education.
- The use of duplicate, independent and reproducible data coding of quality measures pertaining to research methodology and reporting.
- To provide a current state of evidence on trial quality, only studies reported from 2012 to 2013 are included in this review.
- Only articles in English are included in this systematic review.

## INTRODUCTION

Medical education as a field has grown during the past 20 years. It has become a billion dollar industry accounting for about US\$100 billion per year worldwide<sup>1</sup> and increasing awareness of linking education to patient outcomes has brought focus on evidence-based medical education.<sup>2</sup> The growing interest is reflected in the rise in the number of publications within this area over the past several decades.<sup>3</sup> However, this is not unproblematic, as several scholars have warned that medical education research lacks methodological rigour.<sup>3</sup> In a study of randomised controlled trials (RCTs) published between 2000 and 2003, a large proportion fell short of the criteria developed by the International Committee of Medical Journal Editors for reporting RCTs.<sup>4</sup> Meanwhile, some argue that judging the quality of research being performed in medical education by any 'objective' checklist is insufficient.<sup>5</sup> Instead, the quality of medical education research should be based on the advancement of our theoretical understanding, rather than on how well a particular research methodology has been adopted.<sup>5</sup> Other viewpoints state that

whatever method is used should comply with the highest standards of practice for that design.<sup>6</sup> Thus, two discourses of evaluating quality have been promoted. One is assessing quality against 'gold standards' such as checklists and guidelines, and another is judging the advancement of theory.

In clinical epidemiological research, RCTs take on a central role when evaluating healthcare interventions. Since 2000, the CONSORT group has provided guidelines to improve the transparency and rigour when reporting randomised trials within biomedicine.<sup>7</sup> Although the CONSORT statement does not include recommendations for designing, conducting and analysing trials, it indirectly affects design and conduct as transparent reporting may expose deficiencies in research if they exist.<sup>7</sup> Furthermore, CONSORT is informed by methodological theorists and practitioners in clinical epidemiology as well as biostatistics. Assessing the quality of RCTs in medical education using the CONSORT statement may, however, not capture advancement of theory. Insufficient use of a conceptual theoretical framework may lead to failure in identifying the active component of training interventions. Furthermore, a poor description of the context of the study as well as trainee characteristics limits the external validity in terms of generalisability to other settings and populations. Reporting should therefore also relate the study to a relevant theoretical context to justify how it uses and advances existing theory<sup>5</sup> including thorough descriptions of context, educational intervention and control circumstances and trainee characteristics.<sup>8</sup> However, these aspects are not assessed using the CONSORT statement, and other measures to evaluate study quality within medical education research may be warranted. To further evolve our understanding of the quality of RCTs conducted in medical education, we aim to explore the adherence to standardised quality criteria as well as the use of theory in recent literature. The research question of this review is:

In randomised controlled trials in medical education reported between 2012 and 2013, what characterises the quality of papers published in journals dedicated to medical education compared to papers published in biomedical journals with respect to objective quality criteria?

## METHODS

This systematic review is designed according to the seven-step approach recommended for conducting systematic reviews in medical education<sup>2</sup> and reported according to the PRISMA statement.<sup>9</sup>

### Study eligibility

Broad inclusion criteria are used to obtain a broad range of randomised trials in medical education. Studies published between 1 January 2012 and 31 December 2013 in English are included. This period is chosen as new guidelines for reporting randomised trials were

published in June 2010 and previous studies argued that the evaluation of reporting guidelines should first be evaluated 18–24 months following publication.<sup>10 11</sup> All research papers in medical education using randomised designs are included. Medical education research is defined as 'any original research study pertaining to medical students, residents, fellows, faculty development or continuing medical education for physicians.'<sup>12</sup> Using this definition, studies on veterinary, nursing, pharmacist, physiotherapist and dentistry education research are not eligible. Parallel group studies, crossover studies and non-inferiority and equivalence studies are all included, whereas pseudorandomised studies are not.

### Search

The search strategy is developed with the help of experienced librarians to search MEDLINE, EMBASE, CINAHL, PsychINFO, ERIC, Web of Science and Scopus for key terms. These terms include truncated search on random\* and MeSH terms relating to medical education (eg, Education, Professional). Related domains are also included in the search to account for research not categorised under medical education (eg, health profession education, simulation, undergraduate medical education, technology-enhanced education, clinical reasoning, skills assessment, education professional, student health occupation, internship and residency, curriculum planning, instructional method, self-directed learning, etc). The search is supplemented by adding the reference lists of recent reviews in simulation-based medical education and with authors' records of studies published in the period of interest. The authors' records are used to refine the search strategy in an iterative way so that as many relevant randomised studies as possible are included in the online search.

### Study selection

All of the identified studies are screened based on their titles and abstracts individually and compared in pairs to assess agreement so that all studies have been screened by two authors. Potential disagreement is solved by discussion until consensus is reached. If the title or abstract is insufficient for determining eligibility, the full text is reviewed. If consensus cannot be reached by two of the coauthors, the whole author team will decide whether to include the paper or not. The agreement between the raters is determined using intraclass correlation coefficients.

### Data collection process

Data are extracted from included studies by duplicate and independent scoring of each study using a data collection form. The data collection form consists of four steps. Step 1 includes confirmation of study eligibility; step 2 consists of the CONSORT checklist; step 3 consists of the Medical Education Research Study Quality Instrument (MERSQI) framework; and step 4 consists of a Medical Education Extension to the CONSORT statement developed by the review group.

*Step 1:* The first step includes confirmation of study eligibility, extraction of study ID (created by the review author) as well as the name and focus of the journal (medical education/biomedical).

*Step 2:* The dichotomous CONSORT checklist is completed by ticking off each item when either present (=1), absent (=0) or not applicable (NA). The CONSORT statement recommends that researchers provide a scientific background for the study as well as present specific objectives and hypothesis, and thoroughly describe the intervention and control conditions, randomisation procedure, data analysis and interpretation of results.

*Step 3:* The Jadad Scale for reporting randomised trials<sup>13</sup> is used to assess the methodological rigour of the included studies. It consists of three items pertaining to randomisation procedure, blinding and participant withdrawal or dropouts.

*Step 4:* The MERSQI<sup>12</sup> is used to provide an established measure of study quality and scores are compared in pairs and discussed until consensus. Evidence of validity of the MERSQI framework has been established in a previous study.<sup>12</sup> The MERSQI framework provides a measure of trial size (single or multiple institutions), validity of assessment instruments used, and the Kirkpatrick level of outcome measures used (a taxonomy for classifying training programmes). Hence, studies of a certain size that focus on patient outcomes would receive higher scores than single-institution studies that assess the impact of interventions on healthcare professionals' knowledge or behaviour in a simulated setting.

*Step 5:* The Medical Education Extension (MEdEx) is developed by the study group through a literature review of relevant quality research in medical education. To further advance our understanding of the use of theory in the scientific background of the RCTs, the reporting of specific hypotheses, clarity of description of interventions<sup>14</sup> and controls, and the use of theory in the interpretation of the observed results, we chose to include these factors in a MEdEx to the CONSORT checklist. In step 4, the following items are therefore included: (1) Scientific background,<sup>5</sup> (2) Explanation of rationale,<sup>5</sup> (3) Objectives or research question,<sup>4 6</sup> (4) Hypotheses,<sup>4 6</sup> (5) Description of the intervention and control circumstances,<sup>6 8 14</sup> (6) Interpretation<sup>4 5 12</sup> of results (see online supplementary appendix).

### Statistical analysis

Inter-rater reliability is calculated using Intra-Class Correlation Coefficients. In the event of disagreement, assessments will be solved by consensus. Descriptive statistics for each of the three different quality measures will be performed. Logistic regression will be performed using journal type as dependent and CONSORT scores, MERSQI scores and MEdEx scores as predictor variables. Multiple regression using the same predictor

variables and journal impact factor will also be performed to assess the relation between quality measures and the journal impact factor.

### DISCUSSION AND DISSEMINATION

In parallel with the rise in publications in medical education over the past decades, increasing interest is being paid to systematically evaluate the quality of research conducted within this field. We chose to include three different quality measures in the data collection form for this systematic review. To evaluate the quality of reporting, the CONSORT checklist is included as a measure of the degree to which current medical education research adheres to the guidelines endorsed by the *World Association of Medical Editors*, the *International Committee of Medical Journal Editors (ICMJE)* and the *Council of Science Editors*.<sup>15</sup> The second-quality measure includes the MERSQI framework, which has been used extensively in several recent reviews.<sup>16–18</sup> Although MERSQI scores have been shown to correlate with the journal impact factor,<sup>12</sup> it provides limited information on the use of theory or clarity in the description of interventions. Hence, to account for the use of conceptual theoretical frameworks in medical education RCTs, we plan to include a third quality measure in terms of the MEdEx framework.

We hypothesise that this review may demonstrate differences between different quality measures in RCTs reported in biomedical journals compared to those published in journals dedicated to medical education. We expect that RCTs reported in biomedical journals adhere more strictly to the CONSORT statement and use outcome measures that relate to the upper Kirkpatrick levels than RCTs reported in medical education journals. Finally, we hypothesise that RCTs published in medical education journals use theory in the rationale for their research question, methods and in their interpretation of results, whereas this may be missing in research published in biomedical or clinical journals.

The review results will be submitted for publication in a peer-reviewed general medical journal and disseminated through relevant international conferences.

The results of this review will help clarify the state of quality of education research by using common quality standards. The comparative analysis with clinical epidemiology will provide feedback for medical education researchers and contribute to raising the quality of research and improve the reporting of studies within this field.

### Author affiliations

<sup>1</sup>Centre for Clinical Education and the Juliane Marie Centre, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark

<sup>2</sup>The Wilson Centre, University of Toronto and University Health Network, Toronto, Ontario, Canada

<sup>3</sup>Department of Surgery, The Wilson Centre, University of Toronto and University Health Network, Toronto, Ontario, Canada

<sup>4</sup>Department of Family, Community Medicine and The Wilson Centre, University of Toronto and University Health Network, Toronto, Ontario, Canada

<sup>5</sup>Department of Medicine, The Wilson Centre, University of Toronto and University Health Network, Toronto, Ontario, Canada

<sup>6</sup>Department of Anesthesia, The Wilson Centre, University of Toronto and University Health Network, Toronto, Ontario, Canada

**Contributors** All authors contributed to the design of the review and approved the final manuscript. MGT and CR were responsible for conception of the review. CK, KMK, RB, NNW, MGT and CR were responsible for designing the review and the MEdEx. MGT drafted the first version of the protocol and CK, RB, NNW, KMK and CR critically revised the paper.

**Funding** This review has been supported by the Laerdal Foundation.

**Competing interests** None.

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

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

1. Frenk J, Chen L, Bhutta ZA, *et al.* Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet* 2010;376:1923–58.
2. Cook DA, West CP. Conducting systematic reviews in medical education: a stepwise approach. *Med Educ* 2012;46:943–52.
3. Lee K, Whelan JS, Tannery NH, *et al.* 50 years of publication in the field of medical education. *Med Teach* 2013;35:591–8.
4. Todres M, Stephenson A, Jones R. Medical education research remains the poor relation. *BMJ* 2007;335:335–5.
5. Eva KW. Broadening the debate about quality in medical education research. *Med Educ* 2009;43:294–6.
6. Stephenson A, Todres M, Jones R. Reply to Dorman *et al.*'s 'On evidence'. *Med Educ* 2009;43:390–1.
7. Schulz KF, Altman DG, Moher D.; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010;152:726–32.
8. Frambach JM, van der Vleuten CP, Durning SJ, *et al.* AM last page. Quality criteria in qualitative and quantitative research. *Acad Med* 2013;88:552.
9. PRISMA statement. <http://www.prisma-statement.org> (accessed 5 Jan 2014).
10. Prady SL, Richmond SJ, Morton VM, *et al.* A systematic evaluation of the impact of STRICTA and CONSORT recommendations on quality of reporting for acupuncture trials. *PLoS ONE* 2008;3:e1577.
11. Fuller T, Pearson M, Peters JL, *et al.* Evaluating the impact and use of transparent reporting of evaluations with non-randomised designs (TREND) reporting guidelines. *BMJ Open* 2012;2:pii:e002073.
12. Reed DA, Cook DA, Beckman TJ, *et al.* Association between funding and quality of published medical education research. *JAMA* 2007;298:1002–9.
13. Jadad AR, Moore RA, Carroll D, *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.
14. Hoffmann TC, Eructi C, Glasziou PP. Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials. *BMJ* 2013;347:f3755.
15. Altman DG. Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. *BMJ* 2005;330:1056–7.
16. Cook DA, Hatala R, Brydges R, *et al.* Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 2011;306:978–88.
17. Cook DA, Brydges R, Zendejas B, *et al.* Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med* 2013;88:1178–86.
18. Cook DA, Levinson AJ, Garside S. Method and reporting quality in health professions education research: a systematic review. *Med Educ* 2011;45:227–38.

## Appendix 1.

### Extension to the CONSORT statement

*The column on the left-hand side is selected CONSORT elements. The column on the right represents the coding extensions specific for this study. All of these additional items are rated on three-point scales.*

<p><b>2a Scientific background and explanation of rationale</b></p>	<ul style="list-style-type: none"> <li>• <u>Scientific background</u> (maximum score = 3) include the use of             <ol style="list-style-type: none"> <li>1) Educational instruments. Score = 1.0 <i>Example: Simulation-based medical education, use of assessment instruments with validity evidence.</i></li> <li>2) Educational concepts. Score = 1.0 <i>Example: Deliberate practice, self-directed learning.</i></li> <li>3) Educational theories. Score = 1.0 <i>Example: Cognitive load theory, developmental frameworks.</i></li> </ol> </li> <li>• <u>Explanation of rationale</u> is the clinical rationale or justification for conducting the study. Maximum score = 3.             <ol style="list-style-type: none"> <li>1) Clinical background. Score = 1.5. <i>Example: "laparoscopic surgery has long learning curves and complications occurs more frequently with inexperienced surgeons."</i></li> <li>2) Justification of the use of intervention. Score = 1.5. <i>Example: "Simulation-based training has been shown to be useful for initial training and may therefore reduce the number of complications..."</i></li> </ol> </li> </ul>
<p><b>2b Specific objectives or hypotheses</b></p>	<ul style="list-style-type: none"> <li>• <u>Objectives or research question</u> (maximum score = 3) include specifications of             <ol style="list-style-type: none"> <li>1) Setting and population (Each = 0.5)</li> <li>2) Intervention and control (Each = 0.5)</li> <li>3) Outcome measures (Each = 0.1)</li> </ol> </li> <li>• <u>Hypotheses</u> are proposed effects or mechanisms of action.             <ol style="list-style-type: none"> <li>1) Score = 3 if stated clearly as a hypothesis <i>Example: Our hypothesis was that...</i></li> <li>2) Score 1.5 if potential mechanisms of actions are stated but not explicitly called a hypothesis <i>Example: "Simulation-based training has previously shown improved operative performances and may therefore also reduce complications..."</i></li> <li>3) Score=0 if no mechanism of action is proposed or no specific hypothesis is suggested.</li> </ol> </li> </ul>

	<p>Example: Effective communication is difficult. We aimed to explore if a simulated patient programme improved students' confidence in...</p>
<p><b>5 The interventions for each group with sufficient details to allow replication, including how and when they were actually administered</b></p>	<ul style="list-style-type: none"> <li>• <u>Description of the use of the intervention and control</u> (maximum score = 3) include             <ol style="list-style-type: none"> <li>1) Detailed description of the type of intervention and control conditions. Score = 1 <i>Example: Type of simulation or type of learner interaction.</i></li> <li>2) Detailed description of instructions/information available to participants. Score = 1 <i>Example: Verbal or written instructions available prior to and during the intervention and additional resources such as textbooks, web-based learning material etc.</i></li> <li>3) Detailed description of the supervision/ assessment/ feedback provided, the amounts available and the qualifications/training of the persons providing supervision/ assessment /feedback. Score = 1. <i>Example: How much feedback was provided, how was it provided, by whom and for how much time?</i></li> </ol> </li> </ul>
<p><b>Interpretations</b></p>	<ul style="list-style-type: none"> <li>• <u>Interpretation of results</u> (maximum score = 3) includes             <ol style="list-style-type: none"> <li>1) Reported consistent with the observed results (Score = 1.5). <i>Example: "These significantly higher performance-scores suggest that simulation-training of junior surgeons may lead to superior performance in the OR".</i></li> <li>2) Integration of results and interpretation into existing educational theory. (Score = 1.5) <i>Example: "These results are consistent with cognitive load theory suggesting that..."</i></li> </ol> </li> </ul>