

REVIEW ARTICLE

PROS AND CONS OF RANDOMIZED CONTROLLED TRIALS AND BENCHMARKING CONTROLLED TRIALS IN REHABILITATION: AN ACADEMIC DEBATE WITHIN THE EUROPEAN ACADEMY OF REHABILITATION MEDICINE

A. MALMIVAARA¹, M. ZAMPOLINI², H. STAM³ and C. GUTENBRUNNER⁴

From the ¹National Institute for Health and Welfare, Helsinki, Finland; Orton Orthopaedic Hospital, Scientific Unit, Helsinki, Finland, ²Department of Rehabilitation, Foligno Hospital, Foligno, Perugia, Italy, ³Department of Rehabilitation Medicine, University Hospital Erasmus MC, Rotterdam and ⁴Department of Rehabilitation Medicine, Hannover Medical School, Hannover, Germany

The European Academy of Rehabilitation Medicine (EARM) held a debate in Hannover, Germany, on 1st of September 2016 on the pros and cons of randomized controlled trials (RCTs) and observational effectiveness studies (benchmarking controlled trials; BCTs). The debate involved a chairperson, a person presenting the substance of the debate, an opponent, and a rapporteur. The academicians participated in the discussion. Eight propositions and proposed statements formed the substance of the debate. There was agreement that a study question should be the starting point of an effectiveness study, and not the study method, i.e. RCT or BCT. The term “benchmarking” was questioned: does it mean market-oriented medicine? It was clarified that benchmarking refers to the methodological features of this study design: there must always be a comparison between peers. It was agreed that BCTs might be better than RCTs for use in rehabilitation studies, in which one often needs multi-centred studies, such as in the assessment of the effectiveness of pathways when there is complexity of processes, health systems, organizational issues, structures and facilities; or where interactions between therapists, doctors and patients differ between centres; and when assessing the implementation of rehabilitation. In addition, BCTs may deal with ethical issues, e.g. the acceptability of interventions, more easily than RCTs. Recommendations regarding the different approaches (RCTs or BCTs) should be provided by the scientific rehabilitation societies. Concern over the validity of BCTs was considered justified, as the validity criteria of BCTs cover all those related to RCTs and include the risk of selection bias between treatment arms. Appropriate description of the essentials of the study object, including adequate description of how the interventions were actualized in comparison to the study plan, are essential features for a valid and generalizable study for both RCTs and BCTs. BCTs are necessary to widen the evidence-base of effectiveness in rehabilitation. It was suggested that the rehabilitation field should support the concept of BCTs. It was proposed that education regarding BCTs is indicated, and stakeholders need to be convinced that BCTs are a valid alternative to RCTs.

EARM and other physical and rehabilitation medicine (PRM) bodies could advance the use of BCTs for clinical and health policy decision-making.

Key words: rehabilitation; randomized controlled trial; benchmarking controlled trial.

Accepted Apr 12, 2022; Epub ahead of print July 7, 2022

J Rehabil Med 2022; 54: jrm00319

DOI: 10.2340/jrm.v54.2511

Correspondence address: Antti Malmivaara, Finnish Institute for Health and Welfare, University of Helsinki, Helsinki, Finland. E-mail: antti.malmivaara@thl.fi

LAY ABSTRACT

The European Academy of Rehabilitation Medicine (EARM) held a debate on the strengths and limitations of randomized controlled trials (RCTs) and observational effectiveness studies, also known as benchmarking controlled trials (BCTs), in rehabilitation. The main substance of the debate involved eight propositions and four proposed statements. The term “benchmarking” was questioned: does it mean market-oriented medicine? It was clarified that, as benchmarking refers to the features of the study design; there must be comparison between peers. It was agreed that BCTs might be better than RCTs for use in rehabilitation studies: one often needs multi-centred studies and assessment of the effectiveness of pathways; the rehabilitation processes are complex, and health systems and organizational issues are essential; and the essential interactions between therapists, doctors and patients differ between centres. Also, BCTs may deal with ethical issues more efficiently than RCTs. It was recommended that both RCTs and BCTs should be used in rehabilitation research. An essential feature of a valid and generalizable study (for both RCTs and BCTs) is appropriate description of the essentials of the study object. BCTs were considered necessary for widening the evidence-base of effectiveness in rehabilitation, and the rehabilitation field should support the concept of BCTs. It was proposed that education regarding BCTs is indicated, and stakeholders need to be convinced that BCTs are a valid alternative to RCTs. The EARM and other physical and rehabilitation medicine (PRM) bodies should advance the use of BCTs for clinical and health policy decision-making.

BACKGROUND TO THE DEBATE

Academic Debates within the EARM are structured discussions between 2 experts who take a different position concerning a single relevant topic in the field of rehabilitation medicine (1). Academic Debates in Rehabilitation Medicine are a Collaborative Initiative of the EARM and the Journal of Rehabilitation Medicine. Based on an initiative by Bengt H. Sjölund and Gerold Stucki in the Foresight Committee and a decision of the General Assembly in 2015, this Academic Debate was held within the EARM in Hannover, Germany, on 1 September 2016.

The topic of the Debate was the pros and cons of randomized controlled trials (RCTs) and benchmarking controlled trials (BCTs). The general proposal of what is the role of BCTs in scientific research and, particularly in rehabilitation research, was discussed.

CURRENT DEBATE

The debate was chaired by Christoph Gutenbrunner. The pros and cons of RCTs and BCTs in rehabilitation were presented by Antti Malmivaara, Henk Stam commented on the issues under debate, and Mauro Zampolini was the rapporteur for the session. All academicians could contribute to the debate.

The debate was introduced by Christoph Gutenbrunner, emphasizing how goal setting and rehabilitation programmes should be based on evidence, and that it is useful to share decisions with the person experiencing activity limitation and participation restriction. He stated that the knowledge we have is based on current standards, which are based on RCT proof of evidence of the effects of certain interventions. However, he stated that we have difficulty (in using RCTs) in studies in the field of physical and rehabilitation medicine (PRM) because rehabilitation is different from prescribing drug therapy. Exercises and rehabilitation techniques depend on the health professionals' skill, and there is difficulty in applying the double-blind paradigm. There is no single solution, but this debate discusses the possibilities available, using the concept of BCTs.

Antti Malmivaara was invited to clarify for which questions experimental studies (RCTs) can provide answers for and which they cannot. He stated that an alternative study design to RCTs is the observational effectiveness study, a BCT.

PROPOSITIONS AND PROPOSED STATEMENTS

The debate issues were presented by Antti Malmivaara and were based on previous literature on experimental and observational effectiveness studies (2–10).

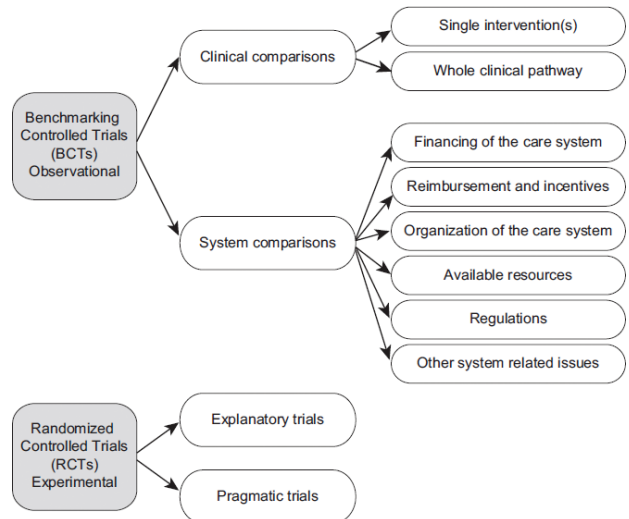


Fig. 1. Benchmarking controlled trials (BCTs) and randomized controlled trials (RCTs) cover all genres of effectiveness research.

The issues included a total of 8 propositions with explanations (indented below), and 4 proposed statements.

Study questions

Proposition 1. RCTs and BCTs cover all the study designs, which can provide evidence on effectiveness.

There are 2 options for assessing the effectiveness of interventions: an experimental study (randomized controlled trial, RCT) or an observational, study (benchmarking controlled trial, BCT). There are no other options besides doing experiments or just observing differences in effects (Fig. 1).

BCTs utilize comparisons between peers; healthcare providers treating similar patients, and there is always benchmarking involved. This is the reason for the term “benchmarking”.

The 6 impacts of healthcare are: accessibility, quality, equity, effectiveness, safety, and cost-effectiveness (Fig. 2).

Proposition 2a. RCTs can provide evidence on effectiveness, safety, and cost-effectiveness, but rarely on the other 3 impacts: accessibility, quality, and equity.

RCTs can provide effectiveness, safety and cost-effectiveness estimates, but accessibility, quality (mainly dependent on the competence of healthcare professionals) and equity issues are context dependent. These issues can be studied in each context by observing and comparing healthcare providers' performance.

Proposition 2b. In ordinary healthcare circumstances, comprehensive evidence of all 6 impacts must be based on observational effectiveness studies; BCTs.

In ordinary healthcare circumstances, estimates of effectiveness, safety and cost-effectiveness between peers treating similar patients can be obtained through observational effectiveness studies; BCTs. In addition, place and time dependent data on accessibility, quality and equity of services can be obtained only by BCTs.

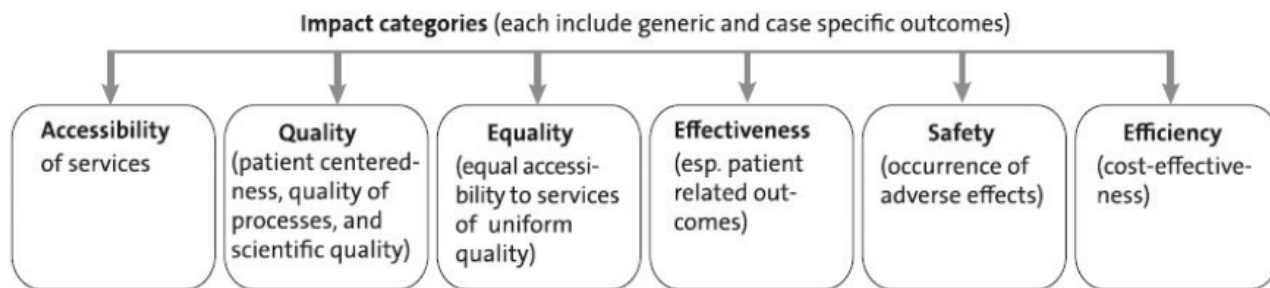


Fig. 2. The 6 impact categories of healthcare.

Clinical impacts

Experimental studies, RCTs, can answer certain, but not all, study questions regarding effectiveness. Therefore, in general, evaluation and research on effectiveness should start from the study question, following the choice of the best method (RCT or BCT) to answer the question (Fig. 3).

Proposition 3. Ethical, study question and feasibility issues (related often to rare and heterogeneous indications, heterogeneous interventions, and poor adherence in RCTs) are justifiable reasons for choosing a BCT design instead of a RCT design.

Study questions regarding healthcare impacts (including effectiveness) can be categorized into 3 groups: single intervention(s), clinical pathways, and performance comparisons between peers. An RCT is usually the most

valid method for obtaining data on the impacts of single interventions. However, BCTs can sometimes be used to complement evidence from RCTs, and in some cases, the BCT is the best or even only study design to answer the study question. The instances in which a BCT may be indicated can be placed in 3 categories: ethical reasons, contextual (study question), and feasibility reasons (Fig. 3). Ethical issues may be a contraindication for an experimental study; study questions may focus on time- and place-dependent issues, feasibility issues may be related to rare and heterogeneous indications, heterogeneous interventions, and poor adherence in RCTs. These factors may form justifiable reasons for choosing a BCT instead of an RCT design.

Proposition 4. RCTs can rarely study the impact of a clinical pathway.

The effectiveness of a clinical pathway (mainly when encompassing primary and secondary healthcare and social services) can rarely be studied with an RCT. This is because

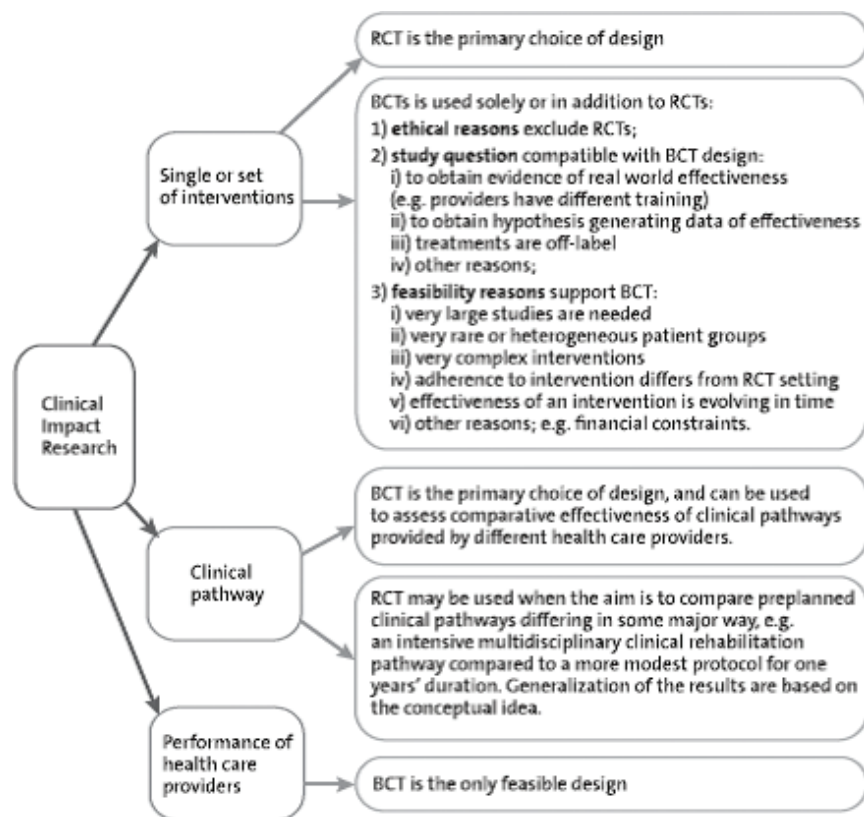


Fig. 3. Clinical impact research. Choosing the most appropriate study design when assessing: (i) impact of a single intervention or set of interventions; (ii) impact of a clinical pathway; and (iii) performance of healthcare providers (in routine healthcare circumstances) in relation to each other. RCT: randomized controlled trial; BCT: benchmarking controlled trial.

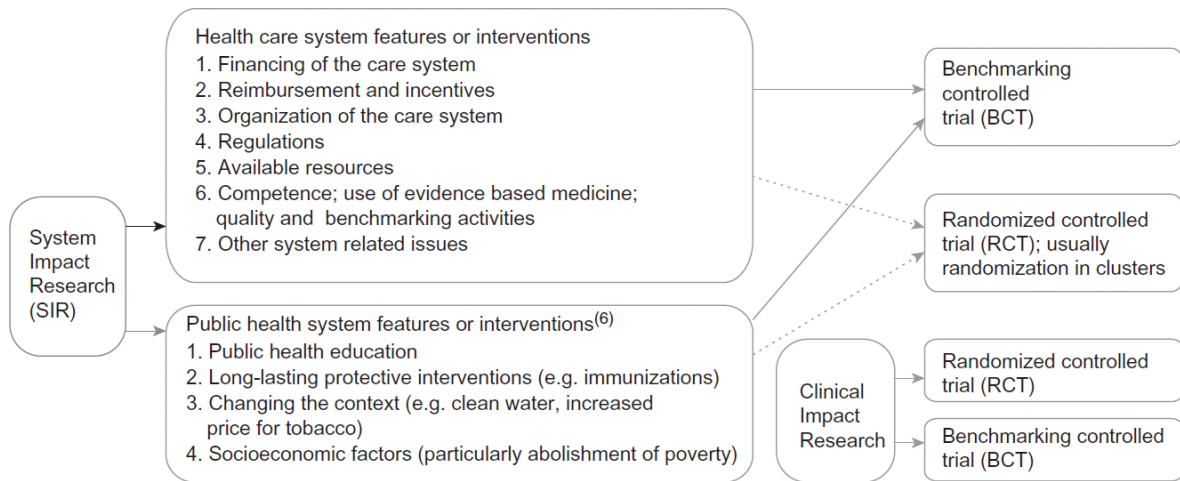


Fig. 4. System Impact Research includes all studies assessing performance of the health care or public health systems. All study objects are feasible for Benchmarking Controlled Trials, while many cannot be studied using a Randomized Controlled Trial design. The Clinical Impact Research is placed in the bottom right corner of the figure only to illustrate another category of impact research; i.e. that of assessing impact of interventions targeting individuals.

one would have to randomize different healthcare providers to either the experimental or the control clinical pathway. After that, the providers should implement the clinical pathways according to the experimental study protocol and only after successful implementation could the trial begin. An alternative would be to compare the existing clinical pathways using a BCT design, in which case there would be no need to implement pre-planned pathways as in the RCT design.

Proposition 5. Assessment of relative effectiveness between peers can never be studied by RCTs.

Assessment of the relative effectiveness between peers is, by definition, benchmarking, and randomization is not feasible.

System impacts

Proposition 6. BCTs must almost always be used when studying the impact of healthcare system features on patients.

BCTs can be used to assess the mutual effectiveness of different existing healthcare systems and features of the systems (Fig. 4). RCTs would need randomization in clusters, which is difficult and expensive to conduct; changing the systems to concord with the experimental protocol is demanding. The studies are time-consuming, and the results may be out of date by the time they are published. Moreover, there are usually problems with producing generalizable evidence using cluster-RCTs.

Validity of evidence of RCTs and BCTs.

Proposition 7. The healthcare system features, and staff competence are potential risks of bias in BCTs, but rarely so in RCTs.

All items that introduce a risk of bias in RCTs are also validity concerns in BCTs. In BCTs, the features of the healthcare system and staff competence are potential risks of bias factors, while they are rarely so in RCTs.

Generalizability of evidence from RCTs or BCTs.

Proposition 8a. Generalizability issues are similar for RCTs and BCTs, and generalizability is better the more comprehensive is the reporting of the study characteristics.

A good description of PICO components (Patients, Interventions, Comparison interventions, and Outcomes) in the study protocol and their assessment in the actual study form the basis for assessment of the applicability of the study results both in RCTs and BCTs.

Proposition 8b. Functioning (International Classification of Functioning, Disability, and Health; ICF), comorbid conditions, health behaviour, environmental, and equity issues are usually important for assessment of generalizability.

Functioning (ICF), comorbid conditions, health behaviour, environmental and equity issues are potential modifying factors for the effectiveness of interventions. Documentation of these factors serves 2 purposes. First, only by documenting these factors will one obtain information on to what degree these factors modify effectiveness. The second purpose is to compare these patient characteristics in ordinary healthcare with those of patients in the RCTs or BCTs.

To sum up, the same PICO principles for assessing generalizability refer to both RCTs and BCTs. The descriptive information needed is the same: selection of patients, and full description of patient characteristics (besides age, sex and disease-specific characteristics; also those of functional ability and health-related quality of life; comorbidities; behaviour: lifestyle (smoking, alcohol, exercise); environment (work, leisure); and socioeconomic conditions (education, income).

Statements for research implications

Statement 1. The feasibility, validity and generalizability of RCTs and BCTs should be studied further, both theoretically and empirically.

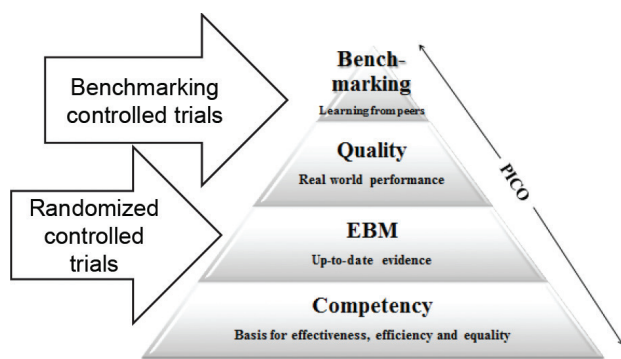


Fig. 5. Real-effectiveness medicine approach for assessing and promoting effectiveness in real-world circumstances (11). Choice of a randomized controlled trial (RCT) or benchmarking controlled trial (BCT) depends on the real-effectiveness medicine framework level of the study question. PICO: Patients, Interventions, Control interventions, Outcomes; EBM: evidence-based medicine.

Statement 2. Recommendations for future research on impact research methodology (feasibility, validity and generalizability) in rehabilitation are needed

Statements for clinical implications

Statement 3. Healthcare staff should be educated to understand the feasibility of RCTs and BCTs in assessing the impacts of healthcare interventions and their respective risks of bias and generalizability of results, in order to be able to appraise the evidence arising from these 2 study genres (Fig. 5).

Statement 4. Internationally uniform quality assessment systems must be instituted, and benchmarking activities promoted.

Response from the Debater and Discussion Among Academicians

The debater Henk Stam presented feedback regarding the issues raised by Antti Malmivaara. He considered that we are apparently at the beginning of a new way of doing research. He suggested that it would be helpful to make a list of the pros and cons of BCTs. Also, to consider how to make these propositions known to journal editors. In addition, he suggested that an algorithm would help in decision-making about which method to use. Antti Malmivaara agreed with these comments and emphasized that one should start with the study question and not with the method, which is what currently happens. This leads to consideration of the RCT as the gold standard, regardless of its ability to answer the research question. An RCT is usually the design of choice for assessing the effectiveness of individual interventions. However, BCTs are the design of choice for assessing the effectiveness of clinical pathways and organizational features.

Anthony Ward agreed that the starting point is how you formulate and apply the research question. To recruit enough participants in rehabilitation research one often needs multi-centred studies; and hence such studies can assess the effectiveness of pathways. The interactions between therapists, doctors and patients differ between centres. RCTs are considered the gold standard, but the control group poses a difficulty. BCTs may control the quality of both the index and control groups better than in RCTs. BCTs may discuss the ethical issues more efficiently. Any interaction in the setting potentially has a modifying effect on effectiveness. We should educate ourselves regarding BCTs, then convince stakeholders that the BCT is a valid alternative to RCTs. And, in rehabilitation, a BCT may be a more appropriate gold standard than a RCT. Antti Malmivaara commented that RCTs are the study design of choice when estimating the biological effects of interventions and needing a double-blind design. However, double-blind RCTs do not produce evidence for real-world circumstances, in which the placebo effect adds to the biological effect.

Stefano Negrini gave the example of scoliosis rehabilitation, in which there were no RCTs, and orthopaedic surgeons used this argument for not recommending bracing. There was a high-quality BCT in scoliosis, which showed the efficacy of bracing, but was not acknowledged until an RCT confirmed the findings. Moreover, observational studies should be performed after RCTs to assess the generalizability of the findings. Negrini stated that we really need to embrace this concept of BCT. And not only BCTs, but also other designs. Moreover, in some cases, effectiveness is so evident, e.g. for mobility devices, that not even BCTs are required; in these cases there is full agreement and no equipoise among clinicians, and it is not even possible to perform the research, since ethics committees would not approve it. It is also necessary to form alliances with other disciplines, e.g. surgeons, who encounter the same problems with the classical demand for using only RCTs to provide evidence for practice. Blinding as a risk of bias is a problem in rehabilitation, as it is rarely possible to perform a double-blind study design. In conclusion, we should support this concept.

Antti Malmivaara responded that there are only 2 options for obtaining evidence of effectiveness: experimental or observational study design. RCTs cover experimental studies, and BCTs cover observational studies. Thus, RCTs and BCTs cover all types of effectiveness studies. In terms of study object, they can be clinical or related to the healthcare system. The risk of bias when studying the clinical effectiveness of a single intervention is usually lowest when per-

forming a RCT. However, there are study questions in which the risk of bias is lower with BCTs than with RCTs, e.g. when adherence to the intervention is poor in RCTs.

Jean-Pierre Didier commented on the 6 impacts of healthcare. One must also consider the acceptability of the patients, the ethical dimension. Antti Malmivaara considered this an excellent point: rehabilitation is a matter of helping the patient; if the patients do not accept something, we should not apply that intervention. In RCTs, the intervention is directed to all recruited patients regardless of their acceptance (given that they have given their informed consent for participating in the study). In BCT design, acceptance is inherent.

Jean-Pierre Didier commented that the patient is a participant in the research; a BCT is for the market, not for the patient. Antti Malmivaara stated that he understands the point, but would not have come across the term “benchmarking” if he had not realized that an observational study always occurs in a benchmarking situation, i.e. benchmarking is the core methodological point in observational studies.

Bengt Sjölund stated that the problem of an observational study is not described; one does not have blinding of outcome assessment. One should hire independent outcome assessors to obtain blinded outcome assessment. Also, one should consider whether the effect is clinically meaningful or statistically significant? Bengt Sjölund stated, further, that the mere fact that you express interest in a person acts as treatment, and the biomedical part is very small. Could we replace a rehabilitation department with a hotel? Antti Malmivaara responded that in ordinary care double-blinding does not exist; therefore, if the study question relates to effectiveness in ordinary care, there is no rationale for double-blinding. The questions with outcome measures are relevant both for RCTs and BCTs. If one obtains statistically significant results, one should assess the proportions of patients recovering (number needed to treat figures) in the respective treatment arms and not assess clinical significance based on the mean differences between study arms.

Kristian Borg stated that there is concern over the black box in RCTs. Is one concerned about the closed black box in BCTs and how to open it? Antti Malmivaara stated that the 10 risks of bias criteria for BCTs include all those in RCTs, but also include a further 3 items related to non-randomized design. One criterion, both for RCTs and BCTs is obtaining enough information regarding the treatment process, i.e. factors covering the “black box”. Therefore, opening of the black box is needed both in RCTs and BCTs.

Stefano Negrini stated that how to manage bias is the problem in all scientific work in different research

situations. Johan Rietman stated that BCTs are helpful in cases where it is challenging to apply RCTs. He stated that there is an increasing number of BCTs in the Netherlands. His group performed a RCT in new technology, and realized that the RCT design compromised the ecological validity. Therefore, BCT might be a more suitable design. Jean Paysant stated that there is complexity in performance during the rehabilitation process. Also, participants change their behaviour during the study. Gilles Rhode stated that we need to assess the performance of health services, and that there should be a better way to investigate the complexity emerging in medicine. The approaches should be centred on the patient. Antti Malmivaara responded that he agreed with these important points.

Gerold Stucki stated that, in social science, there are quasi-experimental studies that are as valid as RCTs. The learning system in Canada is looking at the impact. There is an opportunity to link that with quality management. There are huge opportunities for the future. Antti Malmivaara stated that the competence of the units and the whole system should both be considered. Measurements of competence related to how the system works are needed. Continuous monitoring of patients, the interventions they obtain, the system, and the outcomes are also required.

Carlotte Kiekens proposed avoiding dualism between scientific and human aspects of medicine. She stated that research only has meaning if it is applicable to a real-life context. Both can strengthen each other, and we should embrace other types of research, such as studies in social science. The different approaches could be unified between our scientific society (ESPRM) and EARM. In addition, there are already several methods groups working on these issues within Cochrane.

Christoph Gutenbrunner raised the question of what does evidence-based mean: is it a focus on single treatments. Learning from peers, but also health systems and structures used by individuals, may be important. BCTs are more appropriate for answering questions of implementation of rehabilitation. Henk Stam raised issues regarding the validity and impact of results in effectiveness studies.

Anne Chamberlain asked about the competencies of persons and systems. And whether the BCT is feasible also in low economy countries. Antti Malmivaara commented that, in Canada, they have created the CanMeds-framework used primarily by physicians and other healthcare personnel. CanMeds includes 7 broad categories and respective subcategories. Competence is the cornerstone of everything in healthcare and is of utmost importance in rehabilitation. The competence of teams is important and can be audited, for example, in stroke centres.

Christoph Gutenbrunner emphasized the importance of the facilities and the workforce. How this can be proven in a scientific way? We cannot develop a guideline based on the criteria aiming solely at the treatment of a single disease. We cannot have a control country because there are so many differences. In the World Health Organization (WHO) they have no alternative for RCTs. How to use measures of workforce competency? Antti Malmivaara stated that the starting point could be where we are right now. Let us take the example of spinal cord injury (SCI) rehabilitation. We need documentation of baseline, rehabilitation and treatment procedures, and outcomes according to the ICF. How do SCI centres perform compared with each other? There are 6 impacts of healthcare, for all of which one can get information using BCTs.

Stefano Negrini stated that a RCT is good if you compare 2 therapies. A BCT is useful to compare rehabilitation units, but this is not possible with a RCT. Jorge Lains stated that patients are complex with several variables. RCTs do not represent the real world. Do you need big data?

Antti Malmivaara stated that, in BCTs, one needs good definitions of the study object and a good description of the populations, interventions, and outcomes in all comparisons. Also, in big data (extremely large data sets), you must have a design that includes a comparison.

Jean-Pierre Didier asked what is the meaning of benchmarking? He stated that clinical medicine should not be a market based on economics. Gerold Stucki echoed this, stating that transplantation surgeons in Switzerland have rejected the notion of benchmarking because it leads to the wrong conclusion. It is not a matter of industrial work. They use the term “comparative”. The benchmarking is the achievement mechanics. Antti Malmivaara responded that the idea of benchmarking is learning from the best, but some units are the best in some features, and other in other features. Benchmarking means that one strives for the best performance. The term “comparative” is a truism; science always makes comparisons. In comparison with other economists, health economists have a very similar view to that of clinicians of the aim of healthcare, i.e. producing health and well-being for the patients and the population, and they use the word “benchmarking”. The term “benchmarking” has been well accepted in Finland. The term “quasi-randomized study” is often used for observational effectiveness studies. However, this term is problematic: “quasi” means that something is not something and fails to define the concept in question. Antti Malmivaara have ended up with the term BCT because there must always be a benchmarking situation in an observa-

tional setting, i.e. comparison between peers treating similar patients.

Guy Vanderstraeten stated that a BCT constitutes an alternative. However, the outcome should be defined. Bengt Sjölund asked whether in a BCT one accepts historical data. Antti Malmivaara responded that in research in which one has a comprehensive preplanned register, historical data may be as valid as prospective in terms of selection bias if the data covers the population in question altogether.

Christoph Gutenbrunner asked for a simple example of a BCT. Antti Malmivaara presented an example of comparing hip fracture rehabilitation in one hospital with hip fracture rehabilitation in another hospital. There is always a comparison between peers, i.e. benchmarking, involved. BCTs can measure all 6 categories of impact research: accessibility of services, quality of services, equity of obtaining effective services, effectiveness, safety, and cost-effectiveness. For example, the accessibility of cardiac rehabilitation after myocardial infarction is relatively poor in Finland. After accessibility, one strives for quality in rehabilitation and equity of obtaining that quality, then effectiveness, safety and cost-effectiveness.

At the end of the debate Christoph Gutenbrunner gave the floor to the rapporteur Mauro Zampolini. Mauro Zampolini summarized the debate, as follows. In evidence-based medicine, we must refer to the best evidence from RCTs. The experience of the clinician and patients’ values create the 2 other pillars. We cannot always use RCTs. BCT is a way to apply quasi-experimental designs. BCT could be used after RCTs. BCTs can focus on organizational issues, e.g. stroke centres, and analyse which assess important factors for decreased mortality and better functionality. Outcome measures are a problem for RCTs and BCTs from the patient point of view. This debate issue should be a matter of further discussion in the Academy.

The 6 steps in planning, conducting and reporting effectiveness research in rehabilitation based on the presentation by Antti Malmivaara are described in Table I. The main findings and conclusions of the debate are summarized in Table II.

Table I. The 6 steps in planning, conducting and reporting effectiveness research

1	Decide what is the most relevant effectiveness question in your field. Describe it using the PICO-framework.
2	Decide whether an RCT or BCT is the best method for answering this research question.
3	Ensure that description of patient selection, patient characteristics, adherence to interventions and outcomes ensures assessment of applicability of the study results.
4	Ensure that the internal validity of the study is as good as possible.
5	Take care of complying with the study protocol when conducting the study.
6	Report according to suggestions for RCTs and BCTs.

PICO-framework: Patients, Interventions, Control interventions, Outcomes; RCT: randomized controlled trial; BCT: benchmarking controlled trial.

Table II. Main findings and conclusions of the European Academy of Rehabilitation Medicine (EARM) debate on the pros and cons of randomized controlled trials (RCTs) and benchmarking controlled trials (BCTs)

1	Both RCTs and BCTs are needed in effectiveness research, and they cover all effectiveness questions.
2	Some important study questions can be answered mainly or only by BCTs, particularly regarding the effectiveness of clinical pathways, effectiveness between peers, and effectiveness of structural features of healthcare systems.
3	RCTs and BCTs face similar methodological validity issues. The strength of RCTs is in the baseline comparability of the study groups, and the strength of BCTs is in the adherence to the interventions.
4	RCTs and BCTs face similar generalizability issues. Good description of study populations, adherence to interventions, and outcomes are needed for both study designs. Patient-reported outcomes (PROMs), and assessment of functioning (International Classification of Functioning, Disability and Health; ICF) are essential.
5	Education about BCTs should be increased. Editors of medical journals and stakeholders responsible for advancing the effectiveness of healthcare interventions should be convinced of the merits of BCTs.
6	EARM and other physical and rehabilitation medicine (PRM) bodies should promote BCTs for research, and for clinical and health policy decision-making.

CONCLUSION

Benchmarking refers to the feature of the study design of BCTs: a comparison between peers. Both RCTs and BCTs are needed in effectiveness research, and they cover all effectiveness questions. BCTs may be better than RCTs in rehabilitation studies: one often needs multi-centred studies and assessment of effectiveness of pathways; the rehabilitation processes are complex, and health systems and organizational issues may modify effectiveness. In addition, BCTs may deal with ethical issues more efficiently than RCTs. The strength of RCTs is in the baseline comparability of the study groups, and the strength of BCTs is in the adherence to the interventions. Appropriate description of the study object (patients' characteristics, how interventions were actualized in comparison with the study plan) is essential for all effectiveness studies; as well as patient-reported outcomes (PROMs), and assessment of functioning (ICF). The rehabilitation field is recommended to support the concept of BCTs. Education regarding BCTs is suggested to be undertaken, and stakeholders including medical journal editors are to be convinced that BCTs are a valid alternative for RCTs. The EARM and other PRM bodies are recommended to promote the use of BCTs for clinical and health policy decision-making.

ACKNOWLEDGEMENTS

The authors acknowledge the EARM members for their valuable comments in the debate in Hannover, September 2016: Kristian Borg, Professor, MD, PhD; Anne Chamberlain, Professor, MD, PhD; Jean-Pierre Didier, Professor, MD, PhD; Charlotte Kiekens, MD; Stefano Negrini, Professor, MD; Jean Paysant, Professor MD, PhD; Gilles Rhode, Professor MD, PhD; Johan Rietman, Professor MD, PhD; Bengt Sjölund, Professor MD, PhD; Gerold Stucki, Professor MD, PhD; Guy Vanderstraeten, Professor MD, PhD and Anthony Ward, Professor MD, PhD.

The authors have no conflicts of interest to declare.

REFERENCES

- Sjölund BH, Stucki G, Michail X. Debates in rehabilitation medicine: a collaborative initiative of the European Academy of Rehabilitation Medicine and the Journal of Rehabilitation Medicine & dear readers and authors of the Journal of Rehabilitation Medicine. *J Rehabil Med* 2016; 48: 485–480.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283 (15): 2008–2012.
- Vandenbroucke J. When are observational studies as credible as randomised trials? *Lancet* 2004; 363 (9422): 1728–1731.
- Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med* 2008; 5 (3): e67.
- Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; 340: c869.
- 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. *Int J Surg* 2014;12:1495–1499.
- Malmivaara A. Benchmarking controlled trial – a novel concept covering all observational effectiveness studies. *Ann Med* 2015; 474: 332–340.
- Malmivaara A. Assessing validity of observational intervention studies - the Benchmarking Controlled Trials. *Ann Med* 2016; 48 (6): 440–443.
- Malmivaara A. Clinical impact research – how to choose experimental or observational intervention study? *Ann Med* 2016; 48: 492–495
- Malmivaara A. System impact research – increasing public health and health care system performance. *Ann Med* 2016; 48: 211–215
- Malmivaara A. Real-effectiveness medicine – pursuing the best effectiveness in the ordinary care of patients. *Ann Med* 2013; 45: 103–106.