



APPLICABILITY OF EVIDENCE FROM RANDOMIZED CONTROLLED TRIALS AND SYSTEMATIC REVIEWS TO CLINICAL PRACTICE: A CONCEPTUAL REVIEW

Antti MALMIVAARA, MD, PhD

From the Performance Assessment of the Health and Social Service System, Finnish Institute for Health and Welfare, Helsinki, Finland

Background: The value of randomized controlled trials is dependent on the applicability of their findings to clinical decision-making. The aim of this study is to determine a definition and principles for the applicability of evidence from randomized controlled trials and systematic reviews.

Methods: This narrative review searched studies from PubMed and Web of Science databases using Cochrane Collaboration's Qualitative Evidence Syntheses guidance. Empirical studies were excluded. Based on the included studies, a definition for the concept and propositions for principles of applicability were formulated.

Results: A definition and 11 propositions are presented, 6 propositions having additional sub-propositions. Low risk of bias, ability to answer to specific questions, documentation of the details of how randomized controlled trials turned out, reporting of favourable and adverse outcomes, and systematic comparison of randomized controlled trials and clinical data were considered important. Biomedical randomized controlled trials have the widest applicability, while heterogeneity in study characteristics, human perception, behaviour, environmental, equity factors, and health economic issues lessen applicability. Obtaining applicable evidence is a gradual process. Methodological and substance expertise is necessary for assessing applicability.

Discussion: A definition of applicability and requirements for applicable evidence from randomized controlled trials to real-world contexts are presented. Propositions are suggested for any assessment of applicability of findings from randomized controlled trials, systematic reviews and meta-analyses.

Key words: applicability; generalizability; external validity; transferability; randomized controlled trial; systematic review; meta-analysis; benchmarking controlled trial.

Accepted May 4, 2021; Epub ahead of print May 11, 2021

J Rehabil Med 2021; 53: jrm00202

Correspondence address: Antti Malmivaara, Performance Assessment of the Health and Social Service System, Finnish Institute for Health and Welfare, Mannerheimintie 166, 00270 Helsinki, Finland. E-mail: antti.malmivaara@thl.fi

The pivotal question in using the evidence from randomized controlled trials (RCTs) in clinical medicine is contextual: To whom and under what circumstances do the results of this study apply? (1).

The Agency for Healthcare Research and Quality (AHRQ) Effective Health Care Program prefers to

LAY ABSTRACT

Clinicians' need for knowledge about a specific patient (or group of patients) is the underlying principle for applicability. Consequently, randomized controlled trials and systematic reviews should document all essential factors needed for clinical decision-making. Documentation of the study protocol (inclusion and exclusion criteria of patients, description of content of interventions and the outcome measures) is not sufficient. The documentation must also cover what actually happened in the randomized controlled trial, i.e. the characteristics of patients, the adherence to the index and control interventions, and the amount of co-interventions. Clinical registers using uniform documentation with randomized controlled trials increase the applicability of the research findings to clinical practice. The broadest applicability of findings comes from randomized controlled trials that assess the effectiveness of a single biological intervention for a well-defined disease using a valid biological outcome measure. Heterogeneity in study characteristics (patients, interventions and outcomes), and the presence of human perception (diagnosis, interventions and outcomes based on patient perception), and behaviour, environmental and equity factors, lessen the applicability of evidence. Randomized controlled trials must also report probabilities for favourable and adverse outcomes in order to increase the applicability of evidence.

use the term "applicability" rather than "generalizability", and defines it as "the extent to which the effects observed in published studies are likely to reflect the expected results when a specific intervention is applied to the population of interest under 'real-world' conditions" (2).

The international guidelines for reporting intervention studies aim for a uniform and transparent reporting that allows assessment of internal and external validity of study results (3–5). These guidelines have been widely endorsed by the leading general medical and specialty journals, and following these is mandatory for researchers submitting papers. Consequently, the definitions and principles of applicability (external validity, generalizability) in these guidelines influence how questions related to applicability are reported.

The Consolidated Standards of Reporting Trials (CONSORT) statement includes guidelines for reporting parallel group randomized trials, and defines generalizability as "external validity, applicability of the trial findings"; and "external validity", also called generalizability or applicability, is the extent to which the results

of a study can be generalized to other circumstances” (3). The CONSORT statement presents the principles for each major item of reporting, but does not address the question of how the reporting could optimize the generalizability of evidence from RCTs to clinical practice.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement does not include a definition of applicability (generalizability, external validity) (4). The issue of how the reporting could enhance the applicability of evidence from systematic reviews and meta-analyses to clinical practice is not addressed.

The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement provides guidelines for reporting observational studies, and defines generalizability as “external validity” (5). The question of how the reporting could enhance the applicability of evidence from observational studies to clinical practice is not addressed.

The 3 international guidelines listed above do not have a universal definition of applicability (generalizability, external validity) and do not comprehensively describe principles of how to increase the applicability of research evidence to clinical practice. It seems that common principles for applying the evidence of effectiveness from RCTs to clinical practice are lacking.

The preliminary aims of this paper were to search for studies that have pursued a definition and/or principles for applicability (generalizability) of evidence from RCTs and systematic reviews to clinical practice; and to describe the principles that they present. The primary aim is to pursue a definition for the concept of applicability in effectiveness research, and to present principles for how to apply research evidence to clinical practice. The ultimate aim is to facilitate better patient care by more valid interpretations of applicability of evidence from RCTs.

MATERIAL AND METHODS

Studies on conceptual issues related to applicability (generalizability) were searched in a narrative review, and the definitions and principles of how to assess and increase applicability of evidence from RCTs were extracted. The Cochrane Collaboration’s Qualitative Evidence Syntheses guidance was used with the intention to continue the search and extraction of information to the point where no additional information in relation to the aims of the paper was found (6, 7). PubMed and Web of Science databases were used without time or language limitations. Relevant papers were identified using the following key words in different combinations: conceptual, causal inference; applicability, external validity; generalizability; and transferability, transportability. When relevant papers were found, similar papers and papers that referred to the included paper were assessed for whether they should be added to the review. The review process aimed to find all relevant scientific publications related to the definition and principles of applicability of RCTs.

Information from a recent book *Clinical Research Transformed* was also included (8). Empirical studies statistically assessing the concordance between findings from RCTs and findings from real-world data were excluded, as the focus was on conceptual issues forming the basis for all empirical operationalizations.

Based on studies found in the narrative review, the primary aim of definition of the concept and principles of applicability for RCTs was pursued. The principles are presented in the form of propositions and sub-propositions.

RESULTS

Conceptual studies on applicability (generalizability)

High internal validity of a study indicates that the risk of biased findings is low, i.e. that the findings probably represent “the truth” within the specific context of the study (9). If the internal validity of a study is low, it is probable that the study findings are false. The core issue in applicability is that the study findings would also represent “the truth” within a specific clinical context. Consequently, there is a rationale for applying the results of a study only if the risk of bias is low (9). Also, it has been proposed that internal and external validity should be considered as a joint measure, the target validity, expressing an effect estimate with respect to a specific population (10). N-of-1 trials gather evidence of effectiveness from the individual patient to whom the evidence will be applied (11).

As a prerequisite for enabling the assessment of applicability (generalizability) it is suggested that documentation of each RCT (and systematic review) is performed at 2 levels: what the study was designed to be, and what it actually turned out to be (8). The latter level, what the study actually turned out to be, denotes that RCTs should document and report patient selection, patient characteristics, interventions, parameters that modify treatment effect, adherence to all interventions, and the outcome measures (1, 12).

The appraisal of transferability of RCT data to real-world circumstances are suggested to be based on a comparable description of both the source (RCT) and the target (clinical practice) domains (13). There should be sufficient documentation of what actually happens in the real-world context (12, 14–16).

Measured comparability between population datasets and randomized trials will enhance the range of policy-relevant research questions that can be answered (17). Statistical methods may be used to improve the applicability of a randomized trial to a target population (18–22). Propensity scores can be used to quantify the difference between the trial participants and the target population (23, 24).

Differences in adherence to the intervention between the RCT and the target population should be taken into consideration (25). Methods for transporting evidence

of the effectiveness of compound treatments to clinical practice have been proposed (26). Transportability of evidence may also depend on differences in the mechanisms that determine the outcome in the study and the target populations (27).

RCTs aim to assess the probabilities of change that the intervention causes in outcomes (including adverse effects) when it is used instead of another intervention (or lack of intervention) (8). When the outcomes are dichotomic, a Cox proportional hazards model or some newer regression model, such as the Hanley-Miettinen regression model, can be used in the analyses (28). When the outcome is continuous, the minimally clinically significant changes (or differences) in outcomes, and threshold values for good and poor outcomes are suggested to be determined, and the outcomes dichotomized correspondingly in order to determine respective probabilities using a logistic regression model (8).

Double-blind RCTs are indicated if the question is on the biological (or physical) effectiveness of an intervention (intervention effect per se, without placebo effect) (14, 29). If the study question is on the effectiveness of an intervention in the non-blinded circumstances of everyday healthcare, blinding of the patient or the therapist is not indicated (14, 29). The effectiveness of a clinical pathway or a feature of the healthcare system indicates the use of a cluster randomized RCT or, more commonly, an observational effectiveness study, a benchmarking controlled trial (BCT) (30).

Definition of applicability

Definition of applicability (generalizability): the extent to which the magnitude of effectiveness of an intervention for a specific patient (or specific group of patients) in clinical practice is similar to the magnitude of effectiveness in the results of a RCT or a systematic review of RCTs.

Propositions (principles) for applicability

All propositions relate to clinical interventions (directed towards patients) and most of the propositions also relate to interventions directed towards healthcare system features (in order to improve patient outcomes). The main references on which the propositions are based are presented after each proposition. All propositions are considered important by the author, and those without references are based on the thoughts of the author. The propositions are listed below, and a synopsis of the propositions is shown in Table I.

- *Proposition 1.* High internal validity (low risk of bias) of a RCT or a systematic review (including or excluding a meta-analysis) is a precondition for the study findings to be generalizable to clinical practice (3, 4).
- *Proposition 2.* Rationale for assessing applicability (generalizability) is that the clinicians or other decision-makers need knowledge from RCTs in order to get answers for a specific patient or for a specific group of patients (2).

Table I. A synopsis of the propositions for enhanced applicability (generalizability) of findings from randomized controlled trials (RCTs)

Proposition 1. High internal validity (low risk of bias) is needed.

Proposition 2. Rationale: clinicians or other decision-makers need for knowledge.

2.1. From the clinical context one looks retrospectively for the evidence.

2.2. For prospective judgements from RCTs to clinical practice a good description of clinical context is needed.

2.3. Making overall conclusions of generalizability to a patient population of a particular country is not justified.

2.4. N-of-1 RCTs provide applicable evidence for the patient in question.

Proposition 3. Two levels of documentation of RCTs: study design; and what the trial turned out to be.

Proposition 4. RCTs must report also probabilities both for favourable and unfavourable (adverse) outcomes:

4.1. Dichotomic outcomes: probabilities of outcomes between the index and treatment arms.

4.2. Continuous outcomes: probabilities for a clinically important change, an acceptable symptom state, and persistence of symptoms.

4.3. Patient-profile specific effectiveness estimates of RCTs tailored for individual real-world patients.

Proposition 5. Clinical disease/disorder specific registers.

5.1. Systematic comparison of data from RCTs and from clinical practice.

5.2. Statistical methods of transferability needed.

5.3. Population representative clinical registers needed.

Proposition 6. Broadest applicability when biomedical study object.

6.1. Heterogeneity in study population and multidimensionality of intervention lessens applicability.

6.2. Human perception, behaviour, and environmental and health economic issues lessen applicability.

Proposition 7. RCTs produce the best applicable evidence for questions on effectiveness of single interventions

7.1. Double-blind RCTs required for producing evidence of the intervention effect per se.

7.2. Open RCTs required for producing evidence of effectiveness in clinical practice.

7.3. Effectiveness of clinical pathways or features of the healthcare system require benchmarking controlled trials (BCTs; quasi-experimental studies).

7.4. Assessments of differences in effectiveness between healthcare providers require BCTs.

Proposition 8. RCTs produce study by study ever more applicable evidence.

8.1. If no plausible mechanism of action of an intervention, applicability remains uncertain.

8.2. Conclusions of no-effectiveness of interventions cannot be made unless a population-based sample, comprehensive description of the trial, and findings are repeatable.

8.3. If no generalizable research evidence exists, one cannot declare any research based inferences.

Proposition 9. Expert competence of substance, decision-making context, and methodology are all needed.

Proposition 10. All actors bear responsibility for advancing applicability of evidence.

Proposition 11. Principles of applicability cover preventive, curative, palliative and rehabilitative interventions.

- 2.1. The specific patient or group of patients determines the need for applicable evidence from RCTs; and from the clinical or other decision-making context one looks retrospectively for evidence published prior to the decision-making.
- 2.2. The validity of the judgement of generalizability of the evidence from RCTs prospectively to the clinical decision-making situation is dependent on how explicitly and comprehensively the clinical context is described.
- 2.3. The magnitude of effectiveness of results of a particular RCT or a systematic review is not universally generalizable to a wider population, e.g. it is not correct to say that the results of this particular study are generalizable to the patient population of a particular country. Neither, in contrast, is it correct to say that the results of a particular study are not at all generalizable to a particular country.
- 2.4. N-of-1 (number of 1) RCTs, using a before-after design for finding the most effective treatment for an individual patient, provide effect estimates that are applicable to the particular patient for whom the trial has been designed (11).
- **Proposition 3.** Precondition for adequate estimation of the magnitude of intervention effect in clinical practice is that characteristics of a RCT (or RCTs included in a systematic review) are documented comprehensively at 2 levels: what the study was designed to be and what it actually turned out to be (8).
- **Proposition 4.** In addition to differences in outcomes, RCTs must also report between the treatment arms, probabilities for favourable and unfavourable (adverse) outcomes in order to increase the applicability of the evidence.
 - 4.1. In case of dichotomic outcomes (e.g. mortality), probabilities of outcomes between the index and treatment arms should be presented (8).
 - 4.2. In case of continuous outcomes (e.g. pain), a dichotomization is needed to assess 3 probabilities: (i) the minimal clinically important change in the index and control treatment arms; (ii) the probability of reaching a patient-acceptable symptom state (e.g. in pain); and (iii) the probability of persistence of disturbing symptoms. For all 3 outcomes the threshold levels must preferably be determined based on the data of the RCT in question, rather than based on data from previous studies with similar study questions.
 - 4.3. Patient-profile specific effectiveness estimates (tailored for individual real-world patients) from RCTs increase the validity of assessments of the magnitude of intervention effect for a particular patient (or group of patients) (8).
- **Proposition 5.** Clinical registers using uniform documentation with RCTs increase the applicability of the research findings to clinical practice (17). The benchmarking method can be used as the reference for adequate documentation (30) (Table II).
 - 5.1. All relevant documented data from RCTs and all relevant documented data from clinical practice must be compared systematically to reach the most valid interpretations of the applicability of the research data to the clinical context (17).
 - 5.2. Statistical methods of transferability increase the accuracy of assessments of applicability of findings from RCTs to a specific patient population (18–22).
 - 5.3. RCTs undertaken within a population representative clinical register increase the applicability of research findings to clinical practice (17).
- **Proposition 6.** The broadest applicability of findings (in time and place) comes from RCTs that assess

Table II. The benchmarking method for assessment of applicability of evidence from randomized controlled trials (RCTs). The method can be used also for benchmarking controlled trials (BCTs), and RCTs or BCTs in systematic reviews and meta-analyses. (30, 32, 33)

Description of each item
RCT plan: PICOS (patients, intervention, control intervention, outcomes, study design)
Is the study design appropriate for answering the specified aims? Place and time of the intervention and number of patients/centres. Inclusion and exclusion criteria of the patients. What are the clinical interventions or system level interventions that are compared? What is the primary and what are the secondary outcomes?
How the RCT turned out to be
1. Selection of patients, healthcare system features
1.1. Selection of patients/population to the intervention
1.2. Patients' path
1.3. Reasons for exclusions
1.4. Patients declining participation
1.5. Pre-intervention therapy
1.6. Place and time of recruitment. Total number of patients; numbers per recruiting unit per year
1.7. Comprehensiveness of patient population of the catchment area
1.8. Healthcare settings; number of healthcare units
2. Baseline characteristics; how they turned out to be
2.1. Clinically important data relevant to the particular disorder/disease (age, sex, severity)
2.2. Functioning (disease-specific/generic disability, health-related quality of life)
2.3. Comorbidity, (at least 2 comorbid conditions)
2.4. Behavioural factors (smoking, alcohol consumption, substance abuse, exercise, obesity)
2.5. Environmental factors (type of work, living conditions)
2.6. Potential inequity (education, socioeconomic status, ethnic background)
3. Interventions; how they turned out to be
3. Interventions (s)
3.1. Completed index intervention(s) (%)
3.2. Completed control intervention(s) (%)
3.3. Cross over to index intervention (%)
3.4. Cross over to control intervention (%)
3.5. Co-interventions reported
3.6. Staff competence
4. Outcomes and follow-up
4.1. Valid outcome measurements
4.2. Follow-up percentage satisfactory
4.3. Reasons for dropping out reported

the effectiveness of a single biological intervention for a biologically well-defined disease using a valid biological outcome measure (12).

- 6.1. The more heterogeneity there is in the study population and the more multidimensional is the intervention the vaguer is the study object and, consequently, the less applicable are the findings (12).
- 6.2. The more human perception, human behaviour, and environmental and health economic issues are involved in a RCT, the less applicable are the findings (12).
- *Proposition 7.* RCTs are usually able to produce the most valid and best applicable evidence for questions on the effectiveness of single interventions (3).
 - 7.1. Double-blind RCTs produce evidence of the effectiveness of the core element of the intervention, e.g. a drug molecule, as the placebo effect is eliminated by the study design. The evidence of effectiveness may be highly generalizable in terms of the intervention effect per se, which is most important information. However, double-blind RCTs do not generally produce evidence of the magnitude of effect directly applicable to clinical practice, when a placebo effect is present (14, 29).
 - 7.2. Open RCTs, where patients and healthcare professionals know which treatment has been used, produce evidence of effectiveness that includes both the biological or physical intervention effect and the placebo effect, and thus the evidence corresponds to the conditions of clinical practice. However, the placebo effect may vary according to treatment setting and interaction between patient and healthcare provider, thus decreasing the applicability of the magnitude of the treatment effect (14, 29).
 - 7.3. When the study question is on the effectiveness of clinical pathways or features of the healthcare system cluster RCTs are needed. As the randomization in these study designs has been at the level of centres, the findings are primarily valid to the differences in changes within centres, and only secondarily to the differences in effectiveness at an individual level. Therefore, the magnitude of effectiveness at the individual level is less valid and less applicable than that obtained from individually randomized trials. Moreover, due to heterogeneity in the healthcare systems, the applicability of the findings is less than that from individually randomized trials. Due to these limitations, benchmarking controlled trials (quasi-experimental studies) are the design of choice for these study questions (30–33).
 - 7.4. When the study question is on comparing healthcare providers treating similar patients, a RCT is unable to answer the question, but observational effectiveness studies, benchmarking controlled trials (quasi-experimental studies) are needed (30). The aim is to increase the value of healthcare by benchmarking between peers treating similar patients.
- *Proposition 8.* The aim of RCTs is gradually (study by study) to produce ever more evidence applicable to each specific group of patients and, consequently, to progressively increase the magnitude of effectiveness of interventions in real-world settings.
 - 8.1. A key criterion for choosing interventions for RCTs is a plausible mechanism of action. If there is no plausible mechanism of action, the applicability of the research findings is uncertain.
 - 8.2. Conclusions of no-effectiveness of interventions whose effectiveness have been considered clinically plausible cannot be made definitively unless the study patients represent the whole spectrum of the clinical population, the description of the RCT is sufficient (regarding both what the study was designed to be and what it actually turned out to be), and the findings are repeatable.
 - 8.3. If there is no generalizable research evidence for a particular clinical context, it should be made explicit that no research-based interpretations can be made.
- *Proposition 9.* Assessment of the applicability of findings from RCTs and systematic reviews must be undertaken by expert groups that have competence particularly related to matters of clinical substance, decision-making contexts, and methodological issues (3, 4).
- *Proposition 10.* All actors (researchers, methodologists, healthcare professionals, decision makers, etc.) bear responsibility for advancing the applicability of evidence from RCTs to clinical practice.
- *Proposition 11.* Definition and propositions of applicability cover preventive, curative, palliative and rehabilitative interventions.

DISCUSSION

The aim of this paper was to determine conceptual issues (principles) relevant to the applicability of evidence from RCTs. Conceptual principles form the basis for empirical operationalizations, i.e. for studies statistically assessing the concordance between findings from RCTs and findings from real-world data. Thus, the principles presented in this paper should be considered when planning, undertaking and reporting empirical studies on the applicability of results from RCTs.

The definition of applicability presented in this paper considers the clinical context as the starting point, from

which to look retrospectively at previously published RCTs. Consequently, it is not possible to make inferences prospectively from RCTs to clinical situations unless the details of the real-world context are explicitly described.

In this paper the definition is better conveyed by the “applicability” than by the term “generalizability”. Applicability must always be judged on an (*ad hoc*) individual patient level, but the available research evidence can also be considered generalizable to a defined group of patients (to which the individual patient belongs). This thinking opposes attempts to grade generalizability of evidence from RCTs or systematic reviews to clinical medicine without specifying the clinical situation. Even if an illness does not currently exist in a certain country, the results may be generalizable once the illness does occur. And, if an intervention is found effective in 1 country, it may indicate a need to also implement it in another country. Lack of feasibility should not be considered as lack of applicability.

RCTs are suggested to provide case-specific evidence of the effectiveness of interventions for use by clinicians (34). The aim is to provide estimates of effectiveness from clinical research individualized to each particular patient (8, 34). It has also been suggested that evidence that is considered potentially generalizable represents only a working hypothesis to be evaluated within each clinical context (35).

A necessity for the appropriate assessment of applicability is that data from each RCT is documented regarding what the study was designed to be and what it actually turned out to be (8). Documentation of the study object (RCT) should be comprehensive, both with regard to the study plan (inclusion and exclusion criteria of patients, description of the index and control interventions, and outcome measures), and for how the experiment turned out to be (what were the characteristics of patients, including disability, quality of life, behavioural, environmental and equity factors; and what was the adherence to the index and control interventions, what were the percentages of cross-overs and what was the magnitude of co-interventions) (12, 15). The description needed for the assessment of applicability of evidence of effectiveness from RCTs, BCTs, and systematic reviews and meta-analyses is shown in Table II.

Data are needed both from the RCT, and from the source where the knowledge will be utilized. These sources have been called primary and target contexts (36), but it is suggested in this paper that the primary context is the clinical context where the knowledge is needed, and the corresponding RCTs remain the source contexts. Researchers seem to have a consensus that an appropriate documentation of both of these contexts is necessary for the assessment of applicability (13).

Medical records often lack data regarding essential

parameters that modify the treatment effect. For example, data on disease severity assessed on scales used in RCTs are not uniformly recorded in clinical practice. In order to optimize the assessment of applicability from effectiveness research to clinical medicine there must be similar documentation of patient characteristics (including selection), adherence to interventions (including those interventions that were not intended to be included in the study), and outcome measurements. Although a major challenge to produce, there is a strong need for disease-specific clinical registries that are planned and built on the same principles of design and include similar documentation to that of RCTs, from which the evidence of effectiveness is gathered.

Currently, as patient-profile-specific research data are not available from RCTs, clinicians need to gauge the magnitude of effectiveness based on their competence and on their judgement of the effect-modifying influence of the features of a particular patient-profile. Patient-profile-specific effectiveness estimates would decrease the need for the clinical judgement of applicability of findings (8).

Heterogeneity in study characteristics, and the existence of human perception, behaviour and environmental and equity factors lessen the applicability of evidence. However, the clinical relevance of studies in this category may be high, and the choice of study questions should not be based primarily on the degree of applicability of findings, but on the clinicians’ and societies’ need for evidence on effectiveness.

This study has several limitations. The literature search and appraisal of studies was not systematic, and there is no flow-chart describing the number of excluded studies. The aim of this narrative review was to search for studies until no further ideas relating to definition or propositions for applicability emerged, using the principles of the Cochrane Collaboration’s Qualitative Evidence Syntheses guidance. Some relevant studies may have escaped notice. However, the primary aim of presenting a definition and propositions for applicability has been achieved, and these are open for scientific discussion. Some of the propositions question the current thinking, and some present new ideas on the applicability of evidence from RCTs. All of the propositions provide a conceptual basis on which to build operationalization on applicability. Further conceptual research is needed.

CONCLUSIONS

The starting point for defining and assessing applicability (generalizability) has to be from the point of view of a clinician needing knowledge for a specified clinical situation. RCTs must report appropriately what the study was designed to be and what it actually turned out to be. To optimize the applicability of evidence from RCTs, the

essential data in the RCTs and in the clinical practice have to be reported in a similar way, and there are statistical methods to increase the comparability of the data. In addition to reporting the between-group differences in outcomes, the RCTs must report probabilities for favourable and adverse outcomes, and continuous outcomes must be dichotomized according to clinical importance. The concept and principles of applicability (generalizability) cover preventive, curative, palliative and rehabilitative interventions. Scientific and clinical discussion is needed regarding the definition and principles of applicability of evidence from effectiveness research.

ACKNOWLEDGEMENTS

The author developed the idea for this study, gathered evidence from the scientific literature and wrote the paper.

The author has no conflicts of interest to declare.

REFERENCES

- Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005; 365: 82–93.
- Atkins D, Chang SM, Gartlehner G, Buckley DI, Whitlock EP, Berliner E, et al. Assessing applicability when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* 2011; 64: 1198–207.
- 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.
- Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151: 264–9.
- Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; 4: e297.
- Harris JL, Booth A, Cargo M, Hannes K, Harden A, Flemming K, et al. Cochrane Qualitative and Implementation Methods Group guidance series – paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. *J Clin Epidemiol* 2018; 97: 39–48.
- Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC Med Res Methodol* 2013; 13: 118.
- Miettinen OS, Steurer J, Hofman A. *Clinical research transformed*. Cham, Switzerland: Springer; 2019.
- Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010; 152: 726–732.
- Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target validity and the hierarchy of study designs. *Am J Epidemiol* 2019; 188: 438–443.
- Vohra S, Shamseer L, Sam M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. *BMJ* 2015; 350: h1738.
- Malmivaara A. Generalizability of findings from randomized controlled trials is limited in the leading general medical journals. *J Clin Epidemiol* 2019; 107: 36–41.
- Lesko CR, Ackerman B, Webster-Clark M, Edwards JK. Target validity: bringing treatment of external validity in line with internal validity. *Curr Epidemiol Rep* 2020; 7: 117–124.
- Malmivaara A. Pure intervention effect or effect in routine health care – blinded or non-blinded randomized controlled trial. *BMC Med Res Methodol* 2018; 18: 91.
- Malmivaara A. Generalizability of findings from systematic reviews and meta-analyses in the leading general medical journals. *J Rehabil Med* 2020; 52: jrm00031.
- He Z, Tang X, Yang X, Guo Y, George TJ, Charness N, et al. Clinical trial generalizability assessment in the big data era: a review. *Clin Transl Sci* 2020; 13: 675–684.
- Stuart EA, Rhodes A. Generalizing treatment effect estimates from sample to population: a case study in the difficulties of finding sufficient data. *Eval Rev* 2017; 41: 357–388.
- Ackerman B, Schmid I, Rudolph KE, Seamans MJ, Susukida R, Mojtabai R, et al. Implementing statistical methods for generalizing randomized trial findings to a target population. *Addict Behav* 2019; 94: 124–132.
- Dahabreh IJ, Robertson SE, Steingrimsson JA, Stuart EA, Hernan MA. Extending inferences from a randomized trial to a new target population. *Stat Med* 2020; 39: 1999–2014.
- Dahabreh IJ, Robertson SE, Tchetgen EJ, Stuart EA, Hernan MA. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics* 2019; 75: 685–694.
- Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci U S A* 2016; 113: 7345–7352.
- Ackerman B, Schmid I, Rudolph KE, Seamans MJ, Susukida R, Mojtabai R, et al. Implementing statistical methods for generalizing randomized trial findings to a target population. *Addict Behav* 2019; 94: 124–132.
- Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *J R Stat Soc Ser A Stat Soc* 2001; 174: 369–386.
- Borah BJ, Moriarty JP, Crown WH, Doshi JA. Applications of propensity score methods in observational comparative effectiveness and safety research: where have we come and where should we go? *J Comp Eff Res* 2014; 3: 63–78.
- Westreich D, Edwards JK. Invited commentary: every good randomization deserves observation. *Am J Epidemiol* 2015; 182: 857–860.
- Hernan MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology* 2011; 22: 368–377.
- Cinelli C, Pearl J. Generalizing experimental results by leveraging knowledge of mechanisms. *Eur J Epidemiol* 2021; 36: 149–164.
- Hanley JA, Miettinen OS. An 'unconditional-like' structure for the conditional estimator of odds ratio from 2 x 2 tables. *Biom J* 2006; 48: 23–34.
- Malmivaara A, Armijo-Olivo S, Dennett L, Heinemann AW, Negrini S, Arokoski J. Blinded or nonblinded randomized controlled trials in rehabilitation research: a conceptual analysis based on a systematic review. *Am J Phys Med Rehabil* 2020; 99: 183–190.
- Malmivaara A. Benchmarking Controlled Trial – a novel concept covering all observational effectiveness studies. *Ann Med* 2015; 47: 332–340.
- 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. Assessing validity of observational intervention studies – the Benchmarking Controlled Trials. *Ann Med* 2016; 48: 440–443.
- Miettinen OS. On progress in epidemiologic academia. *Eur J Epidemiol* 2017; 32: 173–179.
- Polit DF, Beck CT. Generalization in quantitative and qualitative research: myths and strategies. *Int J Nurs Stud* 2010; 47: 1451–1458.
- Munthe-Kaas H, Nokleby H, Lewin S, Glenton C. The TRANSFER approach for assessing the transferability of systematic review findings. *BMC Med Res Methodol* 2020; 20: 11–15.