Open Access Full Text Article

COMMENTARY

# Efficacy and effectiveness trials have different goals, use different tools, and generate different messages

Franz Porzsolt[1,2]
Natália Galito Rocha[3]
Alessandra C Toledo-Arruda[3]
Tania G Thomaz[3,4]
Cristiane Moraes[4]
Thais R Bessa-Guerra[4]
Mauricio Leão[5]
Arn Migowski[6,7]
André R Araujo da Silva[8]
Christel Weiss[2,9]

[1]Health Care Research, Department of General and Visceral Surgery, University Hospital Ulm, Ulm, Germany; [2]Institute of Clinical Economics (ICE) eV, Ulm, Germany; [3]Department of Physiology and Pharmacology, Biomedical Institute, Universidade Federal Fluminense, Niterói, [4]Cardiovascular Sciences, Universidade Federal Fluminense, Niterói, [5]Institute of Nuclear Medicine, University Hospital Antonio Pedro, Niterói, [6]National Cancer Institute (INCA), Rio de Janeiro, Brazil; [7]National Institute of Cardiology (INC), Rio de Janeiro, Brazil; [8]Department of Mother and Child, Faculty of Medicine, Universidade Federal Fluminense, Niterói, Rio de Janeiro, Brazil; [9]Department of Medical Statistics, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany

Correspondence: Franz Porzsolt
Klinik für Allgemein-und Viszeralchirurgie, Universitätsklinikum Ulm, Albert-Einstein-Allee 23, 89070 Ulm, Germany
Tel +49 173 329 9774
Email franz.porzsolt@uniklinik-ulm.de

**Abstract:** The discussion about the optimal design of clinical trials reflects the perspectives of theory-based scientists and practice-based clinicians. Scientists compare the theory with published results. They observe a continuum from explanatory to pragmatic trials. Clinicians compare the problem they want to solve by completing a clinical trial with the results they can read in the literature. They observe a mixture of what they want and what they get. None of them can solve the problem without the support of the other. Here, we summarize the results of discussions with scientists and clinicians. All participants were interested to understand and analyze the arguments of the other side. As a result of this process, we conclude that scientists tell what they see, a continuum from clear explanatory to clear pragmatic trials. Clinicians tell what they want to see, a clear explanatory trial to describe the expected effects under ideal study conditions and a clear pragmatic trial to describe the observed effects under real-world conditions. Following this discussion, the solution was not too difficult. When we accept what we see, we will not get what we want. If we discuss a necessary change of management, we will end up with the conclusion that two types of studies are necessary to demonstrate efficacy and effectiveness. Efficacy can be demonstrated in an explanatory, ie, a randomized controlled trial (RCT) completed under ideal study conditions. Effectiveness can be demonstrated in an observational, ie, a pragmatic controlled trial (PCT) completed under real-world conditions. It is impossible to design a trial which can detect efficacy and effectiveness simultaneously. The RCTs describe what we may expect in health care, while the PCTs describe what we really observe.

**Keywords:** randomized controlled trial, pragmatic controlled trial, explanatory trial, pragmatic trial, ideal study conditions, real-world conditions

## Introduction

The efficacy of an intervention demonstrated under ideal study conditions (explanatory trial) will not necessarily predict the effectiveness of the same intervention described under real-world conditions (pragmatic trial). The pragmatic-explanatory continuum indicator summary (PRECIS) group presented a model that predicts that trials will generally be somewhere on a continuum between the two extremes of explanatory and pragmatic studies.[1–3] The PRECIS concept is based on the widely accepted assumption that randomization is an essential prerequisite for clinical trials, be it explanatory or pragmatic.

Randomization is a scientific tool, which is ideal to guarantee similar distribution of risk factors in the groups of an experimental trial and so reduce selection bias. However, randomization is not easy to apply in a clinical setting as it will compete with the patient's expectation. Patients trust their doctors to select and recommend the best

possible solution of health problems. There is sufficient evidence on the ethical, psychological, and legal significance of the need for doctor–patient communication[4,5] and some, but no consistent, evidence on the use of shared decision-making strategies.[6] Researchers and clinicians who work with patients will confirm the significant difficulties of recruiting patients for the participation in randomized controlled trials (RCTs). Five reasons may explain the difficulties of recruitment: misconceptions about trials, lack of equipoise, misunderstanding of the trial arms, variable interpretations of eligibility criteria, and paternalism.[7] Several reports from different areas of health care confirm these findings. The common denominator for the difficulties to recruit patients for RCTs is disagreement of study conditions with patient preferences and values. Although we know that only 5% of adults but 70% of children are included in oncology trials,[8,9] it is not yet possible to identify prospectively the adults who will finally participate in a clinical trial. Neither do we know methods that increase the motivation for participation.[9–11] The refusal of trial participation increases the sampling bias and affects the external validity of clinical trials.[12,13] Although this bias has been known for at least 15 years, so far, no solution has been found to avoid it.[14]

There is a consensus that we need new strategies to generate valid study results,[15] but there is no consensus on the need to allow the integration of preferences without inducing bias. The limitations of RCTs are theoretically well known, but it is still hard to convince nonclinicians that the enforcement of randomization in a preference-dominated word may induce even more bias than evidence. A solution is possible if we pay more attention to the preferences and involve clinicians as well as patients in the general discussions about the design of clinical trials.[7] An explicit and detailed proposal has been made several years ago,[16] but has so far not yet translated into a specific pragmatic controlled trial (PCT).

## Background

While discussing this topic, we have to describe the definitions we are using for efficacy, effectiveness, efficiency, and more recently also for the value of health care. We use efficacy to express that a biologic effect can be observed under ideal study conditions. Effectiveness means an effect is detected not under ideal but under real-world conditions. Efficiency considers the relationship of input and output, ie, it often considers the relation of efficacy and monetary costs. Value describes the individual perspective that might be quite different in two people with a broken finger: a piano player will be concerned about a complete restitution of a broken finger, while a lawyer may consider a broken finger not too important. These examples demonstrate that the assessment of health outcomes has to be considered from different perspectives and that the health outcomes can be described in different dimensions with different meanings and have different consequences for individual and societal decisions.

The core problem addressed in this project is related to the epidemiological and the clinical difference of efficacy and effectiveness. Epidemiologists are aware of the problem and are trying to find a consensus solution that offers the valid assessment of effects under ideal as well as real-world conditions. Clinicians see big difference of effects under ideal and real-world conditions. From a clinical perspective, there is no continuum between efficacy and effectiveness: a patient is diagnosed and treated following the rules of standard care, ie, under real-world conditions or under experimental conditions approved by an institutional review board. Such differences in efficacy and effectiveness of interventions have been demonstrated in many clinical scenarios.[17–24] We have a widely accepted tool (RCTs) to assess treatment effects under experimental study conditions, but we need a generally accepted tool (eg, a PCT) to assess treatment effects under real-world conditions.

When the aim of assessment is the description of effects under real-world conditions, any artificial modification of the natural history of care should be avoided. There are two possible ways to find a solution for the assessment of effectiveness. One may go for a compromise and accept some artificial modifications. This is the model the PRECIS group prefers. A second option is not to accept any artificial interference. In this case, the assessment has to be restricted to observation only. The essential requirements for reporting results of observational studies have been published by the Strengthening the Reporting of Observational studies in Epidemiology. group.[25] It may be necessary to include additional rules in the evaluation of observational studies to avoid various forms of bias such as sampling, selection, performance, attrition, and detection bias.[12,14,26]

In co-operative projects with other groups, we tried to propose solutions for the assessment of effects under real-world conditions without introducing modifications that may affect the outcomes.[27,28] This proposal was confirmed by Gaus and Muche,[29] but was rarely noticed by the scientific community probably because the proposal supports the concept of nonrandomized studies, which is generally not considered reliable. Indeed, in most situations, nonrandomized approaches are unable to generate meaningful data. We try in our research proposal to overcome this problem by recommending the inclusion of some additional steps in nonrandomized trials.

# Aim of the study

As explanatory trials should measure efficacy – the benefit a treatment produces under ideal conditions – and pragmatic trial should measure effectiveness – the benefit a treatment produces in routine clinical practice,[16] we expect only two optimal study designs to provide answers to these two questions.

The aim of this study is to describe the commonalities and differences of these two study designs.

# Methods

On the basis of clinical experience requests from many clinical colleagues and suggestions from the literature, we describe the most frequently followed sequence of steps conducted in an explanatory and in a pragmatic trial. These well-known steps were supplemented by additional steps that help to avoid various forms of bias. As far as possible, scientific evidence is provided to support the suggested supplements.

# Results

The 13 steps to assess efficacy and effectiveness are described in Table 1. The differences between explanatory and pragmatic trials are described in eight of the 13 steps. Five steps, #1–#3, #10, and #13, are identical in explanatory and pragmatic trials, while the remaining eight steps #4–#9, #11, and #12 are different in explanatory and pragmatic trials.

# Steps that are identical in explanatory and pragmatic trials

Step #1 is the structured study question according to the principles of evidence-based medicine.[30] It should be emphasized that the structured study question can be asked for two different reasons, either to appraise an already published paper or to design a planned study. For appraisal of a published paper the original PICO (Patient and her problem, Intervention, Control intervention, Outcome) system is widely adapted and translated.[13,31] When a new study is designed, it is important to define the intended but not the recorded outcome. For assessment of a published study, it is important to record the reported but not the intended outcome. Of course, these two outcomes should be identical in high-quality studies, which, unfortunately, is not always the case. Second, the intended outcome should be defined before defining the interventions. For designing a clinical trial, we recommend to change the sequence of the four parts from PICO to POIC, which describes the sequence of Patient–Outcome–Intervention–Control.

Step #2 is related to the distinction of primary and secondary outcomes. Secondary outcomes are often added to the main study question without formal power calculation. Primary outcomes need to be associated with a formal hypothesis including the dimension in which the selected

**Table 1** Thirteen steps to assess efficacy (explanatory trial) and effectiveness (pragmatic trial)

| Step | Explanatory trial (conclusions derived from experimental studies completed under ideal conditions) | Pragmatic trial (conclusions derived from observations of real-world health care) |
|---|---|---|
| #1 | Phrase the primary and secondary study questions according to the principles of evidence-based medicine | |
| #2 | Distinguish between primary and secondary outcomes of the study | |
| #3 | Define inclusion criteria | |
| #4 | Define exclusion criteria | Define any important risk factors related to any of the primary outcomes |
| #5 | Define treatment options | Identify the most frequently used treatments from existing database |
| #6 | Define appropriate study design according to primary study questions (superiority or equivalence or noninferiority study and set the limits for each study type) | Any pragmatic trial is a descriptive trial. There is only one common design for descriptive trials. Mention 95% CI |
| #7 | Define the hypothesis (expected difference of experimental and control), mention α-/β-error, calculate the needed number of patients to confirm the hypothesis | Any of the selected individual treatments is considered the best possible treatment (in the individual situation for the individual patient) |
| #8 | Ask eligible patients to sign informed consent for randomization, evaluation, and publication of data | Ask eligible patients to sign informed consent for evaluation and publication of data |
| #9 | Allocate the patients randomly to the treatment options of the trial | Allocate patients to treatment options according to individual preferences and results of shared decision making |
| #10 | Guarantee follow-up long enough to detect outcomes and most of the adverse effects | |
| #11 | Compare the results of the randomized groups | Compare only results of groups with identical baseline risks (ie, stratified according to high, intermediate, and low risk). Include results of the "any other treatment" group for specificity control |
| #12 | Apply the intent-to-treat (ITT) principle according to step #9 | Application of the ITT principle is not necessary as the risk groups were stratified |
| #13 | Confirm statistical significance only if clinical effect is relevant (save statistical energy) | |

**Note:** The differences of explanatory and pragmatic trials are described in 8 of the 13 steps.
**Abbreviation:** CI, confidence interval.

endpoint will be assessed, the definition of the expected difference between experimental and control groups, and the calculation of the patients needed to demonstrate the expected significant difference (or equivalence or noninferiority of the experimental and control group). In most RCTs, there will be only a single primary endpoint but several secondary endpoints. In pragmatic trials, usually several primary outcomes have to be considered, such as mortality, cost of treatment, and side effects. When multiple primary outcomes are included in a study, three additional aspects have to be considered: 1) separate power calculations for each endpoint, 2) a correction for multiple testing, eg, Bonferroni correction, and 3) allocation of individual patients to different baseline groups depending on the assessed outcome (eg, a patient with four coronary by-passes will be at high risk for the endpoint "mortality" but not necessarily for the endpoint "side effects of the study treatment").

Step #3 addresses the definition of the inclusion criteria in both explanatory and pragmatic trials. The inclusion criteria define the necessarily existing health problem(s) that qualify a patient to be considered for inclusion in the study. The difference of inclusion and exclusion criteria will be discussed below.

Step #10 points out that the follow-up has to be long enough in both types of studies to detect the majority of defined endpoints and most of the side effects.

Step #13 suggests "to save statistical energy". Conclusions from any trial will be valuable if two conditions are met. The detected effect has to be statistically significant and clinically important. Results that are supported by only one of the two necessary conditions should not be used in daily health care. Any statistical test will consume power, ie, "statistical energy", but decisions about the clinical importance will not. Therefore, it is recommended to decide first if a result is clinically important or not. In case of a clinically irrelevant result, the use of statistical power is a waste of time and energy. This time and energy can be saved by skipping the statistical test when a result lacks clinical importance. This consideration may be worthwhile especially in pragmatic trials because these trials will need more statistical power than explanatory trials (see step #2).

## Steps that are different in explanatory and pragmatic trials

Step #4 describes the exclusion criteria in an explanatory trial.

In pragmatic trials, no exclusion criteria should be defined. Any patient with a particular health problem who asks for health service under real-world conditions has to be served (eventually "watchful waiting") and, consequently, has to be included in the trial. A pragmatic trial that is supposed to describe the real-world condition has to include any served patient, which means there should be no exclusion criteria in pragmatic trials. As there will be no randomized control groups in a pragmatic trial like in an explanatory trial, it is necessary to allocate all patients who are included in a pragmatic trial to different risks groups. The patients have to be allocated to different risk groups at the time of inclusion in a pragmatic trial, but before start of treatment (to avoid confusion). The allocation to the risk groups is important for the evaluation of results; it is usually made electronically and depends on two types of information. First, it depends on the primary outcomes defined in step #2. The set of important risk factors may be different for each of the defined primary outcomes. Second, the risk factors have to be selected according to clinical evidence, should be easy to assess, and have to be assessed in any patient. These factors classify low-, intermediate-, and high-risk patients. With respect to the practicability of the study, the number of selected risk factors per endpoint should be kept as small as possible. Ideally, there will be some factors that have to be met for high risk and some that must not apply for low risk. Patients who qualify neither for high nor low risk should be classified as intermediate risk.

Step #5 defines the treatment options in an explanatory trial usually confined to one or two experimental and/or control treatments.

In pragmatic trials, the number of treatment options can neither be predicted nor limited. It is recommended to define special rules for evaluation and interpretation of pragmatic trials. To generate valid study results, the most frequently used and the most expensive treatment options should be identified from pre-existing databases. This information is helpful to generate realistic study questions that can be answered. Treatment groups that are too small for evaluation as separate groups have to be combined with other groups to a mixed group ("any other treatment"). The evaluation should be supported by propensity score mating procedures.[32]

Step #6 defines the appropriate study design according to the primary study question. This design will be different in superiority or equivalence or noninferiority trials. In addition, the necessary limits for confirmation of differences have to be defined.

In pragmatic trials that are descriptive trials, the mean values and 95% confidence intervals should be calculated for each of the defined outcomes and the investigated treatment groups.

Step #7 describes the hypothesis of explanatory trials, which is usually the expected difference of the outcomes in the experimental and control group. The selected α- and β-error and the number of patients needed to confirm the hypothesis have to be presented.

In pragmatic trials, any treatment selected for a particular patient is considered the best possible treatment for this individual patient in this special situation. The evaluation will demonstrate how frequently the intended goal could be achieved by different treatment options. As several outcomes can be assessed, it will be possible to complete a cost effectiveness analysis based on real-world data.

Step #8 requests to sign an informed consent for randomization, evaluation, and publication of an explanatory trial.

In pragmatic trials, the informed consent is necessary to justify for evaluation and publication of data.

Step #9 describes the random allocation of patients to treatment options in an explanatory trial.

In pragmatic trials, patients will be allocated to treatment groups according to their individual preferences and/or doctors' recommendations. This selection process presents advantages and disadvantages. The two advantages are the low dropout rate because everybody's preferences will be respected. Second, almost all patients will get a placebo effect in addition to the biomolecular effect of the treatment. As long as this placebo effect will not cause harm, there is no reason not to provide it to everybody to the best available treatment. There is solid evidence supporting the hypothesis that the induction of hope or the vocal confirmation of patient expectations is sufficient to improve the reported outcomes. An interesting experiment that supports this hypothesis was done at the Massachusetts Institute of Technology.[33] This experiment was the first placebo-controlled study in the scientific literature, which demonstrated an impressive effect using only placebos but no true treatment arm; the participants of both study arms received placebos confirming that the effective principle in this study was the transmission of vocal information. Of course, this experiment is never applicable to patients due to ethical limitations and medical doctors do not like to discuss it, but nevertheless it definitely demonstrates the power of information on human decisions. Whether we like it or not, it is not according to the rules of science when we just accept the results we like.

The disadvantage of this selection procedure is the lack of control for unknown risk factors. This is a difficult topic – the power of known and unknown risk factors – as it is hard to be compared directly. However, there is interesting indirect evidence that may answer this question.

Step #11 is related to the compared groups within the studies. In an explanatory trial, the comparisons have to be defined when the trial is designed.

In a pragmatic trial, this request can be met only partially. The most frequently used treatment options will usually be known at the time of trial design. Any other treatment options are summarized as a single important comparator called "any other treatment". Within each of the groups, the patients are stratified according to baseline risks (high, intermediate, low risk). The mixed group includes any treatment that is not represented in the other compared groups. Depending on the defined comparators, a particular patient may be allocated for evaluation to one of the comparator groups or to the "any other treatment" group. The comparison of nonrandomized groups is closely related to the power of known and unknown risk factors in clinical studies. As in a pragmatic trial, only groups with comparable baseline risks – each related to a specific study question – will be compared; the remaining uncertainty will be related to the unknown risk factors. If the effects of these unknown risk factors will be as large as the effects of known risk factors, we should see it in appropriately designed experiments. If we do not see it, we have to discuss the potential consequences.

Step #12 requests to apply the intent-to-treat (ITT) principle according to step #9 in explanatory trials.

In pragmatic trials, the application of the ITT principle is not necessary as the allocation of patients to the risk groups is defined by risk factors and will not be affected even if the treatment will change.

## Discussion

A PCT should be different by definition from an RCT. The differences in objectives and contents of the 13 consecutive steps in explanatory and pragmatic trials are listed in Figure 1. These differences can be standardized and lead to advantages and disadvantages of PCTs and RCTs, which have to be weighed.

The internal validity is easier to control in an RCT than in a PCT, but external validity can be controlled better in a PCT than in an RCT. Assuming that internal validity is as important as external validity, we have to conclude that both aspects of validity have to be considered. As there is no study design that can guarantee both types of validity within the same study,[29] there will probably be no other option than designing two types of studies to control for both types of validity. Other topics are related to the effects of confirmed or suppressed preferences and to the power of known versus unknown risk factors. These topics need to be discussed in more detail.

| Explanatory trial | Step | Pragmatic trial |
|---|---|---|
| Study question | 1 | Study question |
| Primary/secondary outcomes | 2 | Primary/secondary outcomes |
| Inclusion criteria | 3 | Inclusion criteria |
| Exclusion criteria | 4 | Risk factors |
| Treatment options compared | 5 | Treatment options frequently used |
| Select appropriate study design | 6 | Use descriptive trial design |
| Define individual study hypothesis | 7 | Apply selected treatment |
| Informed consent (randomization) | 8 | Informed consent (publication) |
| Allocation to treatment (random) | 9 | Allocation to treatment (preference) |
| Follow-up | 10 | Follow-up |
| Compare groups of an RCT | 11 | Compare groups of same baseline |
| Apply intent-to-treat principle | 12 | Intent-to-treat principle not necessary |
| Statistics if clinically relevant | 13 | Statistics if clinically relevant |

**Figure 1** Objectives and contents of the 13 consecutive steps in explanatory and pragmatic trials. The objectives and the contents of these two types of trials are identical in steps #1, #2, #3, #10, and #13. In steps #5, #8, #9, #11, and #12 the objectives are identical but the contents are different in explanatory and pragmatic trials. In the remaining steps #4, #6, and #7 the objectives (and consequently the contents) are different in explanatory and pragmatic trials.
**Abbreviation:** RCT, randomized controlled trial.

## Possible consequences for health care systems

If a new standard can be established to assess effectiveness and if it effectiveness can be distinguished from efficacy, we will be able to close the efficacy–effectiveness gap. By closing this gap, we will contribute to the solutions of several problems. It will become evident that some questions can be answered by efficacy data (generated under ideal study conditions), while others will need effectiveness data (generated under real-world conditions) to be answered.

The goal of basic research is to demonstrate the efficacy of a new principle under ideal study conditions. It makes sense to confirm this basic observation in a second independent experiment under the same ideal study conditions just to reduce the risk of random error according to the principle of repetition. It does not make sense to investigate additional questions such as the optimal target groups or optimal doses under ideal study conditions, as only a small minority of study participants will suffer from a single health problem and will get treatment only for a single health problem. Most study participants present several problems and get several treatments, which means the real world is much more complex – and we wish it should be the same in a clinical trial. Nobody contradicted when it was expressed very clearly that most of published research may be wrong.[34]

To deal with this complex problem, the optimal consequence may not be to adapt the target population and study conditions to the requirements of an explanatory trial, ie, excluding most of the variables and randomly distributing the remaining confounders. It may be possible to get closer to reality (real world) when the design of the study can be adapted to manage the diversity of variables and to avoid bias. Completing more real-world studies will not only increase the variability within a single study (as compared to an explanatory study with many exclusion criteria) but will also decrease the interstudy variability, as any real-world studies aim to cover the heterogeneity of study populations. If high-quality observational studies will become the standard in outcomes research in the next decades, these changes will eventually reduce the interstudy variability and also reduce the problem of structural heterogeneity of meta-analysis. Structural heterogeneity of a meta-analysis can be demonstrated by counting the number of included comparisons (eg, types of treatments), outcomes (eg, 3 or 5 year survival), and subtitles (eg, target populations and drugs).[35]

An additional point of discussion is related to the use of efficacy data (generated under ideal study conditions) for calculation of efficiency (ie, cost effectiveness) because the results of cost–benefit ratios should be applicable to real-world but not ideal study conditions. In contrast, efficacy data should be sufficient to get temporary approval for a new drug. This temporary approval should be used to complete PCTs under real-world conditions to generate the data that describe the "added patient value".

In summary, it can be predicted that the differential view on efficacy and effectiveness will affect research and political decisions such as the policies for approval, the demonstration of added patient value, the pricing regulations, and the public financing of health care.

## Conclusion

The RCT will remain the basic study design to confirm the efficacy of a new intervention under ideal conditions. It is recommended to select patients for this RCT who are likely to benefit from the investigated intervention. However, the RCT can never confirm effectiveness under real-world conditions.

To demonstrate patient benefit it is essential, first, to demonstrate efficacy by using an RCT and second, to demonstrate effectiveness by using a PCT. This second part of the innovative process is not yet established.

In our opinion, it is an important challenge to the scientific community to define the necessary steps and to recommend the appropriate methods for demonstrating the patient-related benefit.

## Acknowledgments

## Disclosure

## References

1. Loudon K, Zwarenstein M, Sullivan F, Donnan P, Treweek S. Making clinical trials more relevant: improving and validating the PRECIS tool for matching trial design decisions to trial purpose. *Trials*. 2013; 14:115.
2. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *CMAJ*. 2009;180:E47–E57.
3. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol*. 2009;62:464–475.
4. Fasola G, Macerelli M, Follador A, Rihawi K, Aprile G, Della Mea V. Health information technology in oncology practice: a literature review. *Cancer Inform*. 2014;13:131–139.
5. Spitzberg BH. (Re)Introducing communication competence to the health professions. *J Public Health Res*. 2013;2:e23.
6. Blair L, Légaré F. Is shared decision making a utopian dream or an achievable goal? *Patient*. Epub 2015 Feb 14.
7. Howard L, de Salis I, Tomlin Z, Thornicroft G, Donovan J. Why is recruitment to trials difficult? An investigation into recruitment difficulties in an RCT of supported employment in patients with severe mental illness. *Contemp Clin Trials*. 2009;30:40–46.
8. Dainesi SM, Goldbaum M. Reasons behind the participation in biomedical research: a brief review. *Rev Bras Epidemiol*. 2014;17: 842–851.
9. Gillies K, Elwyn G, Cook J. Making a decision about trial participation: the feasibility of measuring deliberation during the informed consent process for clinical trials. *Trials*. 2014;15:307. doi: 10.1186/1745-6215-15-307.
10. Keusch F, Rao R, Chang L, Lepkowski J, Reddy P, Choi SW. Participation in clinical research: perspectives of adult patients and parents of pediatric patients undergoing hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2014;20:1604–1611. doi: 10.1016/j.bbmt.2014.06.020.
11. Umutyan A, Chiechi C, Beckett LA, et al. Overcoming barriers to cancer clinical trial accrual: impact of a mass media campaign. *Cancer*. 2008;112:212–219.
12. Metge CJ. What comes after producing the evidence? The importance of external validity to translating science to practice. *Clin Ther*. 2011; 33(5):578–580.
13. Windeler J. Externe Validität [External validity]. *Z Evid Fortbild Qual Gesundhwes*. 2008;102:253–259. German.
14. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ*. 2001;323:42–46.
15. Ioannidis JP. How to make more published research true. *PLoS Med*. 2014;11:e1001747. doi: 10.1371/journal.pmed.1001747.
16. Roland M, Torgersen DJ. Understanding controlled trials. What are pragmatic trials? *Br Med J*. 1998;316:285.
17. Bauer MS, Williford WO, Dawson EE, et al. Principles of effectiveness trials and their implementation in VA Cooperative Study #430: "reducing the efficacy-effectiveness gap in bipolar disorder." *J Affect Disord*. 2001;67:61–78.
18. Delva W, Temmerman M. The efficacy-effectiveness gap in PMTCT. *S Afr Med J*. 2004;94:796.
19. Eichler HG, Abadie E, Breckenridge A, et al. Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat Rev Drug Discov*. 2011;10:495–506. doi: 10.1038/nrd3501.
20. MacDonald KM, Vancayzeele S, Deblander A, Abraham IL. Longitudinal observational studies to study the efficacy-effectiveness gap in drug therapy: application to mild and moderate dementia. *Nurs Clin North Am*. 2006;41:105–117, vii.
21. Maher L. Hepatitis B vaccination and injecting drug use: narrowing the efficacy-effectiveness gap. *Int J Drug Policy*. 2008;19:425–428. doi: 10.1016/j.drugpo.2007.12.010.
22. Scott J. Using Health Belief Models to understand the efficacy-effectiveness gap for mood stabilizer treatments. *Neuropsychobiology*. 2002;46 (Suppl 1):13–15.
23. Tandon R. Bridging the efficacy-effectiveness gap in the antipsychotic treatment of schizophrenia: back to the basics. *J Clin Psychiatry*. 2014;75:e1321–e1322. doi: 10.4088/JCP.14com09595.
24. Weiss AP, Guidi J, Fava M. Closing the efficacy-effectiveness gap: translating both the what and the how from randomized controlled trials to clinical practice. *J Clin Psychiatry*. 2009;70: 446–449.

25. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bull World Health Organ*. 2007;85:867–872.

26. Miller R, Hollist CS. Attrition bias (2007). Faculty publications paper 45. Available from: http://digitalcommons.unl.edu/famconfacpub/45. Accessed September 17, 2015.

27. Porzsolt F, Eisemann M, Habs M. Complementary alternative medicine and traditional scientific medicine should use identical rules to complete clinical trials. *EUJIM*. 2010;2:3–7. doi 10.1016/j.eujim.2010.02.001.

28. Porzsolt F, Eisemann M, Habs M, Wyer P. Form follows function: pragmatic controlled trials (PCTs) have to answer different questions and require different designs than randomized controlled trials (RCTs). *J Publ Health*. 2013;21:307–313. doi: 10.1007/s10389-012-0544-5.

29. Gaus W, Muche R. Is a controlled randomized trial the non-plus-ultra design? A contribution to discussion on comparative, controlled, non-randomized trials. *Contemp Clin Trials*. 2013;35:127–132.

30. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions [editorial]. *ACP J Club*. 1995;123:A12–A13.

31. Nobre MR, Bernardo WM, Jatene FB. A prática clínica baseada em evidências. Parte I – questões clínicas bem construídas [Evidence based clinical practice. Part 1 – well structured clinical questions]. *Rev Assoc Med Bras*. 2003;49:445–449. Portuguese.

32. Austin PC. An introduction to propensity score methods for reducing the effects of confounding. *Multivariate Behav Res*. 2011;46:399–424.

33. Waber RL, Shiv B, Carmon Z, Ariely D. Commercial features of placebo and therapeutic efficacy. *JAMA*. 2008;299:1016–1017.

34. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124.

35. Brögger C. Validität von systematischen reviews der Cochjrane Collaboration [Validity of systematic reviews of the Cochrane Collaboration]. Doctoral Thesis. Ulm, Germany: Medical Faculty, University of Ulm; 2015.