Clinics in Respiratory Medicine

Review Article
Published: 12 Dec, 2020

Key Considerations in the Conduct of Randomized Controlled Trials

Maeve P[1], Reena M[1] and Aanchal K[2]*

[1]Department of Sleep Disorders, Cleveland Clinic Lerner College of Medicine, USA

[2]Department of Critical Care Medicine, Respiratory Institute, Cleveland Clinic, USA

Abstract

In the realm of clinical research, the Randomized Controlled Trial (RCT) study design-which is most often characterized by random assignment of a participant to a treatment or control arm-has long been considered the gold standard to assess the impact of an intervention. The current review is structured to serve as a resource for the clinician to understand the salient aspects of the conduct and interpretation of clinical trials. Specifically, the following will be reviewed: 1) including an a priori hypotheses with clearly defined primary and as indicated secondary endpoints, 2) utilizing a design in accordance with standards Consolidated Standards of Reporting Trials (CONSORT) inclusive of a statistical approach, 3) considering ethical aspects such as clinical equipoise, 4) implementing standardized reporting of results with publication of both statistically significant and non-significant findings, 5) objective oversight of trial conduct and tracking of adverse events, 6) transparency of the reporting of conflict of interest and funding source and 7) registry of clinical trial approach and results during study initiation and conduct.

Key Points

- The Randomized Controlled Trial (RCT) study design has been used for centuries and remains the gold standard to assess the impact of an intervention.
- RCTs can assess efficacy and effectiveness. Efficacy trials focus on the biologic impacts of an intervention, where effectiveness trials focus on real-world feasibility of the intervention.
- RCTs are categorized into four phases, where each phase indicates a progression in the size of the trials, understanding of the treatment intervention, and proximity to clinical use of the intervention.
- Trials must only be conducted for interventions that exist in a state of clinical equipoise. Institutional Review Boards (IRBs) and Data Safety and Monitoring Boards (DSMBs) ensure that trials are conducted efficiently and ethically.
- RCTs are characterized by the randomization of participants to treatment or control arms, but the designs for randomization and allocation vary greatly.
- The reporting of RCTs should be conducted in accordance with the Consolidated Standards of Reporting Trials (CONSORT) and should be maximally transparent.
- Statistical and analytic decisions should be made in order to minimize biases and maximize the power and generalizability of the trial's results.

Introduction

In clinical research, the Randomized Controlled Trial (RCT) study design, characterized by random assignment of participants to treatment or control arms, has long been considered the gold standard to assess the impact of an intervention. Under ideal circumstances, RCT results are incredibly informative, valuable, and form the basis to support causality and inform clinical decision-making. That said, there are many threats to the validity of the RCT design and, accordingly, many aspects of the design which require careful consideration. The current review aims to aid the clinician's understanding of salient aspects of the conduct and interpretation of clinical trials.

Background

Although the first RCT of note was a 1948 trial of streptomycin to treat tuberculosis in the United

OPEN ACCESS

Correspondence:
Reena Mehra, Department of Sleep Disorders, Cleveland Clinic Lerner College of Medicine, Cleveland, 9500 Euclid Avenue Cleveland Clinic, Cleveland, Ohio 44195, USA, E-mail: mehrar@ccf.org

Received Date: *20 Nov 2020*
Accepted Date: *09 Dec 2020*
Published Date: *12 Dec 2020*

Citation:
Maeve P, Reena M, Aanchal K. Key Considerations in the Conduct of Randomized Controlled Trials. Clin Respirat Med. 2020; 2(2): 1016.

Kingdom, the conceptualization of the RCT had been discussed in forums long before this time. In the 1750s, James Lind, a Scottish surgeon, set the stage for comparative trial methods through study that examined the effectiveness of different treatments for scurvy [1]. Although he did not conduct a RCT in the truest sense and the study involved only 12 sailors as participants, multiple arms of the study were used to test remedies such as different herbs and citrus fruit. The basic components of the clinical trial, as characterized by multiple study arms with specific interventions, were then improved upon in the 1940's when Sir Austin Bradford Hill attempted to mitigate bias stemming from the awareness and predictability of the alternate allocation approach with concealed randomization [1]. This then allowed clinical trials to shield both the researcher and the participant to the specific intervention used and allowed for the ability to preserve the integrity of the trial.

Planning Randomized Controlled Trials

Efficacy and effectiveness

Efficacy and effectiveness are distinct in that efficacy refers to ability of the intervention to work in an ideal setting, whereas effectiveness refers to how the treatment works in the "real world," i.e. outside of a controlled setting. Efficacy trials investigate the biologic impact of the intervention. These trials are characterized by high levels of adherence with the intervention and thereby have the ability to draw more direct conclusions about the true impact of the intervention under the most ideal of circumstances. Effectiveness investigations are oftentimes pragmatic and look beyond efficacy to evaluate the feasibility of the intervention. These trials have a higher likelihood of reduced/non-adherence to the intervention and include other real-world factors, such as barriers to care. Ideally, treatments must both be biologically efficacious as well as effective to be viably translatable to the clinical setting.

Phases

Clinical significance depends not only on the type or findings of the trial, but also the trial phase. RCTs are categorized into four phases, where each phase indicates a progression in the understanding of the treatment intervention. Phase I trials explore possible toxic effects and determine the tolerance of the intervention, as well as an appropriate. These trials are smaller (about 10 people to 30 people) and focused on safety. Phase II trials determine if the treatment has a therapeutic effect or if the benefits may outweigh the risks, with a slightly larger sample size (20 to 50 people) and a focus on treatment impact on the outcome of interest. Phase III trials are larger still, utilizing a sample size of 100 to 1000 people, and compare the new treatment to the standard therapy for the condition of interest. If there is no standard therapy, a control or placebo may be used. Finally, Phase IV trials generally use many different centers and hundreds to thousands of participants to obtain long-term, large-scale information about morbidity and late-effects of the treatment. Generally, these studies take place after the therapy has been approved for marketing and the manufacturer requires more long-term safety information.

Ethics

Clinical equipoise

Aside from randomization, perhaps the most important concept when designing an RCT is the standard principle of clinical equipoise. A term coined by Benjamin Freedman in 1987, "clinical equipoise" refers to the intention of a meaningful ethical appraisal of the involvement of human subjects in clinical trials on a case-by-case basis [2]. In terms of RCT design, clinical equipoise exists when there is genuine uncertainty within the professional or scientific community as to which of the treatment arms is superior due to insufficient rationale for a choice between options or approaches of care. At an individual level, personal equipoise exists when the provider has no intervention preference and is uncertain about the harm or benefit conferred by the intervention, allowing for ethical enrollment of patients in the clinical trial.

Trial oversight

Throughout the design and conduct of the RCT, an assigned and blinded Data Safety and Monitoring Board (DSMB) reviews the ongoing conduct of the trial to ensure continuing patient safety, validity, and merit of the trial. In 1998, all NIH funded trials were required to have a Data and Safety Monitoring plan, and the FDA also has a recent guidance document that will most likely result in more industry sponsored trials with a DSMB [3].

DSMBs are often very similar to Institutional Review Boards (IRBs), but IRBs perform their duties from an internal standpoint, while DSMBs are comprised of external personnel, often well acquainted with the subject matter of the trial. Additionally, IRBs conduct continuing review, but their focus is primarily on document and process review, whereas DSMBs continuously evaluate trial feasibility, objectives, logistics, and data. Both boards are committed to patient safety and evaluate adverse events. Adverse events are undesirable health outcomes that occur during the trial and may or may not have a causal relationship to the trial treatment, and Serious Adverse Events (SAE) are life-threatening, require hospitalization, and/or create significant disability. All adverse events should always be reported and depending on the severity and frequency of adverse events, investigators and data safety monitors may decide to terminate the trial prematurely.

Design

Rationale for randomization

The purpose and advantage of randomization lies in the balance across arms. Through randomization, various confounding influences, such as concomitant medications, comorbidities, and other key characteristics, will ideally be equally distributed across the potential allocation arms. This principle applies not only to measured variables, but unmeasured variables, such as environmental exposures, occupations, socioeconomic status, and other, more intangible, potential confounding factors as well. In order to verify balance, the first table for any manuscript of an RCT will show the characteristics of the participants in each randomization group.

The balance of characteristics between groups achieved through randomization serves to minimize or eliminate bias in the treatment assignment, thereby reducing selection bias even at the point of study entry. Randomization also facilitates the blinding of the treatment from the investigators, participants, and study team, reducing biases to the extent it is applied to personnel involved with these roles within the study. The reduction of bias is particularly key in studies with subjective outcome measures, and so randomization is an especially important tool for enhancing the likelihood that any difference that occurs in the outcome of interest occurs solely by chance.

Consolidated standards of reporting trials (CONSORT) standards

Although many RCTs are conducted in accordance with the sound methods described in their initial protocols, trial outcome reporting

is often incomplete, biased, and inconsistent with protocols, which makes their corresponding articles unreliable [4]. The Consolidated Standards of Reporting Trials (CONSORT) 2010 explanation and elaboration document gives evidence of this, citing that only 34% of 616 trial reports indexed in PubMed in 2006 reported patient group assignment methods, only 53% defined a primary end point, and only 45% reported a sample size calculation [5]. Inaccurate reporting is an issue as well. Of 119 reports stating that all participants were included in the analysis in the groups to which they were originally assigned (intention-to-treat analysis), 15 (13%) excluded patients or did not analyze all patients as allocated [5]. Because of the lack of detail and overall underreporting in RCTs, when reviewing and critiquing an article, one can use the CONSORT statement to ensure that all of the key components are present. The CONSORT checklist provides guidelines to standardize study reporting and serves as a useful tool to critically evaluate published articles and ensure better clinical trial publications.

In order to report an investigation completely, investigators must document the flow of participants through each stage of the trial. Documentation of study flow should show the people evaluated for potential enrollment, people who did not meet the inclusion criteria, the people who met exclusion criteria, those who were approached for participation but declined, and then finally those who enrolled. This information helps with the external assessment selection bias and whether or not the trial participants are likely to be a representative sample of the population. Additionally, the investigator must also note if any participants were lost to follow up, seeing as the integrity of the trial could be affected and bias could be introduced due to imbalance between study arms. The rest of the documented study flow should detail the randomization, treatment allocation, follow-up, and analyses of the participants (Figure 1). These steps provide rationale for the trial size, internal validity, and intention to treat. At each step, any deviations from the protocol should be explained, as they can impact the results, conclusions, and overall validity of the study. No investigation is perfect, but the integrity and validity of the investigation can be preserved through accurate reporting.

Typical designs

There are three broad trial types: Superiority, equivalence, and non-inferiority. Superiority trials help determine if a new treatment is different from (or better than) a placebo or existing treatment (active control). The null hypothesis in this situation is that there is no difference between the treatments, and the alternative hypothesis is that the new treatment is different from (two tailed) or better than (one-sided) the control. In an equivalence trial, the null and alternative hypotheses are reversed, so that the null hypothesis assumes some degree of difference between the control and the treatment and the alternative hypothesis is that the treatment is not significantly different than the control. Equivalence trials are generally conducted when an effective treatment already exists, i.e. a situation in which it would be unethical to conduct a placebo-controlled trial, if the treatment is for a life-threatening illness, or the new treatment is not suspected to be substantially better than the existing treatment. The least common trial type is the non-inferiority trial, where investigators want to evaluate whether or not the intervention is worse than an existing intervention.

In addition to general trial types, there are also various specific trial designs, namely parallel, crossover, and factorial designs. In parallel trials, participants are randomized to different study arms and move through the trial in parallel, with the only difference being that one group receives the treatment and the other receives a placebo (Figure 2). Crossover trials involve setting up a parallel-type trial at the outset, measuring outcomes part of the way through, then having the participants go through a washout period and switch interventions

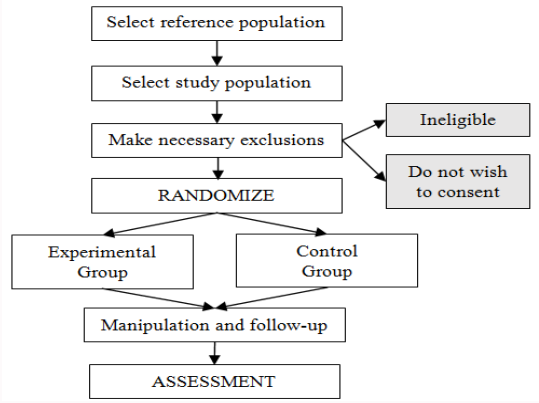

Figure 1: This figure shows the overall study design of a randomized controlled trial. In order to conduct a complete trial, investigators must first identify a population of individuals to investigate, and then select an appropriate study sample, making exclusions as necessary. The remaining individuals must then be randomized to treatment or control study arms, where they will proceed with the interventions and provide data for investigators to analyze and from which to draw conclusions.

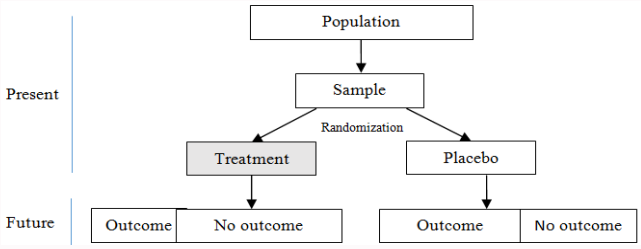

Figure 2: The parallel design is a common and simple randomized controlled trial design. This design leads participants through the study in parallel after randomization to separate study arms.

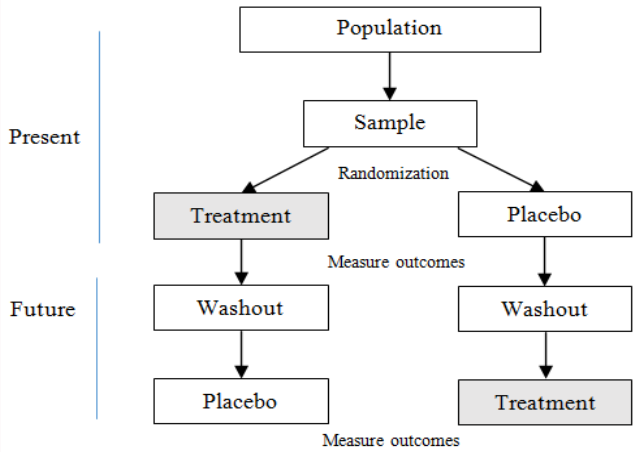

Figure 3: The crossover design for randomized controlled trials involves setting up a parallel-type trial at the outset. Study outcomes are measured part of the way through the trial, and then participants undergo a "washout" period, where they are titrated off of their respective interventions, and then "crossover" to the intervention initially used by the participants from the other study arm.

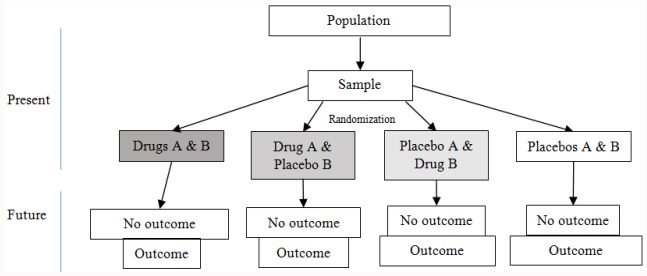

Figure 4: The factorial design for randomized controlled trials utilizes various combinations of interventions to examine an outcome of interest. Each study arm utilizes more than one intervention to assess the relative efficacy of different interventions and intervention combinations.

(Figure 3). The washout period is absolutely necessary in this case, as it allows the investigators to minimize any lingering impact the treatment may have, as well as evaluate carryover effects, if any. The crossover design enhances efficiency in terms of number needed to be randomized, allows for enhanced power as each participant serves as their own control and minimizes concern of confounding by static biologic factors. However, the longer duration of the trial increases participant burden and likelihood of dropouts. The last major trial design is the factorial design, which utilizes various combinations and permutations of different interventions to examine the outcome of interest (Figure 4).

Allocation and Randomization Methods

The quasi-experimental design is an allocation method based on a certain factor, such as a patient's rank in an alphabetized list of names. This design is used when it is not logically feasible or ethical to conduct true randomization, such as with randomization of a patient with limited availability or testing a previously approved therapy. The most common form of allocation is individual allocation, in which each individual participant receives a random study arm assignment. The ratio of allocation may vary, e.g. the likelihood of randomization to any given arm may be higher or lower than another arm, but this ratio is predetermined and would not affect the randomization blinding.

The most common form of randomization is simple randomization, which is based on a simple ratio, such as a 1:1 experimental-to-control group assignment. Although this strategy lends itself to a well-balanced trial in the end, it may not facilitate evenly balanced groups throughout the study. Stratified randomization and restricted randomization can help with this issue. Stratified randomization ensures balance between study arms and confounders along the course of small trials. Restricted randomization also maintains balance throughout the course of the trial by using a blocked randomization pattern, where individual participants are assigned to "blocks" of participants and then randomized to different study arms in a given ratio within those blocks. In order to avoid predictability of assignment, the block sizes are random and often varied throughout the study. Of all of the randomization possibilities, minimization is the least common. This method minimizes the imbalance of selected patient factors between groups, but due to those criteria, participant assignments are predictable, which defeats the purpose and advantage of randomization.

Novel trial designs

Cluster randomization, one of the less common trial designs, is rising in usage. This method allows for randomization of intact social units or clusters of individuals rather than independent individuals. This type of randomization is advantageous when working with different cultures or institutions that have operational differences. Because of different cultural or administrative practices, clustering these individuals into one representative group may avoid treatment group contamination and improve investigator and participant cooperation. For instance, if a participant already receives care at a particular study site, clustering based on a particular institution may improve patient understanding through their relationship with the institution's providers. However, inter-cluster correlations can pose significant statistical analysis challenges, and thus this strategy is not as common as individual randomization.

Pragmatic trials, using Routinely Collected Health Data (RCD), are also rising in usage. With the integration of electronic medical records into standards of care, RCD serve as a rich avenue for research, as they can reduce the costs of time, money, and other resources while providing opportunities for point-of-care/routine-care trial designs. However, using RCD for trials may incur initial infrastructure set-up costs, as well as introduce detection bias and unique ethical implications. As such, future use of RCD in trials should focus on data quality validation, the impact of alternative research designs on data analysis, and transparent reporting [6].

Concealment of allocation and blinding

After an allocation sequence has been generated, the allocation concealment mechanism must be developed to maintain the blinding of the investigator and participant throughout the study. This generally entails numerical assignments to study materials that will be distributed to different study arms, such as numbered bottles or envelopes. Concealment is distinct from blinding in that concealment of allocation is implemented before the patient enters the trial, and blinding serves to mask the treatments after randomization. However, infringements on either can result in biased trials. Failed concealment from either the patient or investigator results in selection bias and renders treatment assignments non-random.

Blinding, although not always feasible, prevents response bias from patients as a result of the knowledge of their treatment and outcome assessment bias on the part of the investigator. Blinding can help the investigator understand the true effects of the treatment by using a placebo group, a group which has an inactive treatment, as comparison to evaluate the difference between patient-expected effects (those experienced by the placebo group) and actual treatment effects (those experienced by the treatment group). Open trials have no blinding, and as such are the most subject to bias. Single blind trials only blind either the participant or the investigator, and double blind trials blind both. Triple blind trials exist as well, drawing in those who assess outcomes as the third, blinded party; their blinding may be the most important of all when it comes to mitigating bias.

Analytic Considerations

Statistical analysis

To understand if a study will have sufficient participants for the potential to allow for statistically significant conclusions; investigators must first calculate the sample size and then screen an appropriate amount of candidates to achieve that sample size. Often based upon prior literature, the sample size calculation makes a reasonable estimate as to how many people the study must enroll in order to have sufficient power to examine a certain effect size or difference

between groups. Most studies make this calculation assuming that not all who are randomized will complete the trial, providing a safety net for when participants drop out or are lost to follow-up. Loss to follow-up of less than 5% leads to little bias, while loss to follow-up of 20% or more compromises the trial's validity significantly [7]. The effect size (d) is calculated as the "Cohen's D," or the standardized mean difference between two groups. A small effect size, where the value of "d" is around 0.20, will require a large sample size to be able to detect that small difference on a significant level. Conversely, if the "d" is greater than 0.80, the study will not need as many participants to show the significant difference between groups.

Even though a research finding may be statistically significant, investigators must also assess whether or not a finding is clinically important. The "Number Needed to Treat" (NNT) is a useful concept for interpreting trial findings and determining the clinical application when testing an intervention that addresses a certain health issue. The NNT represents the number of patients who need to receive the intervention to produce one good outcome as compared to control and is widely used as an index of clinical significance.

Analyses and biases

When it comes to analyzing RCTs, there are two main approaches: Intent-to-treat *vs.* per protocol analysis (also called on-treatment analysis). Intent-to-treat analysis is generally preferred and is based on the patient's randomization. In order to fully preserve the benefit of randomization, researchers should include all randomized participants in the analysis, retained to the group to which they were allocated. However, strict intent-to-treat analysis is oftentimes difficult to achieve due to missing outcomes and protocol deviations. This raises the consideration of per protocol, or on-treatment, analyses. On-treatment analyses are performed according to the actual treatment received by the participant. This more accurately reflects trial conditions and conduct, but can be affected by physician and/or patient preferences, reducing the power of randomization. Secondary outcomes can also provide additional complications. When conducting analyses, all outcome variables should be specified and corrections for multiple testing should be implemented to mitigate type-1 statistical error. Other analysis complications, like poor compliance to the intervention and concomitance with other medications and procedures, can lead researchers to conduct interim analyses to get a more complete picture of the trial.

Early stopping

Interim analyses can reveal reasons for trial extension, but various factors can necessitate early trial stopping. RCTs can end naturally when they reach their sample size goal, length of follow-up goal, or scheduled date of closure. In the latter situation, the trial will stop in a manner independent of its results and the stopping is unlikely to introduce bias in the results. However, a trial may stop earlier than planned due to results of an interim analysis showing larger than expected benefit or harm in the experimental intervention. Trials may stop early if no longer viable due to lack of accessibility to patients or study interventions, decreased funding, or other studies that render the research question irrelevant. There are multiple options for interim analyses, including the Haybittle-Peto, where all interim analysis have a high-significance level and a final significance level of 0.05, and the Pocock, in which all interim and final analyses have a high level of significance. However, the most common is the O'Brien-Fleming analysis, which decreases the significance level with each interim analysis with a goal of having the final test close to a 0.05 level of significance.

Transparency

Transparency is paramount, and even if a study is not an RCT, it is now NIH-mandated that clinical trials be registered and the data included at clinicaltrials.gov. This website helps maintain transparency, as data must be reported periodically and the results are accessible to anyone as it resides in the public domain. Both journal editors and those from the general public have access to RCT results through clinicaltrials.gov, and journals in particular will assess when investigators update their study so as to ensure that study conditions did not change after the study commencement. This information also includes any conflict of interest disclosures from the investigators and study team.

Summary and Conclusion

Given its ability to generate data that provide bases to support causality, the RCT design remains the gold standard for intervention assessment. The background, planning, ethics, design, methods, analytic considerations, and reporting transparency are all key considerations of an RCT, and so this review has been structured to serve as a cursory overview of these components. With these considerations in mind, clinicians may understand the salient aspects of the conduct and interpretation of clinical trials, which can have implications for their understanding of the current state of research, and therefore, for their clinical practice.

Disclosures

Dr. Mehra reports that she has received NIH funding for which she has served as Principal Investigator (NHLBI R01HL109493, R21HL108226). Her institution has received positive airway pressure machines and equipment from Philips Respironics, Resmed and Natus for use in NIH-funded research. She has received honorarium from the American Academy of Sleep Medicine for speaking.

References

1. Bothwell LE, Podolsky SH. The emergence of the randomized, controlled trial. N Engl J Med. 2016;375(6):501-4.
2. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med. 1987;371(3):141-5.
3. United States Department of Health and Human Services. E6 (R2) good clinical practice: integrated addendum to ICH E6(R1). Guidance for Industry Procedural Contains Nonbinding Recommendations. 2018:6.
4. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. JAMA. 2004;291(20):2457-65.
5. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.
6. Mc Cord KA, Al-Shahi Salman R, Treweek S, Heidi G, Daniel S, Whiteley W, et al. Routinely collected data for randomized trials: Promises, barriers, and implications. Trials. 2018;19(1):29.
7. Sackett DL, Richardson WS, Rosenberg W, Haynes RB, Henderson AR. Evidence-based medicine--how to practice and teach EBM. Clin Chem. 1997;43(10):2014.