Phase II Clinical Research Design in Cardiology

Learning the Right Lessons Too Well: Observations and Recommendations From the Cardiovascular Cell Therapy Research Network (CCTRN)

Joshua M. Hare, MD; Roberto Bolli, MD; John P. Cooke, MD, PhD; David J. Gordon, MD, PhD; Timothy D. Henry, MD; Emerson C. Perin, MD, PhD; Keith L. March, MD, PhD; Michael P. Murphy, MD; Carl J. Pepine, MD; Robert D. Simari, MD; Sonia I. Skarlatos, PhD; Jay H. Traverse, MD; James T. Willerson, MD; Anita D. Szady, MD; Doris A. Taylor, PhD; Rachel W. Vojvodic, MPH; Phillip C. Yang, MD; Lemuel A. Moyé, MD, PhD; for the Cardiovascular Cell Therapy Research Network (CCTRN)

Clinical trials demonstrate the best of medical expertise and epidemiological elegance. From the simple building blocks of contemporaneous control groups, randomization, and blinding, they assemble a clear picture of the nature of the treatment-effect relationship. This accomplishment has earned them the star ascendant position in cardiovascular research. Their advantage was demonstrated with Bradford Hill’s work on streptomycin, and as knowledge of the pathogenesis of atherosclerotic disease produced possibilities for new treatments, cardiovascular researchers applied this new research tool to identify effective therapies in a sequential approach (Figure 1).

This work accelerated with clinical trials demonstrating treatment benefits for chronic diseases such as hypertension, lipid abnormalities, and heart failure, cementing their role in identifying new therapies to prevent the sequela of cardiovascular disease in vulnerable populations.

However, the limitations of early clinical trial interpretation also appeared. The findings of the Multiple Risk Factor Intervention Trial (MRIFT), the International Verapamil SR/Trandolapril Study (INVEST), the Early Versus Late Intervention Trial With Estradiol (ELITE), the Cardiac Arrhythmia Suppression Trial (CAST), and the US Carvedilol program controversy together served to undermine the confidence of cardiologists in the interpretation of clinical trial results. Thought leaders in the field identified the interpretation problem (ie, the hyperreliance on P values to reflect positive results for any analysis in a clinical trial) and called for a fundamental change in the design and analysis of clinical trials.

Subsequent work produced the following clinical trials principles: (1) There should be clear and prospective declaration in the written protocol about all important aspects of the clinical trial with particular emphasis on the end points of the study; (2) end-point assessments must be planned with clear and declared assessments of type I error penalties; and (3) the rules of the protocol must be followed in the trial (“First say what you will do, then do what you said”).

The widespread acceptance of the 0.05 type I error level, in concert with the multiple testing issue, generated clinical trials with only a small number of primary end points or confirmatory analyses, buttressed by a larger number of prospectively declared secondary or supportive evaluations. The remaining post hoc analyses, data dredging, and subgroup analyses were now unpersuasive, giving way to the prospectively declared primary analyses with their attendant spending functions and multiplicity corrections. (Figure 2).

Over time, these principles were absorbed by the cardiovascular community. Contemporaneous Protocol Review committees, Data Safety and Monitoring boards, the federal Food and Drug Administration, top-tier journals, and knowledgeable audiences of international cardiology meetings now expect these conditions to be met.

Phase II Investigations Living in a Phase III World

Executing research in accordance with these confining principles is a daunting, expensive, and time-consuming task, yet the investigators, epidemiologists, and biostatisticians who conduct phase III studies are well equipped for the mission. A wealth of preliminary data in animals and in humans is commonly available to serve as the foundation for the selection of a small number of community-accepted primary end points. This leads to the identification of appropriate end points and their effect sizes that, if is adequate availability of finances and subjects, can produce an executable phase III study. This is clinical trial design at its finest.
It is only natural for the designers of phase II clinical trials to embrace these same research standards. However, phase II clinical trials themselves have a much weaker foundation. Traditional end points may not have been implemented in the early pilot studies or case series on which the phase II study itself relies. There may be insufficient evidence for the selection of effect sizes necessary for the computation of a sample size. Traditionally accepted biomarkers may not adequately detect the underlying biological effect of a novel therapeutic.

The human ability to successfully embed the conclusions from preclinical studies into the design of phase I/II studies is both intricate and rare, its absence aggravated by the press of time. An uncertain funding future, along with the finite patent period, drives the perceived need for speed as investigators hasten from phase I/II to phase III studies. Yet it is time itself that is required to ensure that the best population, the best therapy dose, and the best statistical estimates are obtained. The lack of human and financial capital, amplified by the lack of time, is a combination that injects structural weakness into the research enterprise.

This may not be the case in all areas of early clinical research (eg, low-density lipoprotein cholesterol reduction or the development of new cephalosporins) that continue to benefit from well-developed, time-tested in vitro and in vivo models that serve the phase I/II community well.

However, novel therapies or targets subject to the standard clinical trial metric (ie, driving a single primary end point to a 0.05 statistically significant level) suffer when that novel therapy has no natural biomarker, possesses mechanisms of benefit that might differ between animals and humans, and is researched by consortiums with limited resources that cannot afford sequential studies, each with a single primary end point.

Nowhere are these considerations more crucial than in the burgeoning study of cell therapy, where enormous enthusiasm overlays a young and immature area in which preliminary end-point findings, while promising, are based on a relatively small number of subjects. Thus, the issue of end-point selection, a problem that has bedeviled large clinical trials, is even more challenging for this nascent field.

The situation is complicated by a change in the standard paradigm of clinical trial progression in the evaluation of a new therapy. In the traditional paradigm, first-in-human studies are historically conducted in normal volunteers in whom dose escalation can take place in well-monitored circumstances, with testing terminated if there is a sign of harmful effects. However, as well demonstrated in oncology, medications that are believed to hold out efficacy can be expected to produce considerable side effects. Because it is unethical to subject normal individuals with no disease to this anticipated level of risk without benefit to the normal population, these first-in-human studies can be conducted only in the target population. Such individuals with ongoing, commonly progressive disease can be willing, under the right safeguards and oversight, to assume the burden of risk to have an opportunity to receive benefit, yet there are no human data on which to base the measure of efficacy in these trials.

Because large first-in-human studies are oxymorons, investigators who both are limited to a small number of subjects for their phase II design and must labor under current phase III research expectations are compelled to choose from 2 unpalatable alternatives: select modest and reasonable effect sizes for which the small study is likely to be underpowered, or select large effect sizes that lead to achievable sample sizes, a decision that all but ensures that the effect the trial was designed to detect does not exist. Each of these options increases the likelihood that the phase II study will miss a statistically and clinically significant finding. However, the rationale for the phase II cell therapy clinical trial is to identify, perhaps for the first time, the potential benefits of the therapy being assessed. This truism has been recognized by the pharmaceutical industry, in which phase II clinical trial results are scanned for signals on which phase III studies are based with little regard for P values. In the development of a novel therapeutic modality with uncertain biomarkers, the rigors and requirements of phase III methodologies are inappropriate.

It is as though we in academia cannot distinguish between the mission of a (phase III) army with its corps of human and material resources and the charge of a smaller (phase II) reconnaissance unit. They are both part of the same team but have different tasks. The role of the phase II study is to broadly survey the possible delivery mechanisms and possible benefits of the study product. This knowledge is then passed to the phase III trial investigators, who, with the target provided by the phase II investigators, can now direct the appropriate resources for a well-directed advance (Figure 3).

Phase II studies certainly have their limitations; their very size makes them an unreliable foundation for therapy guidelines. However, large studies must have a data-based foundation, and that foundation is built from small studies. The role of a phase II study is not to confirm benefit for a new therapy...
but instead to identify its first signal, leaving confirmation to the following larger studies (Figure 4).

Because the tasks of these 2 trial phases are different, should not their metric of success be different as well?

**Recommendations**

The time-tested guidelines for a successful phase III trial are sound and have served well; we do not advocate any alteration in many of these established metrics for phase II studies. Phase II studies should be based on prospectively declared, detailed, and well-written protocols. These protocols should continue to include close inspection by Protocol Review committees/Data Safety and Monitoring boards, the Food and Drug Administration, and local internal review boards. However, there is growing appreciation that innovative designs of phase II trials are required so that they can be true to their mission, that is, to test a range of efficacy domains in the early clinical testing of novel therapeutics to identify the range of potential beneficial outcomes.

Specifically, we recommend that the measure for success for a phase II clinical trial should be decoupled from the achievement of statistical significance for a small number of primary end points, a metric that is more in alignment with the goal of a phase III study, but to provide phase II studies with the ability to assess many primary end-point analyses without the overriding concern of limiting the likelihood of a false-positive finding. Because in this new domain false-negative findings can stifle future research, we propose the following metrics set as a first step that will help to maximize the utility of phase II studies.

1. Evidence describing the safety of the study product and its delivery coincident within the limitations of the study. Small studies can only begin to define the safety profile of an intervention, yet they are an essential first step. The investigators must present all significant adverse events to the community through publication or to the regulatory bodies that oversee its execution.

2. Many primary end points should be permitted; each should be prospectively declared and then its findings reported. The investigators who will use the findings of the phase II study as a basis for phase III design must have access to the findings of all end-point results as set out in the phase II study protocol. To assess the coherence of the findings, the phase II investigators should select end points from different categories of effects (domains). In cardiac cell therapy studies, useful domains would be the following:

   - Structural evaluations: Measures of left ventricular function, for example, ejection fraction, end-systolic and end-diastolic volumes, stroke volumes (and indexes), infarct size, and ventricular sphericity.
   - Cardiovascular physiological measurements: Measures of contractility, for example, pressure-volume loops, rate of rise of left ventricular pressure (peak +dP/dt), diastolic performance, and loading conditions.
   - Biomarkers: Atrial natriuretic protein, brain natriuretic protein, cardiac enzymes, microRNAs, creatinine, C-reactive protein, transcriptomic-based biomarkers.
   - Functional capacity: Maximal oxygen consumption, peak walking time, and 6-minute walk distance, which are key components of an individual’s ability to function.
   - Quality of life: Well-established measures in a given field, for example, decreased need for target vessel revascularization and recurrent myocardial infarction. In addition, the Short Form-36, and the Minnesota Living With Heart Failure questionnaires are well established in the field.

An example of this approach is the recently published Calcium Uprogulation by Percutaneous Administration of Gene Therapy in Cardiac Disease (CUPID) trial, in which the investigators prospectively selected end points using this domain approach, permitting the study to contribute estimates of effects of biomarkers, organ function, and organism well-being in a nonexploratory research environment. First Mononuclear Cells injected in the United States (FOCUS) study selected end points of both left ventricular function and global function to provide useful and surprising data on the effect of bone marrow mononuclear cells in the heart failure population. Transplantation in Myocardial Infarction Evaluation 2 to 3 weeks Following Acute Myocardial Infarction (LateTIME) and Transplantation in
Myocardial Infarction Evaluation (TIME) reported the first data on a new measure of left ventricular regional wall motion. Similarly, the Percutaneous Stem Cell Injection Delivery Effects on Neomyogenesis Pilot Study (POSEIDON) examined end points in 3 domains: quality of life, functional capacity, and left ventricular structure and function, providing important insights and guidance for future trials. Intelligently crafted end points from a spectrum of well-chosen domains will serve well as prospectively declared end points by revealing the potential effects of cell therapy from the biomarker level to that of the organism, providing data on which subsequent phase III trials may be designed.

3. End-point event rates or mean changes must be established with predetermined precision. Although the investigators of a phase II study may not be able to predict a priori how large or small the effect (eg, the change in left ventricular ejection fraction over time in the cell therapy group) will be, they can and should be expected to determine that effect size, in the end (whatever its size), with precision. The standard errors, anticipated mean changes, and event rates must be determined a priori, and a comparison of observed versus expected variability may be considered one of the best measures of the success of a phase II research effort. Because technology (eg, cardiac magnetic resonance) changes rapidly, investigators must have the flexibility to use the most recent procedures, even if they are introduced after the protocol is written, to provide the most accurate and precise estimates of the effect of therapy. To avoid control groups that are inordinately small, full consideration should be given to apportioning patients 1:1 active:placebo. The use of a dose-escalation design and the availability of the crossover option can enhance the likelihood of potential subject interest in enrollment.

4. Hypothesis testing requiring small P values should not be the primary goal of phase II studies, the goal of which is to provide direction, not decision. The assessment of safety and efficacy in the general population is the goal of phase III studies; hypothesis testing with its multiplicity correction provides a measure of the likelihood that the efficacy findings of these study are not due to sampling error and that the patient population, who will pay the financial cost and bear the side effect burden, is likely to experience efficacy. Positive phase II studies do not lead to approval and community use but instead produce subsequent studies by providing foundational evidence and generating hypotheses. One could argue that hypothesis testing in these underpowered environments serves little use. Conventionally, phase 2 trials are designed to elucidate the mechanism of action of the therapeutic, to explore the dose-response relationship using some quantitative measure of drug effect in vivo, and to begin to explore the factors that contribute to variability of drug response. One can, in these 3 cases, apply (distribution free) conventional statistical approaches to “small” samples sizes. However, interpretations must encompass all of the data. When grossly underpowered clinical end points are measured in phase II to infer safety or efficacy or when 2 biomarkers are measured but only 1 is believed (eg, cholesteryl ester transfer protein), the field can be misled. Phase II studies should have the option to carry out testing at nominal 0.05 levels or to carry out hypothesis testing at $\alpha$ levels $>0.05$, for example, 0.15 or 0.20, while simultaneously being released from the requirement of corrections for multiplicity. In addition, the Bayes perspective should be considered because its use of prior information and loss functions provides a different framework by which to assess the results of clinical trials.

Proper evaluation of these phase II studies requires the community to remind itself of the necessary restraint that must accompany interpretation of the promising results of these smaller studies. Phase II research with its small sample size can cast only a weak spotlight in its initial illumination of potential clinical benefit of cell therapy. Its results are both frequently promising and frequently reversed by subsequent larger studies with their tighter focus and more precise estimators. This is not a fault in the process but a required built-in check to ensure that only the safest products with the strongest assurance of efficacy move forward to be used in larger populations. In the end, interpretations of phase II clinical research are based on impressions as much as hard evidence. This is precisely why the question of the causal nature of the exposure-outcome relationship must be settled in a well-designed, well-executed confirmatory phase III clinical trial. Moderating our expectations of smaller studies requires that, on their conclusion, we call for additional work, not rapid regulatory approval.

We recommend that the criteria that the community should use to assess phase II studies be the following:

1. Strength of association: Is there greater benefit in the cell therapy group than in the control group?
2. Consistency and concordance: If there are other studies in the field, do the findings of this study align with those? A persuasive argument for causality is much more clearly built on a collection of studies involving different patients and different protocols, each of which identifies the same relationship between the intervention and disease. Do a majority of the end-point findings move in the same direction? This concordance or internal alignment of the findings of a single study can substantially ease the learning curve of the community. If consistency and concordance are not present, are there biological reasons for the differences that can readily account for the effects (cells treated differently before injection, dose, etc)?
3. Coherence: If there is a mechanistic component to the study, do its results line up with other outcomes in a comprehensible way? Is there any well-accepted scientific principle that would argue against the effect? A cell therapy that improves left ventricular volumes and improves functional outcome (eg, improves performance on the 6-minute walk) brings coherence to the results through its physiological link of left ventricular and organism performance.
4. Dose response: If there is dose-response component to the study, is there a gradient of responses that track naturally with the gradient in dose or duration of therapy?
5. Safety: Risk must be determined objectively (i.e., through the use of blinding, at least of safety outcome determinations) and with appropriate circumscription. Small studies cannot determine whether a therapy is safe, only that the incidence of events cannot exceed an upper bound identified by the size of the study. A phase II study with 100 patients and no deaths cannot conclude that the therapy does not produce deaths, only that the death rate is <1 per 100.

These criteria are based on those of Bradford Hill,31 the father of the same randomized, clinical trials on which the cardiology research community relies. However, perhaps the greatest legacy of Dr Hill’s work is to remind us that, regardless of the controversial nature of the research, the interpretation requires careful and independent thought. Just as justice is more than reading from a rule book, the correct interpretation of a clinical trial requires more than a mere calculation of \( P \) values assessing orthodox end points.

Acknowledgments

The CCTRN would like to thank Dr Milton Packer for participating in seminal discussions that led to the writing of this paper.

Source of Funding

Funding for the Cardiovascular Cell Therapy Research Network was provided by the National Heart, Lung, and Blood Institute under cooperative agreement U11 HL087318-06.

Disclosures

Dr Hare has grants funded by the National Institutes of Health to his institution and has ownership interest and has done consultant work for Vestor. Drs Bolli, Henry, March, Murphy, Pepine, Perin, Simari, Willerson, and Taylor all reported grants funded by the National Heart, Lung, and Blood Institute to their institutions. Dr Traverse has grants funded by the National Institutes of Health. The other authors report no conflicts.

References


**Key Words:** clinical trials | phase II | research design | stem cell | therapy
Phase II Clinical Research Design in Cardiology: Learning the Right Lessons Too Well: Observations and Recommendations From the Cardiovascular Cell Therapy Research Network (CCTRN)


Circulation. 2013;127:1630-1635
doi: 10.1161/CIRCULATIONAHA.112.000779

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/127/15/1630