Published in final edited form as:

Am Heart J. 2012 September ; 164(3): 277–284. doi:10.1016/j.ahj.2012.07.002.

Use of alternative methodologies for evaluation of composite end points in trials of therapies for critical limb ischemia

Sumeet Subherwal, MD, MBA^a, Kevin J. Anstrom, PhD^a, William S. Jones, MD^a, Michael G. Felker, MD^a, Sanjay Misra, MD^b, Michael S. Conte, MD^c, William R. Hiatt, MD^d, and Manesh R. Patel, MD^a

^aDuke Clinical Research Institute, Duke University, Durham, NC

^bMayo Clinic, Rochester, MN

^cUniversity of California San Francisco, San Francisco, CA

^dUniversity of Colorado School of Medicine and CPC Clinical Research, Aurora, CO

Critical limb ischemia (CLI) represents advanced atherosclerosis of the lower extremities; results in severe hemodynamic compromise at rest; and manifests with lower extremity ischemic rest pain, ulceration, or gangrene. Patients with CLI have strikingly high morbidity and mortality rates, with 1-year mortality rates ranging from roughly 10% to 25% and 1-year amputation rates ranging from 10% to 20%. 1–4 Given the high morbidity and mortality associated with this CLI, goals of therapy include (1) reduction in overall mortality, (2) limb preservation with wound healing (in patients with wounds), and (3) improvement in quality of life with greater mobility and less pain. Attempts at medical therapies for CLI have included use of intravenous prostanoids, strict control of diabetes, smoking cessation, and wound care. Increasingly, the use of biologic therapies to induce angiogenesis (i.e., protein therapy, gene therapy, and cell therapy) is being investigated for treatment of advanced peripheral artery disease. 2, 5–9 Furthermore, advances in technologies (i.e., drug-coated balloons, atherectomy devices, and improvements in stent technology) for revascularization are increasingly being examined to improve outcomes in CLI. 10–12 As these therapies evolve, a reappraisal of study designs and end points to evaluate newer biologics and device therapies will be needed. Ideally, study designs with end points that capture the major clinical goals of CLI therapy with sufficient sensitivity to detect meaningful differences should be sought.

Many trials use a time-to-event analysis to compare therapies with respect to their impact on composite end points (including fatal and non-fatal components). Trial designs have also incorporated both safety and efficacy end points into the composite end point in hopes to provide clinicians a more comprehensive understanding of the effect of the investigational therapy. In addition, single-arm trials using Objective Performance Goals (OPGs) as

Address correspondences to: Sumeet Subherwal, MD, MBA, Duke University Medical Center, PO Box 17969, 2400 Pratt St. Rm. 0311 Terrace Level, Durham, NC 27705.

Disclosures

Sumeet Subherwal: None.

Kevin J. Anstrom: None. W. Schuyler Jones: None. Michael G. Felker: None.

Sanjay Misra: Consulting fees from Flexstent and Arterioocyte (<\$5,000), and research grant support from NHLBI.

Michael S. Conte: Consulting fees from Aastrom-Advisory Board, Baxter-Advisory Board, and Humacyte-Advisory Board.

William R. Hiatt: Research grant support from the following sponsors involved in clinical trials in PAD: Aastrom, AstraZeneca, Dनावे, GSK, Medrad Possis, Pleuristem.

Manesh R. Patel: Research grants from NHLBI, AHRQ, AstraZeneca, Pleuristem, Johnson and Johnson, and serves as a consultant/advisory board member for Genzyme, Bayer Healthcare, OrthoMcNeil Jansen, Baxter, theheart.org.

therapeutic benchmarks are now accepted for select therapies (usually device therapies) in the absence of randomized, controlled trials of new therapies in CLI.^{13,14} However, clinicians should be aware of several major limitations and challenges in using the time-to-event analysis and OPGs. This article addresses limitations inherent in the interpretation of trial results using these methodologies. Meanwhile, an ideal evaluation of an investigational therapy should allow comparison of clinically important end points, capture clinically meaningful recurrent events (events that occur after the first event) and assign relative values to the end points of interest such that end points of more clinical significance contribute to a greater extent to the final statistical comparison between therapies than less meaningful events. Consequently, we introduce alternative methodologies that achieve these three goals allowing for a more comprehensive evaluation of the effect of a therapy. We focus our discussion on one of these methodologies referred to as Global Rank Method for end point analysis and provide an outline on how to perform such analysis using a hypothetical example. No extramural funding was used to support this work, and the authors are solely responsible for the design, drafting, and editing of this manuscript and its final contents.

Assessment of traditional time-to-event analysis of composite end points

Although the choice of end points for trials of medications, devices, and biologics remains controversial, Conte et al recently identified those individual and composite end points that are among the most important for CLI.¹³ Use of such composite end points has several advantages. First, the composite end points are intended to capture the meaningful clinical effects of the proposed therapy for patients with CLI. From a clinician's perspective, the composite end point captures important outcomes besides death as a "net clinical outcome," thus allowing clinicians to consider the efficacy of the therapy on the aggregate of important outcomes. This may be beneficial in disease states like CLI where various nonfatal clinical findings and measurements (such as extent of wound healing, assessment of rest pain, repeat revascularization) contribute to significant morbidity, have the ability to be linked to a physiologic effect/improvement and are important to both patients and clinicians. Second, use of composite end points potentially reduces the number of patients needed to demonstrate statistical differences between two therapies and allow reasonable sample sizes and shorter duration. This is relevant within CLI trial populations, where there are significant challenges in identifying and retaining patients and the trial costs of biologics and devices are high.

Despite the benefits of composite end points listed above, there are several limitations that regulatory agencies and clinicians should consider in evaluation of trials using time-to-event analysis. First, only the initial event during the trial is considered for statistical analysis and subsequent events are ignored. This limitation is illustrated in the following example: in a randomized trial of patients with CLI being treated with a drug-coated balloon versus standard angioplasty for tibial artery disease with a composite end point of death, amputation, or revascularization, a subject enrolled may require multiple revascularizations within the first 3 months and subsequently require a major amputation. In the traditional time to even analysis, only the first revascularization is captured in the primary efficacy analysis; thus, the patient who eventually underwent amputation is considered equivalent to another subject who required a single repeat revascularization procedure in the primary statistical analysis. However, from a clinical perspective, the investigational device was less effective in the first patient compared with the second patient. Importantly, recurrent events after the first event are censored, which may be particularly concerning in CLI trials where nonfatal event rates are very high. Ignoring recurrent more serious events in the time-to-event analysis may not provide a comprehensive picture of the efficacy of one therapy compared with another. A second major limitation of the traditional analysis is that it

assumes each component of the composite to be of equal significance, although clearly different components of the composite have varying degrees of clinical importance. Within time-to-event analysis, each end point receives a “0” or “1” weighting; therefore, an uncomplicated revascularization would be considered with equal weight as amputation or death. Assessment of the efficacy becomes even more challenging when one outcome occurs with greater frequency and drives the composite towards statistical significance.

Another major limitation complicating interpretation of the traditional analysis revolves around issues of directionality of the effect, which may not be uncommon when composite end points incorporate both safety and efficacy end points, often referred to as a “net clinical benefit.” However, interpretation of a trial using this approach is challenging if the effect of the therapy on the individual safety and efficacy end points go in different directions. Such an example is illustrated with results of the CREST trial where the primary composite end point of periprocedural stroke, myocardial infarction, or death, or ipsilateral stroke in the following four years was not different between the carotid stenting and carotid endarterectomy arms.¹⁵ However, periprocedural stroke favored the endarterectomy arm, whereas myocardial infarction favored the carotid stenting arm. Consequently, the differing directionality of effect of the therapies among the various components of the composite end point makes interpretation of the overall no difference between therapies potentially difficult to interpret. Furthermore, the different directionality of the components may decrease power as well. Some have advocated separately identifying efficacy end points and safety end points. For CLI, such a separation would be attractive because it allows for a clearer signal to define the clinical benefits of a therapy that only targets the limb while allowing key safety concerns to be separately identified and highlighted. This is of particular relevance in CLI therapies given that these trials are often under-powered to detect safety signals.¹⁶ Thus, the true safety profile of a new therapy may not be fully characterized in a single study but could be assessed when the safety data are combined across the development program. However, this would require a complete understanding of the directionality of effect of the therapy on the components of the composite, which is often limited in early phase therapies. Additionally, both the use of “net clinical benefit,” as well as separately reporting safety and efficacy components continues to have major limitations in that recurrent events are not considered in the statistical analysis and all end points are given equal weight in the statistical consideration.

Assessment of trials using objective performance goals

The use of OPGs, or therapeutic benchmark controls, in a single arm study design has also been accepted for novel therapies of CLI to allow for trials to be performed with smaller sample sizes.^{13,17} From a regulatory standpoint, the use of well-documented historical controls is a viable alternative to randomized control trials in specific circumstances. For instance, the Food and Drug Administration (FDA) has approved several cardiac valve prosthesis from single-arm studies comparing newer devices with historical controls.¹⁸ Use of OPGs may be feasible for trials investigating newer devices for revascularization in CLI by comparing efficacy of these with historical controls of bypass surgery or traditional angioplasty.¹³ In fact, the use of OPGs has been estimated to reduce sample size by roughly 75% with elimination of the control group and the performance goal being a fixed number with no variability.¹⁹

Although OPGs offer a feasible approach to compare newer therapies without randomized trials, its utility is limited in CLI given the heterogeneous population with respect to clinical presentation, disease severity, and expected outcomes. Consequently, use of historical controls may not be applicable in all circumstances given rapid advancements in therapies and prior historical controls may not be representative of current practice. Importantly, the

selection of the OPG's for comparison may be of debate given that optimal medical therapy and outcomes for CLI is still controversial, although Conte et al have laid a foundation by nine safety and efficacy measures.¹³ In addition, use of OPGs may not be applicable to CLI trials investigating biologics, given these therapies are relatively recent. Therefore, use of such nonrandomized and controlled trial designs with OPGs may not be appropriate in all circumstances, and an alternative randomized trial design that captures recurrent events and allows for adequate power with similar sample sizes is required.

Other statistical approaches for analysis of composite end points

There has been increasing discussion of use of alternative approaches to analysis of composite end points in a randomized fashion given the inherent limitations of the traditional time to first event analysis and OPGs.^{20–23} One such approach that may be ideal for use in trials of biologics and devices for treatment of CLI is the global rank method, which was initially proposed in 1984 as a means of dealing with composite and repetitive end points.²⁴ Unlike the traditional time-to-event analysis, global-rank methodology allows consideration of recurrent events in the final, permits different directionality of the effect of the therapies on the component end points and differentiates severity of events by assigning relative rankings to events.

Using the global rank method, a hierarchy of end points based on clinical importance is created a priori based on consensus of expert opinion. The end points are placed on a continuum, where end points are ranked from least favorable to most favorable. Each subject is then assigned a relative rank based on the events that he or she experiences during the course of the trial. Patients with worse outcome (e.g., death) are given a worse rank than patients with an event that is considered by consensus to be relatively less important (e.g., complete wound healing). For each arm of the trial, the rankings of each patient within each arm of the study are incorporated into a single statistic that is compared between therapies. Typically, the trial design may specify how to deal with “ties,” such as patients who have wound healing earlier after enrollment receive a better ranking than patients that have wound healing later. Alternatively, the trial design may permit ties to exist, such that patients are given the same weight.

Although it was first proposed to be used in cardiovascular trials by Califf et al in 1990,²⁵ it was not widely adopted because of lack of consensus of expert opinion on relative weighting of end points of interest (e.g., myocardial infarction vs stroke in trials of antithrombotics in acute coronary syndrome). However, more recently, global-rank has been of increased interest in heart failure trials^{21,22} and incorporated into exploratory analysis of statistical plans of heart failure trials such as the RELAX trial (ClinicalTrials.gov identifier: NCT00763867). This has been possible in heart failure because investigators have come to a consensus on relative ranking of different end points of trials of mechanical assist devices.

Use of global rank method in CLI

The Global Rank Method may be adopted within CLI trials given that it may be possible to develop consensus on relative ranking of end points within these trials. We review the steps necessary for using the global rank method for a CLI trial below by using a hypothetical example of a new biologic agent for treatment of CLI.

Hypothetical example: investigators seek to determine the efficacy of a new biologic factor for the treatment of CLI patients with ischemic ulceration of lower extremities (Fontaine stage IV). Their current traditional time-to-event analysis incorporates death and major amputation as the primary composite end point; however, the investigators recognize that complete wound healing and complete pain resolution are goals for treatment and wish to

include these two variables within the primary composite as well. They recognize the limitations of the time-to-event analysis given the patients will have subsequent events after the first non-fatal end point. Consequently, the investigators want to know how the global rank methodology may be applicable to their trial, the strengths and limitations of the methodology and how it may affect their power calculations.

Step 1: Choice of end points

The first step for use of global rank is to determine clinical end points of interest. As with any trial methodology, the end points of clinical importance should be measureable. Unlike the traditional trial approach, the frequency with which the end points occur relative to one another is not as important for the global rank since the worst outcome is ranked the worst. The authors, who are vascular medicine specialists and represent different subspecialties, have proposed the following end points for this example of an investigational biologic factor: (1) death, (2) major amputation (above ankle), (3) minor amputation (below ankle), (4) complete wound-healing, and (5) rest pain resolution off narcotics. Certainly, for trials of devices, other end points may be included, such as technical success, procedural success, vessel patency, and complications.²⁶ Similarly, given the importance of quality of life in the CLI population, quality of life estimates may be incorporated within the composite, such that an improvement in quality of life by a certain percentage decided upon a priori with one of the investigational therapies would be considered a positive end point. However, until there are validated quality of life instruments in the CLI population, visual pain scales may be used as a substitute within the composite. For simplicity and illustrative purposes, we have limited to the 5 end points above.

Step 2: Relative ranking of end points

After selection of appropriate end points, it is crucial that experts in the field develop consensus on hierarchical ranking of end points a priori. The Global Rank Method can only be used if a committee of experts is able to provide relative ranks of the components of the composite. For the present example, the authors ranked the end points listed above on a continuum from least favorable events to most favorable (Table I). The group chose hard end points of most clinical importance to patients and clinicians, where death, major amputation, minor amputation, and incomplete healing are unfavorable outcomes, and complete healing with pain and complete pain resolution are more favorable outcomes. As noted above, select trials of devices may include procedural success and complications and patency within the analysis. As these are traditionally thought of as intermediate end points, they would likely be included further down in the hierarchical global rank, though a group of experts will need to determine the consensus ranking relative to other selected end points. Similarly, quality of life and visual pain scales may be included as well within the trial end point analysis provided that the group of experts is able to reach consensus on the relative weighting. Modifications to these end points may be necessary depending upon the population of CLI being investigated (rest pain Fontaine Stage 3/Rutherford Stage 4 vs ischemic ulcer Fontaine Stage 4/Rutherford Stage 4 or 5). For simplicity, we will stick to the use of the end points listed in Table I for this hypothetical example.

Step 3: Power calculation

The global rank method retains much of the statistical power and potentially smaller sample sizes associated with the traditional composite end point analysis,^{27–29} even though it is able to provide a more global picture of the recurrent events that occur with a therapy.

The traditional time-to-event power calculations are provided in Table II. We provide 3 estimates of sample size based on an assumption of 39% 1-year rate of death or major amputation for the control group, and estimated event rate of 29% for the new therapy.

(roughly 25% reduction in events for a large effect), estimated event rate of 31% (roughly 20% reduction in events for a medium effect), and estimated event rate of 31% (roughly 10% reduction in events for a small effect). The composite event rates in this hypothetical group are estimates of current literature.^{1, 30–32}

Estimates of the 1-year placebo (control group) event rates for each component outcome are extrapolated from published CLI literature in Table III,^{1,2,30–33} alongside three possible scenarios of treatment effects where the investigational therapy has a large, medium, or small effect. As the test used for the Global Rank methodology is the Wilcoxon–Mann–Whitney statistic, the power calculations may be computed using computer simulations or using published formulas. The resulting sample size calculations for the Global Rank Method are provided in Table IV and are based on formulas described by Tang.³⁴ For the three scenarios in our example, the power with global rank was preserved with smaller sample sizes. However, it is important to note there will be some cases where the sample size estimates with global rank will be very similar or possibly larger than estimates for time-to-event analyses. There are major advantages of the Global Rank Method over the traditional time-to-event analysis, as follows: it provides a framework for allowing tiers or rankings of end points capture recurrent events, and therefore, may better represent the change in overall clinical burden in response to therapy.

Step 4: Primary analysis upon completion of trial

For the primary analysis, subjects are placed into “buckets” according to the hierarchical order of end points listed in Table I. Patients that experience the worst event, that is, death, are placed into the first bucket and assigned the worst rank. Among the survivors, patients that have the next worst event during the course of the study, that is, major amputation, are placed in the second bucket and given the next worst rank. It is possible to place a hierarchical order within this bucket as well, such that patients that experience an above the knee amputation receive a worse rank than those with below-the knee amputation. However, given the need for simplicity and general clinical consensus, we provide a system that continues the current practice of viewing all major amputations as similar outcomes. Next, patients that have the next worst outcome, that is, minor amputation without a major amputation, are placed into the third bucket. As with major amputation, it may be possible to further categorize those with minor amputations into a hierarchical order as well; however, for simplicity, the group chose to simplify this bucket into broad category of minor amputations assuming their severity to be of similar severity. Easily, a trial can become more granular and use the following hierarchical order: above the knee > below the knee < below the ankle, given goals of therapy typically to preserve as much of the lower limb as possible, since severity of amputation has a significant impact on quality of life.³⁵ As long as it is prespecified in the analysis plan, the outcomes may be ranked in as many levels and much granularity as the committee decides for a particular trial.

Essentially, all subjects are categorized into one of the buckets according to the worst outcome he or she experiences. Patients in whom an unfavorable outcome (i.e., death, amputation) does not occur during study period are then placed into buckets of the favorable outcomes (in our example, this is complete healing with pain and the next is complete healing with complete pain resolution). Those that do not have any favorable or unfavorable outcomes are placed in the bucket of incomplete healing. Ties in this scheme are may be dealt with time to worst event (i.e., patients who die earliest after randomization receive the worst ranks).

After all subjects are ranked according to the prespecified criteria, the two treatment arms are then evaluated using previously described statistical methods³⁶ including the Wilcoxon–Mann–Whitney test. The results from these analyses include a *P* value for the test of

differences in distributions. Others have proposed alternatively summarizing these comparisons between the 2 groups using the so-called unmatched win-ratio,²⁰ which is discussed below in alternative methods.

Potential advantages of the global rank

There are several major advantages of the global rank method for clinical trial design. Unlike the time to first event analysis, most pre-identified end points are captured in the final statistic. Of note, should a worse ranking event occur first, then subsequent minor events would not be considered. However, given that the worse outcome during course of study period is usually the most important event of interest, this limitation is acceptable. Importantly, recurrent events are captured, except of course in cases of death where subsequent events are not possible. Additionally, both favorable and unfavorable outcomes are also captured in the final test statistic.

Another advantage deals with directionality. In the traditional analysis, if the directionality of the benefit of the investigation therapy differs among the components of the composite (e.g., improvement of the investigation therapy on amputation at the expense of increased mortality), then interpretation of the composite is difficult. With global rank, the directionality of the events is not important as all events are captured in the final statistic. Furthermore, events with greater clinical importance (decided a priori by committee of experts) are given greater relative weight and contribute more to the final statistic compared with the time-to-event analysis where the overall effect is assumed to apply to each component of the composite equally.³⁷⁻³⁹ However, trials using global rank need to collect data stringently given that to maintain power, and therefore, it is crucial to have accurate follow-up of these patients and minimize lost-to follow-up or missing data. To help minimize the effect of missing data, it must be prespecified before the trial onset how such missing data will be accounted for. To further see the illustrate the advantages of the Global Rank Method and compare outcomes with traditional time-to- event analysis, future studies should run the analysis from real-world published trials.

Limitations

There are several limitations with using the global methodology as well. First, a true survival curve (e.g. Kaplan-Meier estimate) is not generated for the composite since all events are shown that occur throughout the trial. Certainly, the investigators may show the survival estimates for each component of the end point with the understanding that time to first event is not the primary analytic plan. Additionally, the overall global rank comparison may be statistically significant, though the individual components used for the primary analysis may not be significantly different between the two arms. It is therefore important to display not only the global rank end point but also the individual end points for each treatment under study so the reader has a more comprehensive understanding of the treatment differences. This is also a critique of the traditional analysis of time to first event, but the global rank method somewhat minimizes this deficiency given that the worst outcomes count more toward the final statistic. Similar to the traditional method, it is prudent for investigators to only include those end points that are affected by the treatment of interest to maintain power; Including analysis of end points that are unaffected by the treatment may add “noise” to the treatment effect thereby diminishing statistical power to detect a difference between treatments. Finally, an accepted follow-up time period is required for global rank assessments, and this can be contentious. However, for CLI, most patients and clinicians agree that 1-year outcomes are clinically meaningful.

The relative rankings used for the hypothetical example are based upon the consensus from the authors, which represent vascular specialists from different fields. However, future

trials may consider use of a more formal RAND modified Delphi process for consensus of ranking of trial end points for CLI trials. Such a process should include representatives/key opinion leaders within the field representing different subspecialties, representatives of the FDA, and representatives of industry. Inclusion of the key shareholders will allow more generalizable consensus on scales for relative ranking, though each trial may modify the ranking and end points chosen. A possible solution for achieving agreement between the clinical community, industry, and the FDA for relative rankings include mediums such as the Peripheral Academic Research Consortium, which is currently working to standardize definitions for end points within trials.

Other novel methodologies for analysis of complex end points

As alluded to earlier, there has been increasing discussion on other methodologies for analyzing composite end points. Besides global rank, Pocock et al introduced a slightly different approach to compare outcomes, using a win-ratio method for comparison of composite end points between different therapies.²⁰ In their paper, Pocock et al propose method of evaluating composite end point using a matched (subjects in the 2 treatment groups are matched on pre-treatment risk profile) and unmatched version. The unmatched version is very similar to the Global Rank method. In the matched version, matched pairs of patients are compared first on those end points with greater relative significance (e.g., death) and subsequently on less clinically significant events (e.g., rehospitalization). The calculated win-ratios are compared to assess the treatment effect and give a final statistic of efficacy. However, a significant limitation in using this approach for CLI trials is the lack of a well-accepted risk profile tool (similar to a GRACE or TIMI risk score for ischemic outcomes in acute coronary syndrome^{40,41}) to differentiate pretreatment risk for adverse events among CLI patients. However, use of clinical staging (e.g., Rutherford 4 vs 5 at enrollment) or perhaps anatomical staging (TASC II classification)⁴² can potentially be used to help perform the win-ratio by profiling patients before treatment and randomly comparing patients from the same stage.

Unlike the global rank method and the win-ratio methods that provide a relative rank to the components of the composite end point, Armstrong et al recently introduced the methodology of weighted composite approach to trial end point analysis. In this method, a modified Delphi panel of experts assigns actual weights for each component of the composite end point a priori.^{23,43} Each patient event is incorporated into the analysis by multiplying the weight of the event with the patient's total score from that point forward. The score for each subject within a specific period is evaluated, and the aggregate score provides the number of weighted patients at risk during a specific period (a "severity-weighted survival score"). The modified log-rank statistic is computed by evaluating the total severity-weighted score at each time-point in the follow-up interval.

Conclusion

Although the use of composite end points or OPGs in CLI trials is used to capture meaningful outcomes and potentially help reduce the number of patients required for a study, there are several limitations in interpretation of these results. The global rank methodology addresses some of these limitations by inclusion of recurrent events and by differentiating the hierarchical order of the various components. We demonstrate through the use of a hypothetical example of biologic compound that the global rank can be used within a randomized trial design to capture and analyze meaningful outcomes while preserving statistical power. Such methodology should be considered in the design of future trials of CLI.

References

1. Nikol S, et al. Therapeutic angiogenesis with intramuscular NV1FGF improves amputation-free survival in patients with critical limb ischemia. *Mol Ther*. 2008; 16:972–8. [PubMed: 18388929]
2. Belch J, et al. Effect of fibroblast growth factor NV1FGF on amputation and death: a randomized placebo-controlled trial of gene therapy in critical limb ischaemia. *Lancet*. 2011; 377:1929–37. [PubMed: 21621834]
3. Brass EP, et al. Parenteral therapy with lipo-ecraprost, a lipid-based formulation of a PGE1 analog, does not alter six-month outcomes in patients with critical leg ischemia. *J Vasc Surg*. 2006; 43:752–9. [PubMed: 16616232]
4. Powell RJ, et al. Results of a double-blind, placebo-controlled study to assess the safety of intramuscular injection of hepatocyte growth factor plasmid to improve limb perfusion in patients with critical limb ischemia. *Circulation*. 2008; 118:58. [PubMed: 18559703]
5. Rajagopalan S, et al. Regional angiogenesis with vascular endothelial growth factor in peripheral arterial disease: a phase II randomized, double-blind, controlled study of adenoviral delivery of vascular endothelial growth factor 121 in patients with disabling intermittent claudication. *Circulation*. 2003; 108:1933–8. [PubMed: 14504183]
6. Morishita R, et al. Safety evaluation of clinical gene therapy using hepatocyte growth factor to treat peripheral arterial disease. *Hypertension*. 2004; 44:203–9. [PubMed: 15238569]
7. Tateishi-Yuyama E, et al. Therapeutic angiogenesis for patients with limb ischaemia by autologous transplantation of bone-marrow cells: a pilot study and a randomized controlled trial. *Lancet*. 2002; 360:427–35. [PubMed: 12241713]
8. Jones WS, Annex BH. Growth factors for therapeutic angiogenesis in peripheral arterial disease. *Curr Opin Cardiol*. 2007; 22:458. [PubMed: 17762548]
9. Powell RJ, et al. Safety and efficacy of patient specific intramuscular injection of HGF plasmid gene therapy on limb perfusion and wound healing in patients with ischemic lower extremity ulceration: Results of the HGF-0205 trial. *J Vasc Surg*. 2010; 52:1525–30. [PubMed: 21146749]
10. Werk M, et al. Inhibition of restenosis in femoropopliteal arteries. Paclitaxel-coated versus uncoated balloon: femoral paclitaxel randomized pilot trial. *Circulation*. 2008; 118:1358–65. [PubMed: 18779447]
11. Tepe G, et al. Local delivery of paclitaxel to inhibit restenosis during angioplasty of the leg. *N Engl J Med*. 2008; 358:689–99. [PubMed: 18272892]
12. Stoner MC, et al. Mid-term results with laser atherectomy in the treatment of infrainguinal occlusive disease. *J Vasc Surg*. 2007; 46:289 e1–95 .e1. [PubMed: 17600661]
13. Conte MS, et al. Suggested objective performance goals and clinical trial design for evaluating catheter-based treatment of critical limb ischemia. *J Vasc Surg*. 2009; 50:1462 e3–73 .e3. [PubMed: 19897335]
14. Kumar A, et al. FDA perspective on objective performance goals and clinical trial design for evaluating catheter-based treatment of critical limb ischemia. *J Vasc Surg*. 2009; 50:1474–6. [PubMed: 19897334]
15. Brott TG, et al. Stenting versus endarterectomy for treatment of carotid-artery stenosis. *N Engl J Med*. 2010; 363:11–23. [PubMed: 20505173]
16. Brass EP, et al. Risk assessment in drug development for symptomatic indications: a framework for the prospective exclusion of unacceptable cardiovascular risk. *Clin Pharmacol Ther*. 2006; 79:165–72. [PubMed: 16513441]
17. Grunkemeier GL, Jin R, Starr A. Prosthetic heart valves: objective performance criteria versus randomized clinical trial. *Ann Thorac Surg*. 2006; 82:776–80. [PubMed: 16928482]
18. Chen E, et al. FDA perspective on clinical trial design for cardiovascular devices. *Ann Thorac Surg*. 2006; 82:773. [PubMed: 16928481]
19. Parides MK, et al. Progress versus precision: challenges in clinical trial design for left ventricular assist devices. *Ann Thorac Surg*. 2006; 82:1140–6. [PubMed: 16928569]
20. Pocock SJ, et al. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012; 33:176–82. [PubMed: 21900289]

21. Felker GM, Anstrom KJ, Rogers JG. A global ranking approach to end points in trials of mechanical circulatory support devices. *J Card Fail.* 2008; 14:368–72. [PubMed: 18514927]
22. Felker GM, Maisel AS. A global rank end point for clinical trials in acute heart failure. *Circ Heart Fail.* 2010; 3:643–6. [PubMed: 20841546]
23. Armstrong PW, et al. Refining clinical trial composite outcomes: An application to the Assessment of the Safety and Efficacy of a New Thrombolytic–3 (ASSENT-3) trial. *Am Heart J.* 2011; 161:848–54. [PubMed: 21570513]
24. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics.* 1984; 40:1079–87. [PubMed: 6534410]
25. Califf R, Harrelson-Woodlief L, Topol E. Left ventricular ejection fraction may not be useful as an end point of thrombolytic therapy comparative trials. *Circulation.* 1990; 82:1847–53. [PubMed: 2225381]
26. Diehm N, et al. A call for uniform reporting standards in studies assessing endovascular treatment for chronic ischaemia of lower limb arteries. *Eur Heart J.* 2007; 28:798–805. [PubMed: 17317699]
27. MoyÈ LA, Davis BR, Hawkins CM. Analysis of a clinical trial involving a combined mortality and adherence dependent interval censored endpoint. *Stat Med.* 1992; 11:1705–17. [PubMed: 1485054]
28. Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. *Stat Med.* 1999; 18:1341–54. [PubMed: 10399200]
29. McMahon RP, Harrell FE Jr. Power calculation for clinical trials when the outcome is a composite ranking of survival and a nonfatal outcome. *Control Clin Trials.* 2000; 21:305–12. [PubMed: 10913806]
30. Matoba S, et al. Long-term clinical outcome after intramuscular implantation of bone marrow mononuclear cells (Therapeutic Angiogenesis by Cell Transplantation [TACT] trial) in patients with chronic limb ischemia. *Am Heart J.* 2008; 156:1010–8. [PubMed: 19061721]
31. Adam DJ, et al. Bypass versus angioplasty in severe ischaemia of the leg (BASIL): multicentre, randomized controlled trial. *Lancet.* 2005; 366:1925–34. [PubMed: 16325694]
32. Randon C, et al. Angioplasty or primary stenting for infrapopliteal lesions: results of a prospective randomized trial. *Cardiovasc Intervent Radiol.* 2010; 33:260–9. [PubMed: 19957178]
33. Walter DH, et al. Intraarterial administration of bone marrow mononuclear cells in patients with critical limb ischemia/clinical perspective. *Circ Cardiovasc Interv.* 2011; 4:26–37. [PubMed: 21205939]
34. Tang Y. Size and power estimation for the Wilcoxon–Mann-Whitney test for ordered categorical data. *Stat Med.* 2011; 30:3461–70. [PubMed: 22086799]
35. Peters EJG, et al. Functional status of persons with diabetes-related lower-extremity amputations. *Diabetes Care.* 2001; 24:1799–804. [PubMed: 11574445]
36. Lachin JM. Worst rank score analysis with informatively missing observations in clinical trials. *Control Clin Trials.* 1999; 20:408–22. [PubMed: 10503801]
37. Lim E, et al. Composite outcomes in cardiovascular research: a survey of randomized trials. *Ann Intern Med.* 2008; 149:612–7. [PubMed: 18981486]
38. Cordoba G, et al. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ.* 2010:341.
39. Freemantle N, et al. Composite outcomes in randomized trials. *JAMA.* 2003; 289:2554. [PubMed: 12759327]
40. Antman EM, et al. The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. *JAMA.* 2000; 284:835–42. [PubMed: 10938172]
41. Granger CB, et al. Predictors of hospital mortality in the Global Registry of Acute Coronary Events. *Arch Intern Med.* 2003; 163:2345–53. [PubMed: 14581255]
42. Norgren L, et al. Inter-society consensus for the management of peripheral arterial disease (TASC II). *J Vasc Surg.* 2007; 45:5.
43. Bakal JA, Westerhout CM, Armstrong PW. Impact of weighted composite compared to traditional composite endpoints for the design of randomized controlled trials. *Stat Methods Med Res.* 2012 Epub ahead of print.

Table I

Hierarchical order of end points in descending order of severity

1	Death
2	Major amputation
3	Minor amputation
4	Incomplete healing
5	Complete healing with pain
6	Complete pain resolution

Table II

Sample size calculations for the traditional time-to-event analysis of the amputation free survival end point

Traditional time-to-event: major amputation or death[*]	Total sample size for scenario 1 (large effect)[*]	Total sample size for scenario 2 (medium effect)[†]	Total sample size for scenario 3 (small effect)[‡]
80% Power	702	1106	4502
90% Power	940	1482	6026

^{*} Assuming a 39% 1-year rate of death or major amputation for the control group, and an event rate of 29% for the new therapy.

[†] Assuming a 39% 1-year rate of death or major amputation for the control group, and an event rate of 31% for the new therapy.

[‡] Assuming a 39% 1-year rate of death or major amputation for the control group, and an event rate of 35% for the new therapy. Assumes a 1:1 randomization ratio, a two-sided type I error rate of 0.05, and that all subjects are followed for no more than one year. Calculations were performed in nQuery Advisor 6.0. For the time to death end point under scenarios 1 and 2, the required sample sizes are 2186 and 2926 for 80% and 90% power, respectively. For the time to death end point under scenario 3, the required sample sizes are 6242 and 8356 for 80% and 90% power, respectively.

Estimated event rates for hypothetical example

Table III

Ranking	End point	Placebo 12-month estimated rate of events	Scenario 1: large effect	Scenario 2: medium effect	Scenario 3: small effect
1	Death	25%	20%	20%	22%
2	Major amputation	14%	9%	11%	13%
3	Minor amputation	12%	7%	9%	12%
4	Incomplete healing	7%	6%	6%	7%
5	Complete healing with pain	17%	25%	21%	17%
6	Complete pain resolution	25%	33%	33%	29%

Scenarios 1 to 3 correspond to potential 12-month rates of a hypothetical biologic therapy. Scenario 1 is a large treatment effect; scenario 2 is a medium sized treatment effect; and scenario 3 is a small treatment effect.

Table IV

Power calculations using global rank method

Global rank power calculations*	Total sample size (scenario 1: large effect)	Total sample size (scenario 2: medium effect)	Total sample size (scenario 3: small effect)
80% Power	466	651	3182
90% Power	622	870	4258

* Based on event rates listed in Table II. The sample sizes are estimated using the formulas of Tang.³⁴