# PIES Method of Critique

*Steven Baroletti, PharmD, BCPS, and Paul M. Szumita, PharmD, BCPS*

**Abstract:** Critical evaluation of clinical trials is essential prior to applying results into clinical practice. We have created the acronym, PIES (Population, Intervention, Endpoints, and Statistics) that encapsulates the four basic aspects of a trial. The PIES method creates a systematic approach to critically evaluating a trial, allowing practitioners to formulate opinions as to the applicability to clinical practice.

**Key Words:** trials, evidence-based medicine, PIES

(*Crit Pathways in Cardiol* 2004;3: 205–208)

In today's era of evidence-based medicine, treatment decisions are routinely supported by results from clinical trials. In the setting of many various methods of research, it has clearly been established that the randomized clinical trial (RCT) is the gold standard of testing a scientific theory.[1] Although the RCT is designed to eliminate many of the biases that may occur with observational research, there are still opportunities for flaws within this design. Many times, results of trials are extrapolated to different patients and treatments than are truly tested. Additionally, the true benefit may not always be the same as portrayed or even be as clinically relevant. Therefore, critical evaluation of a trial is essential prior to applying results into clinical practice. Oftentimes, the clinician using the literature does not have the time to efficiently critically analyze a clinical trial.

The purpose of this paper is to introduce a new methodology to systematically approach the key components of clinical trials and give instruction of how to evaluate each of these components. We have created the acronym PIES (Patient Population, Intervention, Endpoints, and Statistics) that simplifies the 4 basic aspects of a trial. This easy-to-use checklist will serve as an aid to remember each of the components to evaluate every time a trial is read (Table 1).

From Brigham and Women's Hospital, Boston, MA, and Northeastern University, Boston, MA.

Reprints: Steven Baroletti, PharmD, BCPS, Clinical Pharmacy Specialist, Brigham and Women's Hospital, Boston, MA. (Tel): 617-732-7165; (fax): 617-566-2396. E-mail: Sbaroletti@partners.org.



ISSN: 1535-282X/04/0304-0205

DOI: 10.1097/01.hpc.0000146868.94502.b0

## P: Population

There are 2 major aspects that a reader should evaluate with regard to the patient population in a study. First, carefully review the baseline characteristics for any factors that are statistically different between treatment groups. Any confounding variables or differences between groups may skew the results of the trial. This usually will not occur when patients are randomized, although the only way to assure this is to statistically compare groups. If statistical differences exist, it is now the decision of the reviewer as to whether this is clinically relevant. If the confounder identified has an established impact on the outcome being tested, the validity of the results may be weakened.

The second aspect of the patient population to evaluate is the overall characteristics of the patients enrolled in the trial. The inclusion and exclusion criteria determine patient enrollment in an RCT. Data are often inappropriately applied to patient populations that were not studied. This commonly occurs where minimal or conflicting data exist.

The recently published Pravastatin or Atorvastatin Evaluation and Infection Therapy—Thrombolysis in Myocardial Infarction 22 (Prove-IT TIMI-22) study demonstrated the benefit of high dose 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA) (statin) therapy on reducing cardiovascular endpoints when initiated within 10 days of an acute coronary syndrome (ACS).[2] Patients who fit these inclusion criteria will certainly benefit from this therapy. However, to extrapolate these results and initiate high-dose statin therapy in patients without a history of ACS would be inappropriate use of these data. Although these patients may benefit from statin therapy based on other data, the evidence does not directly support initiation of the high-dose statin therapy in all patients with a history of coronary artery disease but not recent ACS.[3]

In an ideal situation, data would not be extrapolated beyond the studied patient population. Many large RCTs exclude specific patients and disease states to create a homogenous population. In the absence of patient specific evidence, these data often serve as a guide for clinical decisions in special populations. There are no RCTs directing treatment of hypertension following solid organ transplant. There are, however, multiple RCTs guiding therapy for treatment of hypertension in other populations that may serve as

**TABLE 1.** The PIES Method of Critically Evaluating Clinical Trials

| | |
|---|---|
| **P**atient Population | Are there any major differences in patient characteristics that may confound the results?<br>Evaluate inclusion and exclusion; criteria serve as a guide to the patients for whom the results may be applicable |
| **I**ntervention | Is the intervention being tested representative of current practice or derived from previous well-conducted studies? |
| **E**ndpoints | Do the endpoints of the trial truly represent what is claimed as being studied?<br>Is the endpoint used in the trial clinically significant?<br>If a surrogate endpoint is used, is it validated for correlation to a hard clinical endpoint? |
| **S**tatistics | What type of data ae being assessed (nominal, ordinal, continuous)?<br>Are the statistical tests used to evaluate the data appropriate?<br>Is the effect size clinically relevant?<br>Evaluate the results in absolute values and calculate number needed to treat (NNT)* |

*NNT = (1/ARR) × 100.

an aid in choosing an antihypertensive agent.[4,5] If used judiciously, extrapolation may be beneficial.

Although extrapolating data may influence a clinical decision, caution should be used when attempting to extend the use to a therapy beyond the population in which it was studied. The effect of drotrecogin alfa (activated protein C) on mortality was studied in patients with severe septic shock and multiorgan dysfunction.[6] Since this is the only patient population with proven efficacy, stringent use of activated protein C is indicated. The inclusion and exclusion criteria guide the appropriate use of the intervention in a specific population.

## I: Interventions

The second major component of a trial is the interventions, which are the therapies being tested. Multiple trials conducted in similar patient populations may produce variable results. To adequately compare these results, it is important to note the following: therapy, dose, titration, dosing schedule, and timing of initiation with relation to a specific procedure or event.

In a RCT, results can only be validated if the control group is receiving optimal therapy. The use of $\beta$-blocker therapy has clearly been established for treatment of heart failure.[7,8] However, controversy existed as to if the benefits of the agent carvedilol are in part attributed to the $\alpha$ antagonistic properties of the drug. Thus, an RCT was designed to test this hypothesis compared with metoprolol (selective for the $\beta$-1 receptor) another agent with established benefit in this population.[9] The results of this study demonstrated a significant reduction in mortality in the patients randomized to the carvedilol arm over the patients in the metoprolol arm. At face value, one may believe that carvedilol should now be used instead of metoprolol for treatment of heart failure. However, upon critically analyzing the interventions, it becomes clear that the regimen used in the metoprolol group (immediate release metoprolol tartrate) is substantially different from those used in the prior conducted trials (metoprolol succinate extended release) that exhibited benefit. This may now weaken the validity of the results.

The intervention being tested should be representative of either current practice or from previous well-conducted studies. The interventions implemented for the control group should represent standard of care currently practiced.

## E: Endpoints

The endpoints are the measure of the clinical impact of the intervention. Chosen endpoints should represent the desired outcome of the intervention. The first factor to evaluate is the clinical relevance of the given endpoint.

For life-threatening disease states, prevention of mortality would be a good choice. Other choices for endpoints may be an actual event, procedure, or other measure of clinical deterioration (need for surgery or rescue catheterization, or recurrent ischemia in patients presenting with a myocardial infarction).[10] However, to achieve a statistical difference using hard endpoints, these events must occur frequently enough. Many times, this is not the case, and a more prevalent clinical event must be chosen as a marker for the more severe outcome originally desired to study. This endpoint is known as a surrogate endpoint.

If a surrogate endpoint is used, it must be assessed for validity in relation to the hard endpoint. This is usually supported by a prior study demonstrating a correlation. Judgment must also be used to assess the clinical significance of the endpoint. Using a surrogate endpoint that has no causality to a hard clinical endpoint may weaken the strength of a study. TIMI grade blood flow, a measure of reperfusion following an infarct, has been used as a surrogate endpoint for mortality following MI.[11] This is an example of an appropriate representation of hard endpoints because it is supported by trial data.

Subgroup analysis is another aspect that needs to be approached with caution when evaluating the endpoints of a trial. A subgroup analysis is a group of patients stratified from the total patient population by certain characteristics (eg, age, gender, concomitant medications) and analyzed accordingly.[1]

These data may be beneficial to identify a group of patients that may have different outcomes. However, in most cases a trial is not powered (see Statistics section) to detect a difference between 2 subgroups. Therefore, these results should be reviewed as a hypothesis that needs to be either confirmed of disproved in a future trial.

The endpoints of a trial are the representation of what is truly being studied and what effect can be expected from the given intervention. It is imperative that the endpoint used be clinically significant.

## S: Statistics

The final component of the evaluation of a trial is the statistical analysis. Unless you are designing a clinical trial, a superior understanding of statistics is generally not necessary. However, a basic understanding of the tests used is important. As discussed before, the endpoints chosen in a trial represent clinical effects of a given therapy. Statistical tests are used to determine if a significant difference exists (reject the null hypothesis) between the treatment and the control group.

To determine what is the appropriate statistical test, first you must identify the type of data that are being assessed. Data are measured 1 of 3 ways, nominal, ordinal, or continuous. There are specific statistical tests to analyze these different types of data that account for confounders and the properties of these data. Usually, the statistical tests used in large RCTs are appropriate.

As important as choosing the right endpoint is defining how to measure this endpoint. The method chosen for assessment of statistical significance should be reflective of what is valued in the clinical setting (eg, diagnosis of pulmonary embolism with computed tomograph, opposed to symptoms alone).[12] Additionally, the difference between groups that is chosen to be statistically significant (effect size) should possess clinical importance (eg, reduction in systolic blood pressure of 2 mm Hg in a hypertension trial). To assess that the statistical significance is clinically relevant, the above criteria should be based on previous well-conducted RCTs and current clinical practice.

The next major aspect of statistics to assess is the presentation of the results. Results can be expressed as reductions in risk from one group over another. Several of the commonly used terms may be misleading. This may appear as an odds ratio (OR), absolute risk reduction (ARR), relative risk reduction (RRR), or confidence interval (CI). The ARR is the only expression that directly compares gross results between groups. The others are comparisons of changes made relative to that respective group. It is important not to place extreme emphasis on one value without analyzing all of them. For example, in a study comparing group A with group B, if 1% fewer patients in group A were to have an event than in group B, the ARR would equal 1%. However, if only 2% of patients in group B had that event, the RRR may be presented as 50%. This may exaggerate clinical significance.

A valuable calculation to assess the clinical relevance of the statistical outcome is the number needed to treat (NNT). The NNT calculates the number of patients NNT to experience the benefit from 1 intervention. This can be calculated be taking the reciprocal of the ARR and multiplying by 100.

The last things to look for when evaluating the statistics deal with the sample size. A sample size is calculated prior to enrollment based on the number of patients that are needed in each group to achieve a statistical difference if one truly exists. If the sample size is not met, the trial is subject to "type II" error (accept the null hypothesis when a difference does exist). This can be seen in trials in which a "trend" toward statistical significance does not reach the appropriate *P* value or confidence interval.

Statistical analyses are crucial to determining the validity of a trial. Evaluation of statistics includes assuring that the statistical tests are appropriate, that the effect size is clinically relevant, and the proper sample size is reached. The statistical results can be confirmed for clinical relevance by using the ARR and NNT.

## Summary

The PIES approach to evaluating clinical trials is a systematic approach to critically evaluating a clinical trial. Although not inclusive of all potential flaws within a trial, the PIES method of trial evaluation provides an efficient way of attacking the 4 major components of an RCT, allowing clinicians to formulate opinions as to the clinical value of a paper in addition to those of the authors.

## ACKNOWLEDGMENT

We thank Oluponmile (Ponle) Aboyeji for his creative contributions.

## REFERENCES

1. Lader C, Cannon C, Ohman M, et al. The clinician as investigator: participating in clinical trial in the practice setting. *Circulation*. 2004; 109:2672–2679.
2. Cannon CP, Braunwald E, McCabe CH, et al. Intensive versus moderate lipid lowering with statins after acute coronary syndromes: the Pravastatin or Atorvastatin Evaluation and Infection Therapy–Thrombolysis in Myocardial Infarction 22 Investigators. *N Engl J Med*. 2004;350: 1495–1504.
3. Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: a randomised placebo-controlled trial. *Lancet*. 2002;360:7–22.
4. Chobanian AV, Bakris GL, Black HR, et al. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*. 2003;289:2560.
5. Baroletti S, Gabardi S, Magee C, et al. Calcium channel antagonists as treatment of choice for hypertension in renal transplant patients: fact or fiction? *Pharmacotherapy*. 2003;23:788–801.
6. Bernard GR, Vincent JL, Laterre PF, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med*. 2001;344:699.
7. Packer M, Fowler MB, Roecker EB, et al. Effect of carvedilol on the morbidity of patients with severe chronic heart failure: results of the Carvedilol Prospective Randomized Cumulative Survival (COPERNICUS) study. *Circulation*. 2002;106:2194.

8. Hjalmarson A, Goldstein S, Fagerberg B, et al. Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in congestive heart failure (MERIT-HF): MERIT-HF Study Group. *JAMA*. 2000;283:1295.
9. Poole-Wilson PA, Swedberg K, Cleland JG, et al. Comparison of carvedilol and metoprolol on clinical outcomes in patients with chronic heart failure in the Carvedilol Or Metoprolol European Trial (COMET): randomised controlled trial. *Lancet*. 2003;362:7.
10. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial: assessment of the Safety and Efficacy of a New Thrombolytic Investigators. *Lancet*. 1999;354:716.
11. Ross AM, Coyne KS, Moreyra E, et al, for the GUSTO-I Angiographic Investigators. Extended mortality benefit of early postinfarction reperfusion. *Circulation.* 1998;97:1549.
12. Goldhaber S, Elliott C. Acute pulmonary embolism, part I: epidemiology, pathophysiology, and diagnosis. *Circulation*. 2003;108:2726–2729.