
 Perspective

Why do Phase III Trials of Promising Heart Failure Drugs Often Fail? The Contribution of “Regression to the Truth”

HENRY KRUM, MBBS, PhD, FRACP, ANDREW TONKIN, MD, FRACP

Victoria, Australia

ABSTRACT

There has been considerable recent disappointment with the failure of a number of major new pharmacological strategies for the treatment of chronic heart failure. In turn, there has been much speculation as to why trials of these therapies have not shown benefit. Among a number of plausible and scientifically valid reasons, consideration should be afforded to the potential contribution of “regression to the truth.” Regression to the truth derives from the biological concept of regression to the mean, whereby random fluctuations in a biological variable occur over time, such that the true value of the variable is approached with repeated measurements. This same concept can be applied to clinical trial programs for new drugs for heart failure. Because only strongly positive trials generally go on to phase III testing, and some of these early phase studies are positive by chance alone, on retesting in phase III the results are very likely not to be as strongly positive. Numerous examples of regression to the truth apply for trials of heart failure therapies, as well as in other areas.

A major concern is how to minimize negative outcomes in phase III trials. One approach is to perform major rigorous phase II testing. Alternatively, avoidance of phase II testing will minimize “regression to the truth” because there are no data in phase II from which regression might occur. However, this approach does not obviate the need for an evaluation process in the selection of candidate agents (and their appropriate dose) in order to proceed to definitive testing.

Key Words: Heart failure, clinical trials, research methods.

There has been considerable disappointment within the heart failure community recently with the failure to demonstrate benefits in trials of a number of major new pharmacologic strategies for chronic heart failure. These include studies of blockade of tumor necrosis factor- α (RENEWAL),¹ endothelin receptor antagonism (ENABLE),² and vasoepitidase inhibition (OVERTURE).³

From the National Health and Medical Research Council of Australia Centre of Clinical Research Excellence in Therapeutics, Department of Epidemiology and Preventive Medicine, and Department of Medicine, Monash University, Alfred Hospital, Melbourne Victoria, Australia.

The ideas and opinions expressed in Perspective articles in the *Journal of Cardiac Failure* are those of the authors and do not necessarily reflect those of the Editor, the Publisher, the Heart Failure Society of America, or the Japanese Heart Failure Society.

Manuscript received November 11, 2002; revised manuscript received January 6, 2003; revised manuscript accepted January 20, 2003.

Reprint requests: Dr. Henry Krum, National Health and Medical Research Council of Australia Center of Clinical Research Excellence in Therapeutics, Department of Epidemiology and Preventive Medicine, Monash University, Alfred Hospital, Melbourne Victoria 3004, Australia.

© 2003 Elsevier Inc. All rights reserved.

1071-9164/03/0905-0004\$30.00/0

doi:10.1054/S1071-9164(03)00018-6

There has, in turn, been much speculation as to why these pharmacologic strategies have failed. This is particularly so because common to all of these strategies were apparently high mechanistic plausibility, encouraging early phase clinical data and phase III study programs that were carefully planned, rigorously conducted, and involved committed researchers.

A number of reasons for the failure of these trials have been commonly cited. These include subsequent uncertainty on reexamination of the strength of the biologic rationale for these approaches, inadequate rigor in preclinical testing, the use of surrogate measures rather than clinical outcomes as endpoints in phase II studies, inadequate targeting of the patient population, and issues regarding the dose and dosing interval of certain agents. Most commonly proposed is the concept that a pharmacologic “threshold” may have been reached in treatment of heart failure with high rates of usage of proven background therapies (ie, renin-angiotensin blockade and β -blockade) in contemporary heart failure trials.

All of these reasons for these failures are plausible and scientifically valid. They undoubtedly contribute to the majority

of “failures.” Nevertheless, what is often not considered in research programs related to new heart failure therapies in particular (and all trial programs in general), is the concept of “regression to the truth” occurring within the phase III component of the study program.

What is Regression to the Truth?

To understand regression to the truth we must first consider the concept of regression to the mean. This concept derives from the random fluctuations that can occur in a variable over time. As a consequence, a single measurement of that variable more often yields a value removed from the mean, and the “true” value of the variable is approached with repeated measurements. As a corollary, in population studies, a single measurement of the dependent variable—for example, cholesterol—can lead to an underestimate of the strength of its association with an outcome such as coronary heart disease death (“regression distribution bias”).

Consider a theoretical drug (drug x) being studied to determine its benefit in heart failure, as assessed by a surrogate measure, lowering of plasma norepinephrine (Fig. 1). The left panel shows that there is really no difference in plasma norepinephrine levels before and after drug x. However, the investigators went on to perform a subgroup analysis of those patients with norepinephrine levels above the mean (middle panel), and that subgroup demonstrated a significant reduction in norepinephrine levels with drug x. The investigators might therefore claim that drug x is effective in lowering plasma norepinephrine in patients with high norepinephrine levels. Furthermore, it is these patients (ie, patients with high levels) who are those that are particularly in need of a drug that will lower such elevated levels.

Although it is possible that drug x does indeed lower elevated plasma norepinephrine levels, it is equally plausible (if not more so) that the high plasma levels were “captured” as being falsely or atypically high (for the individual patient)

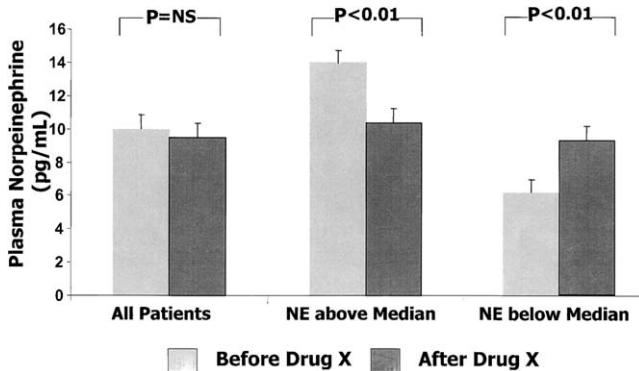


Fig. 1. Theoretical drug (drug x) and its hypothetical impact upon plasma norepinephrine levels for all heart failure patients (left panel) and those with high baseline levels (right panel). These data are illustrative of the concept of regression to the mean, as applied to a biologic variable.

at baseline and then when the same patients were remeasured at a later time point, levels were not as high (ie, classic regression to the mean).

This concept is well-understood for a biologic variable, but how can this concept be applied to that of a clinical trial program for a new drug for heart failure? This is conceptually illustrated in Fig. 2, which depicts early phase trials conducted in the assessment of a variety of potential new drug therapies for heart failure. Each dot represents a trial of a certain drug. As can be seen, some early phase studies will be strongly negative, some strongly positive, but most will cluster around neutrality and, therefore, one can construct a standard bell-shaped curve. We know that many trials of new chemical entities are conducted in the setting of heart failure. Because of the large number of studies conducted, some will be positive by chance and indeed some will be strongly positive by chance. Does this matter? Yes, it does. It is highly likely that only drugs associated with strongly positive trials (ie, those to the right of the vertical dotted line) will go on to phase III testing. Because some of these studies that are positive by chance alone will be among these, then when retested in phase III trials, the results will no longer be strongly positive. This is analogous to the high plasma cholesterol or norepinephrine being retested in the earlier examples.

“Regression to the Truth” in Heart Failure

This concept is true, not just of heart failure trials, but of any drug therapy for any specific indication. What exacerbates the problem in the setting of chronic heart failure is the low percentage strike rate in the development of successful pharmacologic therapies for this condition. Only renin-angiotensin and β -adrenoceptor blocking agents have come to the market over the last 30 years or so.

Therefore, very few promising drugs in early phase would be positive in phase III (if tested) and thus registrable for a heart failure indication. This is illustrated by the open circles below the curved line, interposed on the totality of early phase trials in Fig. 2. This line is curved because, of course, a strongly positive early phase study will make it more likely (but possibly still with low probability) of positive findings in phase III studies. Nevertheless, this still leaves a large number of trials strongly positive in early phase by chance alone (circled cluster) “regressing to the truth.”

Just by way of completion of this concept, it is also possible that there may be a false-negative early-phase result. However, this outcome would be virtually impossible to detect because no sponsor or investigator would take a neutral or negative early phase result and then proceed to phase III testing.

“Regression to the Truth” in Clinical Trials

Does “regression to the truth” occur in the real world of clinical trials? There are a large number of recent examples

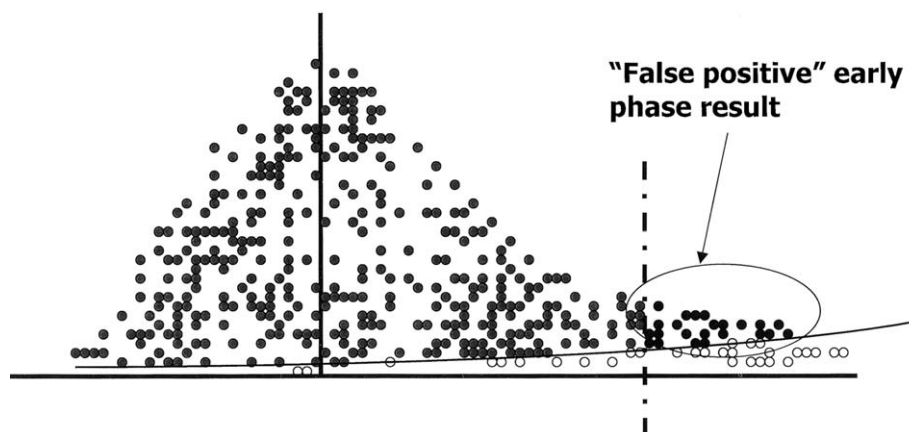


Fig. 2. Theoretical depiction of results of early phase trials conducted in the assessment of a variety of potential new drug therapies for heart failure. Each dot represents a trial of a certain drug. In general, only strongly positive early phase studies (ie, those to the right of the vertical dotted line) will go on to phase III testing. The open circles below the curved line, representing trials positive in phase III (if tested), are thus registrable for a heart failure indication. Trials strongly positive in early phase by chance alone (depicted by the circled cluster of black dots) are false positives (ie, those “regressing to the mean” on retesting).

in which the concept may be operative within programs assessing the potential of new therapies for heart failure. These include both failed phase III trials after positive early-phase programs, as well as further testing after subgroup analysis of major trials. Examples of phase II studies that have gone on to be neutral or negative on phase III testing include the vesnarinone program in phase II⁴ leading to the VEST (Vesnarinone Survival Trial) study,⁵ the REACH-1 (Research on Endothelin Antagonism in Congestive Heart Failure) study⁶ with bosentan leading on to the ENABLE trial,² studies of tumor necrosis factor- α blockade with etanercept⁷ leading on to RENEWAL,¹ and of vasopeptidase inhibition⁸ preceding OVERTURE.³

Examples of subgroup or post hoc analyses of major trials associated with a subsequent neutral or failed formal phase III program include the ELITE-1⁹ trial preceding ELITE-2 (Evaluation of Losartan in the Elderly)¹⁰ and the PRAISE (Prospective Randomized Amlodipine Survival Evaluation)¹¹ trial leading to PRAISE-2.¹² It is noteworthy that in both of these latter examples, the decision to proceed to phase III testing was not based on the positive results of the primary endpoint of the trial, but rather analyses that were not prospectively defined. Furthermore, beneficial effects in subgroups (eg, the nonischemic group in PRAISE¹¹) were not supported by any plausible mechanistic rationale. Thus it should not be surprising that subsequent “definitive” testing of this subgroup yielded a neutral or negative result, at variance to the original observation.

There are a number of specific reasons why “regression to the truth” is a problem in setting of these trials.

Chronic heart failure is a condition notorious for the inability of surrogate markers of disease to reliably predict clinical outcome. Thus large sample sizes are needed to demonstrate definitively clinical efficacy; studies of smaller numbers of patients are much more prone to misleading

results. The effect of candidate drugs on hemodynamic parameters and exercise tolerance measures are clear examples of this phenomenon. Divergent responses have been observed for these and other surrogate markers with regard to long-term clinical outcome in the testing of a number of candidate drug classes.¹³⁻¹⁵ Even ventricular remodeling (seemingly a useful surrogate for clinical outcomes after angiotensin-converting enzyme inhibitor and β -adrenoceptor blocker therapy) may not necessarily be suitable for agents acting via alternative mechanisms.

Another factor that may lead to failure in phase III testing even if the agent is effective in a well-designed phase II trial is geographic and demographic heterogeneity between patients in a trial cohort as well as in their background disease management (of chronic heart failure, as well as important comorbid conditions). This has been noted in a number of recent multinational clinical trials.^{16,17} Variation in use of proven chronic heart failure therapies such as spironolactone and β -blockers clearly compounds this problem.

An important aspect of this discussion is how to minimize negative outcomes in phase III trials. There are two diametrically opposite approaches to addressing this issue.

The first is to perform more rigorous testing in phase II (ie, with greater attention to optimizing the dose and the patient population studied) and to ensure accuracy and precision in estimating treatment effects before proceeding to phase III testing. Although well-designed and well-powered phase II studies may substantially increase the cost of drug development, the “net” cost of inadequate phase II studies may be even higher.

The opposite approach is to avoid phase II testing altogether in heart failure (as has been proposed by others¹⁸) and go straight to phase III studies. This concept is supported by the preceding discussion regarding the absence of useful

surrogate markers in chronic heart failure that might provide reliable guidance as to future clinical success from phase II testing. Indeed, it has been argued¹⁸ that the only reason to perform phase II testing is to obtain information concerning appropriate dosage in the target population to be studied. An “advantage” of this approach, in the context of the present discussion, is that it completely avoids any issue of regression to the truth because there are no data in phase II from which regression might occur. However, this approach does not obviate the need for an evaluation process in the selection of candidate agents (and their appropriate dose) to proceed to definitive testing.

Nevertheless, despite either approach being taken, it seems unavoidable that a high proportion of phase III trials will continue to “fail,” either because of “regression to the truth” or lack of formal previous testing in the relevant patient population. The ultimate safeguard in the drug development process is the current system requiring two pivotal phase III trials to be positive for their predefined endpoint(s) for registration to be obtained. This will hopefully filter the occasional bad treatment that succeeds because of random chance, if not the good treatment that fails.

Conclusions

Regression to the mean is a well-accepted concept in the measurement of biologic variables (eg, plasma norepinephrine), but we should not forget that “regression to the truth” occurs in clinical research programs. Therefore, in addition to the many plausible explanations for the failure of recent phase III heart failure trials, we should consider the concept that “regression to the truth” may be contributing to these disappointing outcomes.

References

1. Coletta A, Clark A, Banarjee P, Cleland J. Clinical trials update: RENNEWAL (RENAISSANCE and RECOVER) and ATTACH. *Eur J Heart Fail* 2002;4:559–61.
2. Kalra P, Moon J, Coats A. Do results of the ENABLE (Endothelin Antagonist Bosentan for Lowering Cardiac Events in Heart Failure) study spell the end for non-selective endothelin antagonism in heart failure? *Int J Cardiol* 2002;85:195–7.
3. Packer M, Califf RM, Konstam MA, Krum H, McMurray JJ, Rouleau JL, et al. Comparison of omapatrilat and enalapril in patients with chronic heart failure: the Omapatrilat Versus Enalapril Randomized Trial of Utility in Reducing Events (OVERTURE). *Circulation* 2002;106:920–6.
4. Cohn JN, Goldstein SO, Greenberg BH, Lorell BH, Bourge RC, Jaski BE, et al. A dose-dependent increase in mortality with vesnarinone among patients with severe heart failure. Vesnarinone Trial Investigators. *N Engl J Med* 1998;339:1810–6.
5. Feldman AM, Bristow MR, Parmley WW, Carson PE, Pepine CJ, Gilbert EM, et al. Effects of vesnarinone on morbidity and mortality in patients with heart failure. Vesnarinone Study Group. *N Engl J Med* 1993;329:149–55.
6. Packer M, Swedberg K, McMurray JJV, et al. , for the REACH-1 Investigators Multicentre, double-blind, placebo-controlled study of long-term endothelin blockade with bosentan in chronic heart failure: results of the Reach Trial. (abstract). *Eur Heart J* 1999;20(Suppl):95.
7. Bozkurt B, Torre-Amione G, Warren MS, Whitmore J, Soran OZ, Feldman AM, et al. Results of targeted anti-tumor necrosis factor therapy with etanercept (ENBREL) in patients with advanced heart failure. *Circulation* 2001;103:1044–7.
8. Rouleau JL, Pfeffer MA, Stewart DJ, Isaac D, Sestier F, Kerut EK, et al. Comparison of vasopeptidase inhibitor, omapatrilat, and lisinopril on exercise tolerance and morbidity in patients with heart failure: IMPRESS randomised trial. *Lancet* 2000;356:615–20.
9. Pitt B, Segal R, Martinez FA, Meurers G, Cowley AJ, Thomas I, et al. Randomised trial of losartan versus captopril in patients over 65 with heart failure (Evaluation of Losartan in the Elderly Study, ELITE). *Lancet* 1997;349:747–52.
10. Pitt B, Poole-Wilson PA, Segal R, Martinez FA, Dickstein K, Camm AJ, et al. Effect of losartan compared with captopril on mortality in patients with symptomatic heart failure: randomised trial—the Losartan Heart Failure Survival Study ELITE II. *Lancet* 2000;355:1582–7.
11. Packer M, O'Connor CM, Ghali JK, Pressler ML, Carson PE, Belkin RN, et al. Effect of amlodipine on morbidity and mortality in severe chronic heart failure. Prospective Randomized Amlodipine Survival Evaluation Study Group. *N Engl J Med* 1996;335:1107–14.
12. Thackray S, Witte K, Clark AL, Cleland JG. Clinical trials update: OPTIME-CHF, PRAISE-2, ALL-HAT. *Eur J Heart Fail* 2000;2:209–12.
13. Gottlieb SS, Kukin ML, Penn J, Fisher ML, Cines M, Medina N, et al. Sustained hemodynamic response to flosequinan in patients with heart failure receiving angiotensin-converting enzyme inhibitors. *J Am Coll Cardiol* 1993;22:963–7.
14. Packer M, Colucci WS, Sackner-Bernstein JD, Liang CS, Goldscher DA, Freeman I, et al. Double-blind, placebo-controlled study of the effects of carvedilol in patients with moderate to severe heart failure. The PRECISE Trial. Prospective Randomized Evaluation of Carvedilol on Symptoms and Exercise. *Circulation* 1996;94:2793–9.
15. Packer M, Carver JR, Rodeheffer RJ, Ivanhoe RJ, DiBianco R, Zeldis SM, et al. Effect of oral milrinone on mortality in severe chronic heart failure. The PROMISE Study Research Group. *N Engl J Med* 1991;325:1468–75.
16. The Beta-Blocker Evaluation of Survival Trial Investigators . A trial of the beta-blocker bucindolol in patients with advanced chronic heart failure. *N Engl J Med* 2001;344:1659–67.
17. Wedel H, Demets D, Deedwania P, Fagerberg B, Goldstein S, Gottlieb S, et al. Challenges of subgroup analyses in multinational clinical trials: experiences from the MERIT-HF trial. *Am Heart J* 2001;142:502–11.
18. Packer M. Current perspectives on the design of phase II trials of new drugs for the treatment of heart failure. *Am Heart J* 2000;139:S202–6.