Meta-Analysis of the Effects of Endothelin Receptor Blockade on Survival in Experimental Heart Failure

DOUGLAS S. LEE, MD,1,2,3,4 QUANG T. NGUYEN, PhD,3 NATHALIE LAPIORIS, MSc,3 PETER C. AUSTIN, PhD,4 ARNE OHLSSON, MD, MSc,1,2,5,6 JACK V. TU, MD, PhD,1,2,4 DUNCAN J. STEWART, MD,1,6 JEAN L. ROULEAU, MD1,3

Toronto, Ontario, Canada; Hamilton, Ontario, Canada

ABSTRACT

Background: Although an initial study of endothelin receptor blockade reported positive findings, subsequent experiments and clinical trials in humans found little or no benefit.

Methods: We applied meta-analytic methods to assess the methodologic rigor of preclinical studies of endothelin blockade and to quantitatively evaluate the totality of evidence regarding the effect of endothelin receptor blockers in experimental heart failure. A total of 396 animals were assigned to control and 594 were assigned to experimental therapy in the pooled analysis. Of the 9 studies identified, no study reported a priori sample size justification. Although there was a tendency to increased mortality with early administration (relative risk 1.39, \( P = .15 \)) and decreased mortality with late administration (relative risk 0.85, \( P = .6 \)), in the overall analysis, there was no significant evidence of benefit or harm (relative risk 1.03, \( P = .9 \)). Studies with a small sample size had estimated effects that tended to deviate further from the pooled estimate of all studies.

Conclusions: Consideration of mortality effects in the totality of studies revealed no significant effect of endothelin antagonists in animal models of experimental heart failure. Given the potential for between-study variability, reliance on studies with small sample size may lead to unrealistic expectations when extrapolating preclinical experimental results to future research.

Key Words: Heart failure, myocardial infarction, meta-analysis, clinical trials, statistics, mortality, prediction.

Observational studies demonstrated that endothelin-1 was increased in patients with myocardial infarction\(^1\) and heart failure,\(^2\)-\(^4\) and was associated with a poor prognosis.\(^5\),\(^6\) Given its pathophysiologic role in promoting vasoconstriction and cardiac remodeling, endothelin receptor blockade represented a biologically plausible therapeutic strategy for heart failure, especially in view of the dramatic successes with inhibitors of other vasoconstrictive neurohormones (eg, angiotensin II, catecholamines).

When initial studies of endothelin blockers in animal models demonstrated marked beneficial effects in delaying adverse left ventricular remodeling after myocardial infarction and in improving survival,\(^7\),\(^8\) an array of endothelin receptor blockers were developed and clinical trials were initiated. However, in contrast to the initially positive animal model studies, results from clinical trials in humans suggested at best a neutral effect (ENABLE study, 2002).\(^9\) Because the use of large numbers of animals necessary for a survival study may not be justified ethically, the effect of an intervention on mortality may require extrapolation from results of studies that may not be primarily designed to establish survival differences.\(^10\) Therefore, insufficient numbers of animals are often employed. The sample size limitations associated with extrapolating results of survival from small experimental studies have, until now, been considered an inevitable limitation of such designs.

Meta-analysis is an approach that has been widely employed in clinical research, to provide a pooled estimate of
Effect. In this paper, we applied meta-analytic methods to quantitatively evaluate studies of endothelin receptor blockade in experimental heart failure. We hypothesized that consideration of the sample size of preclinical studies and statistical evaluation of the totality of evidence would better reflect the effect of the intervention.

Methods

Study Identification

We identified studies using endothelin antagonists in animal models of myocardial infarction and heart failure in the Medline and Embase electronic databases from 1988 to 2002. Studies comparing pharmacologic endothelin blockade with control in animal models were identified. No restrictions to the type of agent were applied. Therefore, both selective and nonselective endothelin antagonists were eligible interventions regardless of dose or time of administration. The analysis was limited to studies of treatment duration that exceeded at least 1 day. The primary outcome of interest was all-cause mortality; therefore, studies were required to have abstractable survival data to merit inclusion. Reference lists from retrieved articles were searched for additional potentially relevant studies. No language restrictions were applied.

Quality Assessment and Data Abstraction

We abstracted data on the duration of administration, timing of drug delivery, and agent employed from the included studies. Outcome data included the number of animals in both the treatment and the control arms and the number of deaths from all causes. Two reviewers (DL, QN) independently abstracted data from each study that met entry criteria, using a data abstraction tool developed for this purpose. Quality assessment guidelines for this study were adapted from the clinical review guidelines of the Cochrane Collaborative Review Group and included the following: (1) randomization, (2) blinding of intervention, (3) completeness of follow-up, and (4) blinding/objectivity of outcome measurement. Any potential disagreement on the abstracted data or quality assessment was arbitrated by 2 of the authors (JR, DS).

Statistical Analysis

We pooled the studies using the Mantel-Haenszel method, stratifying by time of drug administration. The relative risk (RR) and 95% confidence intervals (CI) were initially calculated using the fixed-effects model. Heterogeneity between studies was identified using the chi-square statistic and a P value of less than .10 was considered to represent significant heterogeneity between studies. If there was between-study heterogeneity, the results of both the fixed effects and the random effects model analyses were reported. Analyses were conducted using Review Manager 4.1 (Cochrane Collaboration, Bristol, UK). We conducted stratified analyses of early versus late administration, as defined by the timing of the initiation of therapy (≤3 versus >3 days), and also pooled studies using selective or nonselective endothelin antagonists. In the latter analysis, studies using bosentan were included in both the selective and nonselective receptor antagonist analyses because, in the doses employed, it nonspecifically antagonizes the receptor, but nevertheless has greater affinity for endothelin-A (ET\textsubscript{A}) as compared with ET\textsubscript{B} receptors.

Results

Study Selection and Quality Assessment

Of 16 potential studies, 9 reported on mortality and these results were pooled. In 6 studies, the experimental group received a selective ET\textsubscript{A} antagonist: BQ-123, LU-135252, A-127722, or sitaxsentan. In 4 studies, the experimental therapy was a nonselective antagonist: SB-209670, LU-420627, TAK-044, or bosentan. In 1 study, selective and nonselective antagonists were compared with a similar set of controls and, therefore, the experimental interventions were analyzed as a unit. Another study evaluated a single bolus intervention and was therefore excluded from the analysis. The studies pooled in the analysis are shown in Table 1. There were between-study differences in the animal models employed and variations in drug administration protocols. Although the experimental studies had complete follow-up and reported objective mortality endpoints, the use of randomization and blinding of the intervention to minimize potential biases was either inconsistent or not reported explicitly (see Table 1). In addition, the sample sizes of the studies varied considerably, yet none reported a priori power or sample size calculations.

Synthesis of Study Results

In total, there were 396 animals assigned to control and 594 assigned to experimental treatment. In the fixed effects model, there was no overall effect on mortality with a RR of 1.07 (P = 0.4) and 95% confidence limits that crossed unity (0.92, 1.24). However, there was a significant degree of heterogeneity (chi square 21.2, P = .004); therefore, a random effects model was used to pool studies. The random effects analysis resulted in no overall effect with experimental therapy, with a RR of 1.03 and 95% CI 0.74, 1.44 (P = 9).

Figure 1 shows all studies ordered according to increasing relative study weight, which reflects the size and number of events in the included studies. Of the 5 low-weight studies in the random effects analysis, 3 had point estimates that suggested a tendency toward benefit with endothelin antagonism. The RRs of death were 0.32, 0.36, and 0.88 in the random effects analysis, 3 had point estimates that exceeded at least 1 day. The primary outcome of interest was all-cause mortality; therefore, studies were required to have abstractable survival data to merit inclusion. Reference lists from retrieved articles were searched for additional potentially relevant studies. No language restrictions were applied.
Table 1. Description of Included Studies and Clinical Quality Assessment Criteria

<table>
<thead>
<tr>
<th>Antagonist/Author</th>
<th>Animal Model</th>
<th>Total Sample Size</th>
<th>Initiation Post-MI</th>
<th>Therapy Duration</th>
<th>Length of Follow-Up</th>
<th>Randomization</th>
<th>Blinding of Intervention</th>
<th>Complete Follow-up</th>
<th>Blinding Objective Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQ123*/Sakai7</td>
<td>Male Sprague-Dawley</td>
<td>27</td>
<td>Day 10</td>
<td>12 wk</td>
<td>12.9 wk</td>
<td>None</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Bosentan/Mulder20</td>
<td>Male Wistar</td>
<td>208</td>
<td>Day 7</td>
<td>9 mo</td>
<td>9 mo</td>
<td>Yes</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LU135252*/Mulder14</td>
<td>Male Wistar</td>
<td>60</td>
<td>Day 7</td>
<td>10 wk</td>
<td>10 wk</td>
<td>Yes</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LU135252'/Nguyen13</td>
<td>Male Wistar</td>
<td>296</td>
<td>24 h</td>
<td>4 wk</td>
<td>30 d</td>
<td>Yes</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>BQ123-SB299670'/Sakai14</td>
<td>Male Sprague-Dawley</td>
<td>20</td>
<td>Day 10</td>
<td>12 wk</td>
<td>12.9 wk</td>
<td>None</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mulder20</td>
<td>Male Wistar</td>
<td>64</td>
<td>3 h</td>
<td>6 wk</td>
<td>6 wk</td>
<td>None</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>TAK-044'/Takahashi18</td>
<td>Male Wistar</td>
<td>41</td>
<td>&lt;24 h</td>
<td>3 wk</td>
<td>3 wk</td>
<td>Yes</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sitaxsentan'/Podesser17</td>
<td>Male Wistar</td>
<td>25</td>
<td>Day 3</td>
<td>6 wk</td>
<td>6 wk</td>
<td>Yes</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>LU420627'/Nguyen22</td>
<td>Male Wistar</td>
<td>249</td>
<td>24 h/day 10</td>
<td>14 wk</td>
<td>14 wk</td>
<td>Yes</td>
<td>None</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*Indicates selective ETA antagonism. †Indicates nonselective ETA/ETB antagonism. ‡Bosentan evaluated as both selective and nonselective antagonist.

active treatment arm. In this study, active treatment with endothelin blockade after infarction was harmful, with a relative risk of death of 1.96 (95% CI: 1.31, 2.92).

Stratified Analyses

As shown in Fig. 2, 5 studies initiated pharmacologic endothelin blockade early (within 3 days after infarction) and continued therapy thereafter. The pooled results from these studies suggested that early initiation of therapy tended to increase the risk of death. The RR of death was 1.28 with active treatment (95% CI: 1.07, 1.54; \( P = .007 \)) in the fixed effects analysis, but there was significant heterogeneity (\( P = .02 \)). The random effects model also demonstrated a tendency toward harm (relative risk 1.39; 95% CI: 0.88, 2.19), however, this was not statistically significant (\( P = .15 \)). In studies in which endothelin blockade was initiated later after infarction (>3 days), there was no significant effect on outcome (Fig. 3). The RR was 0.85 with a 95% CI that crossed unity (0.48, 1.52; \( P = .6 \)).

![Fig. 1](https://via.placeholder.com/150)

Fig. 1. All included studies of endothelin antagonism in experimental heart failure listed in order of increasing relative weight. Over all studies, there was no beneficial effect on survival with endothelin antagonism (\( P = .9 \)).
Mortality Analysis in Preclinical Research

Lee et al

Fig. 2. All studies of early administration of endothelin antagonists on survival in rats with experimental heart failure. Analysis showed no statistically significant evidence of benefit but a tendency toward harm; the risk of mortality was 39% higher with endothelin antagonists compared with control (random effects: $P = .15$).

The studies evaluating selective (ET$_{A}$) and nonselective (ET$_{A/N}$) antagonists are shown in Fig. 4 and 5, respectively. In these analyses, the studies of Sakai are separated because selective$^7$ and nonselective$^{14}$ agents were employed. Additionally, the study employing bosentan$^{20}$ was included in both analyses. There was no evidence of heterogeneity ($P = .34$), and the pooled results showed no significant reduction in the risk of death with a relative risk of 0.92 (95% CI: .78, 1.09; $P = .3$). In the random effects analysis, there was a nonsignificant trend to increased mortality effect, with a relative risk of 1.04 (95% CI: .55, 1.97; $P = .9$).

Discussion

Positive findings in animal model studies may provide an impetus for the initiation of clinical studies in humans. Indeed, findings in adequately powered animal studies have led to important advances in our understanding of the pathophysiology of disease and have predicted the response in humans to selected therapeutic interventions.$^{23,24}$ However, despite the usefulness of animal studies, there remains variability in design quality of such work. In the clinical arena, the designs of randomized controlled trials are rigorously scrutinized for bias and validity. Less than 2 decades ago, such rigor was not routinely applied to clinical research, but an appreciation of the impact of trial design on study outcome may have led to a more sound scientific approach. Although animal research forms the basis for applied studies in humans, the same critical analysis of study design has not been widely applied to basic research studies, specifically as they relate to survival. Potential sources of variability between studies evaluating survival in animal models include variations in the mode of pathologic insult, response to therapeutic intervention, and environmental factors. The inherent

<table>
<thead>
<tr>
<th>Study</th>
<th>Endothelin blockade n/N</th>
<th>Control n/N</th>
<th>RR (95% CI Random)</th>
<th>Weight %</th>
<th>RR (95% CI Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Podessor 2001</td>
<td>0/12</td>
<td>1/13</td>
<td>2.0</td>
<td></td>
<td>0.36[0.02,8.05]</td>
</tr>
<tr>
<td>Takahashi 2001</td>
<td>9/22</td>
<td>5/19</td>
<td></td>
<td></td>
<td>1.55[0.63,3.84]</td>
</tr>
<tr>
<td>Pfeffer 2000</td>
<td>10/33</td>
<td>7/31</td>
<td></td>
<td></td>
<td>1.34[0.56,3.08]</td>
</tr>
<tr>
<td>Nguyen 2001</td>
<td>116/138</td>
<td>15/38</td>
<td></td>
<td></td>
<td>2.13[1.43,3.18]</td>
</tr>
<tr>
<td>Nguyen 1998</td>
<td>75/139</td>
<td>83/157</td>
<td></td>
<td></td>
<td>1.02[0.82,1.26]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>210/344</td>
<td>111/258</td>
<td>100.0</td>
<td></td>
<td>1.39[0.88,2.19]</td>
</tr>
</tbody>
</table>

Test for heterogeneity chi-square=11.41 df=4 p=0.022
Test for overall effect z=1.42 p=0.15

Fig. 3. All studies of late administration of endothelin antagonists in experimental heart failure. Data from Nguyen$^{21}$ includes the subset with late administration of endothelin antagonist. There was no demonstrable benefit with late administration ($P = .6$).
Fig. 4. Studies of selective endothelin-A (ET\textsubscript{A}) receptor antagonism versus placebo. The subset of Sakai et al\textsuperscript{7} that evaluated ET\textsubscript{A} receptor blockade and that of Mulder et al\textsuperscript{20} (employing bosentan) are included in this analysis. Overall, there was no significant benefit with selective ET\textsubscript{A} receptor antagonism with relative risk of 0.92 ($P = .3$).

Variability between studies and inadequate consideration of the potential implications of studies with small sample size may inappropriately suggest exaggerated survival benefits, leading to unrealistic expectations in subsequent research.

A key component of the design of randomized controlled trials is the calculation of power and sample size. Studies performed in animal models generally do not report this aspect of study design, but rather tend to employ small numbers. This may result in isolated studies that report an exaggerated treatment effect compared with the totality of evidence, or alternatively, studies that are underpowered to detect a true treatment effect. We pooled data from individual studies to help increase the precision of the estimated effect of experimental endothelin antagonism. When all studies of endothelin antagonism versus placebo were pooled, there was no compelling evidence of a survival advantage with endothelin antagonism. Based on the pooled results, it would be unlikely that embarking on a larger study would find an improvement in survival with endothelin blockade.

Whereas larger studies\textsuperscript{21,22} demonstrated either harm or no effect with active treatment, three of the smaller studies\textsuperscript{14,16,17} tended to show a beneficial outcome. If one were to assume that the overall estimate was reflective of the “true effect,” then reliance on an isolated small sample study might lead to the inaccurate belief that the intervention was beneficial\textsuperscript{25,26}. This is in contrast to the overall effect from the totality of studies, which might be viewed as truly regressed to the null. The problems inherent in the reliance upon research studies of small sample size are further exacerbated by the potential for publication bias and a tendency to favor the publication of positive reports in comparison with small studies that report neutral or negative results\textsuperscript{27}.

Fig. 5. Studies of nonselective ET\textsubscript{A/B} antagonism (including bosentan) versus placebo. Pooled analysis revealed no significant evidence of benefit with nonselective endothelin antagonism with relative risk of 1.04 ($P = .9$).
similarities, common environmental exposures, techniques of disease injury, and drug administration may reduce within-study variability. Thus between-study variations may be large relative to within-study variations (increased $p$) when similar but nonidentical experiments are conducted in multiple laboratories. Therefore, given the potential for variability between studies, it is important to consider the totality of evidence from multiple studies before embarking on the subsequent stage of randomized clinical trials in humans. We propose that meta-analysis of studies conducted in animal models may be a potential method to quantitatively synthesize the findings from multiple experiments.

In addition, we maintain that sample size and power should be key considerations in the design of basic research studies where it is the intention to report outcomes. Based on the observed overall control event rate of 50% mortality, a 2-tailed type I error rate of 0.05 and power of 0.80, we would estimate that a study would require 58 animals per group to detect as large a difference as a 50% reduction in events. Larger samples of 170 and 246 animals per group would be required to detect smaller effect sizes of 30% and 25% reduction in event rates, respectively. None of the small studies would have satisfied the sample size criteria to detect a true effect size as large as 50%. Additionally, high beta error rates (eg, failure to reject the null hypothesis when a true difference exists) from small studies with inconclusive findings would raise the issue of whether a true beneficial or harmful effect was not detected due to lack of power. A power analysis revealed that only 20-22 of the 9 studies had sufficient power to detect the 70% reduction of events that was initially reported. Clearly, if the true effect size was reduced, there would be even less power to detect such an effect with the numbers of animals employed in most studies. This form of analysis is dependent on the design and quality of the studies, which comprise the overall analysis. For example, inclusion of studies with a short duration of therapy or follow-up times may not detect a long-term benefit. A further limitation is that the studies may have been designed primarily for the purpose of reporting a physiologic effect, and not mortality outcomes. However, even if the primary objective was a physiologic endpoint and not survival, it is important to report survival results because a lack of consistency between the surrogate outcome and survival may have important implications for future research.28-32

Previous reports have demonstrated that meta-analyses are concordant with large randomized controlled trials and that meta-analyses may predict the outcome of such trials.33 Although the true effect in animal models may differ from the true effect in humans, we believe that careful and systematic review of animal model studies of sufficient power may help researchers in their decision making about future research directions. Furthermore, the analytic methods of meta-analysis may help to identify sources of heterogeneity that may have biologic implications when experimental therapies are applied to humans. The potential explanations for differences between an individual study and the overall results are speculative, but may include timing of drug administration, variability in extent of myocardial infarction, specificity of ET$_{AB}$ blockade, and design considerations (such as randomization and blinding). Evaluation of the effects of these differences did not, however, find 1 overriding explanation for the differences found among the studies (data not shown).

In summary, systematic and quantitative synthesis of preclinical pharmacologic intervention studies may have utility in the assessment of novel therapeutic agents. In addition, quantitative analysis of results from multiple investigations may lead to estimates of overall effect, which may differ substantially from the results of isolated small studies. Our results suggest that caution should be used against over-interpretation of small, early studies of survival benefit. Furthermore, pooled analysis may supplement the physiologic data assessed in preclinical research. Using this methodology, we found no evidence of a significant survival benefit of endothelin receptor blockade in preclinical studies of experimental heart failure. This mode of analysis may have applicability in the preclinical evaluation of mortality effects of other pharmacologic agents, and warrants further study. Future efforts in quantitative synthesis of preclinical studies will be enhanced by the performance of multiple studies with comparable experimental study designs that merit inclusion in a pooled analysis.

References