Quantitative Evaluation of Drug or Device Effects on Ventricular Remodeling as Predictors of Therapeutic Effects on Mortality in Patients With Heart Failure and Reduced Ejection Fraction

A Meta-Analytic Approach

Daniel G. Kramer, MD,* Thomas A. Trikalinos, MD, PtlD,† David M. Kent, MD, MS,† George V. Antonopoulos, MD,* Marvin A. Konstam, MD,* James E. Udelson, MD*

Boston, Massachusetts

Objectives
The purpose of this study was to quantitatively assess the relationship between therapy-induced changes in left ventricular (LV) remodeling and longer-term outcomes in patients with left ventricular dysfunction (LVD).

Background
Whether therapy-induced changes in left ventricular ejection fraction (LVEF), end-diastolic volume (EDV), and end-systolic volume (ESV) are predictors of mortality in patients with LVD is not established.

Methods
Searches for randomized controlled trials (RCTs) were conducted to identify drug or device therapies for which an effect on mortality in patients with LVD was studied in at least 1 RCT of ≥500 patients (mortality trials). Then, all RCTs involving those therapies were identified in patients with LVD that described changes in LVEF and/or volumes over time (remodeling trials). We examined whether the magnitude of remodeling effects is associated with the odds ratios for death across all therapies or associated with whether the odds ratio for mortality was favorable, neutral, or adverse (i.e., statistically significantly decreased, nonsignificant, or statistically significantly increased odds for mortality, respectively).

Results
Included were 30 mortality trials of 25 drug/device therapies (n = 69,766 patients; median follow-up 17 months) and 88 remodeling trials of the same therapies (n = 19,921 patients; median follow-up 6 months). The odds ratio for death in the mortality trials was correlated with drug/device effects on LVEF (r = 0.51, p = 0.001), EDV (r = 0.44, p = 0.002), and ESV (r = 0.48, p = 0.002). In (ordinal) logistic regressions, the odds for neutral or favorable effects in the mortality RCTs increased with mean increases in LVEF and with mean decreases in EDV and ESV in the remodeling trials.

Conclusions
In patients with LVD, short-term trial-level therapeutic effects of a drug or device on LV remodeling are associated with longer-term trial-level effects on mortality. (J Am Coll Cardiol 2010;56:392–406) © 2010 by the American College of Cardiology Foundation

While the past 2 decades have seen important advances in therapies for heart failure (HF) (1), there have also been some promising agents—endothelin antagonists (2,3), cytokine inhibitors (4), and vasopeptidase inhibitors (5,6)—developed through phase 3 clinical testing only to yield negative or neutral results. Because phase 3 trials are by far the most costly and time-consuming phase of drug development, minimization of potential negative or null results is important for the development of new therapies. Thus, it would be very helpful to obtain an early signal of clinical efficacy in the context of shorter, smaller phase 2 trials.

Assessment of ventricular remodeling (i.e., characteristic changes in ventricular volume and wall thickness and shape) is often referred to as a potential surrogate end point for drug or device effects on HF outcomes (1,7). Left ventricular end-diastolic volume (EDV) and end-systolic volume (ESV) (8), dimensions (9,10), and left ventricular ejection fraction (LVEF) (11–14) are each prognostic when measured at 1 point in time for subsequent mortality risk.

From the *Division of Cardiology and the CardioVascular Center, Tufts Medical Center, and the †Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, Massachusetts. This article was partially funded by the grant NIH/NCRR UL1RR025752, Tufts Clinical and Translational Science Institute. Dr. Konstam is a consultant for Boehringer-Ingelheim, Trevena, Johnson & Johnson, Cardioxyl, Merck, and Forest. Dr. Udelson is a consultant for Cytori, Angioblast, and Medtronic.

Manuscript received December 28, 2009; revised manuscript received May 3, 2010, accepted May 18, 2010.
Moreover, data from some HF trials of individual therapeutic agents suggest a relation between a drug or device effect on remodeling and the therapeutic effect on natural history outcome (9,10,15). How well therapy-induced changes in these parameters predict therapeutic benefit in mortality outcomes, independent of an individual therapy, has not been quantitatively assessed. Herein, we systematically evaluate the degree to which therapy-induced changes in 3 measures often assessed in remodeling studies (LVEF, EDV, and ESV) are associated with therapeutic effect on mortality outcomes in phase 3 clinical trials in patients with HF and left ventricular dysfunction (LVD).

Methods

General approach. Ideally, the assessment of the relation between the effect of a therapy on remodeling and its effect on mortality would be evaluated in large adequately powered outcome trials, in which all of the patients also had early assessment of remodeling by noninvasive imaging. However, given the expense and complexity of imaging in such a setting, very few trials have included measures of remodeling in all patients, with some exceptions (3,16–19). More often, remodeling is assessed in a substudy population selected from the overall population sample of a trial (20–25) or hypotheses are formed based on the results of outcome studies along with smaller remodeling studies from distinct samples of patients (26). We initially identified from the literature drugs or devices for HF patients that were studied in large randomized controlled trials (RCTs) evaluating mortality. Then, we systematically identified from the published literature effects of those drugs and devices on remodeling parameters from imaging studies, and examined associations between trial-level (mean) changes in the remodeling outcome and effects on mortality with the same drug or device.

Identification of eligible interventions. We first performed a systematic literature search to identify all drug and device therapies for patients with LVD, for which mortality was evaluated in at least 1 large placebo-controlled RCT with adequate follow-up. For all treatments identified in this set of RCTs (subsequently referred to as “mortality” RCTs), we systematically identified all published placebo-controlled RCTs describing the effects of those treatments on parameters of LV remodeling (subsequently referred to as “remodeling” trials).

Search strategy. We performed several incremental and overlapping MEDLINE searches (covering January 1966 through April 16, 2007) using the keywords “heart failure” and “double-blind” and “placebo.” We first identified qualifying mortality trials that identified the interventions for which a search for remodeling trials was required. The search strategy for the remodeling trials was performed in 2 steps (1966 to 1999 [27] and January 1, 1999, to April 16, 2007). We limited searches to English language peer-reviewed publications on human subjects. Double-blind analysis was a requisite for all studies, except for the device trials.

Thorough examination of citation lists from all retrieved studies, meta-analyses and review articles was conducted to identify additional relevant publications. We complemented searches with input from field experts. Only published studies in peer-reviewed journals were included. Care was taken to include only the most recent or most complete data in the case of overlapping and duplicated data sets. Data from review articles, case reports, abstracts, reports of trial presentations at conferences, and data from letters were not included.

Eligibility criteria. All included trials were English language, double-blind (except device trials), randomized, placebo-controlled trials on human subjects with LVEF ≤45%. For mortality trials, we prospectively required at least 500 patients and at least 6 months of follow-up for eligibility. For remodeling trials, we required the measurement of at least 1 measure of ventricular remodeling (LVEF, volume, or dimension) in active drug and placebo groups over a period of at least 4 weeks, an arbitrary cut-point meant to eliminate very short-term studies of acute effects. There was no minimum sample size for the remodeling trials.

Data extraction. A single investigator (D.G.K.) extracted data on pre-constructed paper forms, receiving input from an experienced methodologist when needed. Information on the publication, the active intervention, and patient characteristics from each trial were extracted. Mean follow-up duration and the number of patients enrolled and analyzed per study arm were recorded.

For each arm in the mortality trials, the number of deaths from all causes was extracted. For the remodeling trials, we recorded information necessary to calculate the mean net difference in LVEF, EDV, ESV, or dimensions over time across the compared randomized arms (28). Typically, this involved recording the means and standard deviations of the pertinent measures before and after treatment.

Calculation of trial-level (mean) effects on mortality and remodeling outcomes. For each intervention, we calculated the odds ratio (OR) for long-term all-cause mortality. When multiple studies existed, we derived summary effects using standard meta-analysis, the Mantel-Haenszel method for 2 studies (29) and the DerSimonian and Laird random effects method for ≥3 studies (30). We tested for heterogeneity with the Q statistic (considered significant for p < 0.10) (31) and calculated its extent with P (32). The P

Abbreviations and Acronyms

AUC = area under the curve
CI = confidence interval
EDV = end-diastolic volume
ESV = end-systolic volume
HF = heart failure
IQR = interquartile range
LVD = left ventricular dysfunction
LVEF = left ventricular ejection fraction
NYHA = New York Heart Association
OR = odds ratio
RCT = randomized controlled trial
ROC = receiver-operating characteristic

JACC Vol. 56, No. 5, 2010
July 27, 2010:392–406
Kramer et al.
Ventricular Remodeling and Effects on Mortality
expresses the proportion of between-study variability that is attributed to heterogeneity rather than chance, with values >50% indicating high heterogeneity (33).

From each remodeling trial, we calculated the mean net difference of the change in the LVEF, EDV, and ESV over time between intervention and placebo groups. As described in the following text under “Correlation and regression analyses,” main analyses considered effects from the remodeling trials separately. To be inclusive of as much extant published data as possible, we prospectively elected to also include data from studies that published results of LV dimensions instead of calculated LV volumes. For remodeling trials that reported only LV dimensions, we transformed mean changes (and variances thereof) in LV dimensions to the corresponding mean changes (and variances) of LV volumes, by using the delta method (34) and the Teicholtz formula (35). Similarly, mean changes (and variances) in LV ESV and EDV indices were transformed to LV volumes using the Mosteller formula for the body surface area (36) and assuming a height of 1.70 m and a weight of 75 kg. The latter are approximate of the median sex-averaged values in the U.S. population aged similarly to the mean ages in our eligible studies (37). Results were not sensitive to small perturbations in these somatometric assumptions (data not shown). We assessed the validity of the transformations in 5 studies reporting 8 pairs of mean changes both in diameters/indices and in volumes. Discrepancies between the estimated (transformed) and actual mean changes were >10 ml in 4 measurements (maximum discrepancy approximately 25 ml in EDV). The transformed values were less extreme than the actual values in 7 of 8 pairs and always smaller than 10 ml. This finding suggests that the transformations bias findings toward the null.

Correlation and regression analyses. For each of the 3 remodeling outcomes (LVEF, EDV, and ESV), main analyses were performed in 3 steps. All main analyses were unadjusted and considered remodeling effects from trials on the same intervention separately. First, we calculated unweighted Spearman correlations between the trial-level remodeling effects and the corresponding summary OR for mortality. These analyses are largely descriptive as they do not account for the statistical errors that accompany the mortality and remodeling effect estimates.

Second, we explored whether there is a simple linear relationship between the log-transformed OR for mortality and the magnitude of the remodeling effects using a general linear model that accounts for heteroskedasticity in the dependent variable (inverse variance weighting) and the clustering of observations by intervention (with robust standard error estimation). This analysis yields the relative OR for mortality per given change in the remodeling parameter.

Third, we examined with logistic and ordinal logistic regressions whether remodeling outcomes can predict whether the OR for mortality was favorable, neutral, or adverse (i.e., statistically significantly decreased, nonsignificant, or statistically significantly increased odds for mortality, respectively) with the corresponding intervention. An intervention was considered to have a favorable mortality effect if the upper bound of the 95% confidence interval (CI) of the OR was <1, a neutral mortality effect if the 95% CI of the OR included 1, and an adverse mortality effect if the lower bound of the 95% CI of the OR was >1. Again, we used inverse variance weighting and accounted for clustering of observations by intervention with robust standard error estimation. This alternative analysis categorizes the effect on mortality into 3 ordinal categories, and informs on the ability of short-term remodeling outcomes to predict effects on mortality in larger trials.

For illustrative purposes, we plotted predicted probabilities for favorable, neutral, or adverse effects on mortality for different magnitudes of remodeling effects, and produced receiver-operating characteristic (ROC) curves. The area under the ROC curve (AUC) is a simple measure of the discriminatory ability of the mean change in the remodeling outcomes to predict mortality. It represents the percentage of all possible discordant pairs of interventions discordant for their mortality effects in which the variable (change in LVEF, EDV, or ESV, respectively) and correctly assigns a higher probability of mortality benefit to the intervention that actually demonstrated benefit. An AUC value of 0.5 implies no discriminatory ability, and an AUC value of 1.0 implies perfect discrimination.

Sensitivity analyses. In sensitivity analyses, regressions were adjusted for the mean value of the remodeling outcome at baseline and for follow-up duration. We also repeated all regression analyses using unweighted regressions, and after eliminating the shorter-term remodeling studies (<12 weeks of follow-up). We performed main analyses excluding studies where we calculated values of changes in ESV and EDV from results on diameters or indices. Finally, additional sensitivity analyses used the meta-analysis-derived summary effect size of all remodeling trials within each drug/device intervention, rather than considering remodeling trials separately (as was done in the main analyses).

All analyses and graphs were performed in Stata SE version 11 (Stata Corp., College Station, Texas) and in Meta-Analyst (Tufts Evidence-based Practice Center, 2009). Unless otherwise stated, all p values are 2-tailed and considered significant at the 0.05 level.

Results

Results of search algorithms. The search algorithm yielded 1,992 citations, and 248 articles were retrieved and reviewed in full text. In total, 117 nonoverlapping RCT reports were eligible. We identified 30 large RCTs describing the effects of 25 different interventions on mortality (Fig. 1A) that reported on a total of 69,766 patients over a median follow-up of 17 months. For each individual drug/device intervention for which a mortality trial had been identified, we identified between 1 and 22 remodeling trials.
There were 88 remodeling RCTs evaluating 91 distinct drug or device versus placebo comparisons, involving 19,921 patients with median follow-up of 6 months (Fig. 1B). Four of the 30 mortality trials reported serial measurements of remodeling parameters of all patients and, therefore, also qualified as remodeling trials (16,17,19,38), whereas 11 other remodeling trials were substudies of 1 of the mortality RCTs (20–22,25,39–45). To assess remodeling, 39 trials employed radionuclide ventriculography, 32 utilized 2-dimensional echocardiography, and 3 utilized magnetic resonance imaging. Fourteen studies used a combination of radionuclide ventriculography and 2-dimensional echocardiography.

**Effect on drug or device therapies on mortality.** Table 1 describes the effects of the 25 distinct drug/device therapies on mortality (4,16–19,44,46–67). Nine, 12, and 4 interventions had significantly favorable, neutral, and significantly adverse effects on mortality, respectively, in the identified trials. The mortality trials included predominantly male patients (81% on average) with mean ages between 57 and 67 years (median age 63 years). Mean New York Heart Association (NYHA) HF functional class was 2.7 (interquartile range [IQR] 0.7). The median number of patients per individual drug/device intervention was 2,345 (IQR 2,328). The 30 trials followed up patients for a median of 17 months. Two (6%) of the trials evaluated patients with LVD after an acute myocardial infarction; the remaining 29 (94%) trials evaluated patients with chronic HF and LVD.

Of note, we chose to evaluate the use of enalapril to treat asymptomatic LV dysfunction as a separate entity, as this was the only trial in which all of the patients studied had NYHA functional class I HF (57).

**Effect on drug or device therapies on remodeling.** We identified 88 remodeling RCTs that included a total of 19,741 patients (median 53 patients per trial; IQR 133) in which the effect of the 25 drug or device therapies (identified from the mortality trials) on remodeling was evaluated. The majority (n = 85 trials) were parallel arm trials, whereas the remaining 3 trials examined 2 separate drugs versus a single placebo arm. The remodeling trials included predominantly male patients (>57% in all trials), with mean ages between 45 and 71 years (median 60 years). Mean NYHA HF functional class was 2.5 (IQR 0.7). Nine (10%) of the trials evaluated patients with LVD after an acute myocardial infarction; the remaining 82 (90%) trials evaluated patients...
with chronic HF. While the average follow-up was 6 months; 81 trials (89%) had a follow-up of ≥12 weeks.

Mean changes in the LVEF were extracted from 86 trials. Mean changes in EDV and ESV were extracted from only 14 studies; we calculated mean changes in the EDV and ESV from corresponding changes in diameters or indices in 35 of 49 studies and 26 of 40 studies reporting pertinent data, respectively. Tables 2, 3, and 4 describe the summary net changes in LVEF, EDV, and ESV for the 25 eligible interventions compared with placebo (16,17,19–25,38–45,49,68–139). For several interventions, between-study heterogeneity was extensive.

**Change in LVEF and long-term mortality.** Placebo-corrected change in LVEF from each individual remodeling trial was plotted against the mortality OR for the specific therapy. There was a significant correlation between effect sizes in the mortality trials and mean therapy-induced changes in EDV (r = 0.44, p = 0.002) (Fig. 3A). A decrease of 10 ml in the mean change in EDV corresponded to a relative OR of 0.95 for mortality (95% CI: 0.94 to 0.97, p < 0.001).

A decrease of 10 ml in the mean change of EDV was associated with 1.9-fold (95% CI: 1.2 to 3.2, p = 0.012) increased odds that an intervention would have significantly favorable effects on mortality. Figure 3B shows the predicted probability for favorable effects on mortality according to an intervention’s average effect on EDV. In ROC analysis, the AUC for a net change in LV EDV to distinguish favorable from neutral or adverse outcome effects of therapies was 0.76.

**Change in LV ESV and long-term mortality.** Placebo-corrected change in ESV from each individual remodeling trial was plotted against the mortality OR for the specific therapy. There was a significant correlation between effect sizes in the mortality trials and the effect sizes on ESV in the remodeling studies (r = 0.48, p = 0.002) (Fig. 4A). A decrease of 10 l in the mean ESV change corresponded to a relative OR of 0.96 for mortality (95% CI: 0.93 to 0.98, p = 0.01).

Logistic regression analyses suggested a significant association between the mean change in ESV due to therapy and effects on mortality, based on a per 10-ml decrease in the mean change in ESV (OR: 0.56, 95% CI: 0.53 to 0.60). Figure 4B shows the predicted probability for favorable effects on mortality according to an intervention’s average effect on ESV. In ROC analysis, the AUC to distinguish favorable from neutral or adverse therapies was 0.73. It is noted from Table 4 that there are fewer published data on drug/device effects on ESV compared to data on LVEF or EDV, particularly for interventions that were found to have neutral or adverse mortality effects. For this reason, there was insufficient remodeling data for ESV to model the probability for neutral or adverse outcomes individually.

**Secondary analyses.** Adjustments for the average LVEF, EDV, or ESV at baseline (as applicable in the corresponding analyses) and follow-up duration in the remodeling trials resulted in very similar estimates as the main analysis (data not shown). The same was true for the unweighted versions of the regression analyses. Eliminating the shorter-

---

### Table 1: Drug/Device Effects of Mortality Compared With Placebo in Patients With Heart Failure and LVD

<table>
<thead>
<tr>
<th>Intervention (Ref. #)</th>
<th>No. of Studies (n)</th>
<th>OR† (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amiodarone (17)</td>
<td>1 (674)</td>
<td>0.87 (0.64–1.19)</td>
</tr>
<tr>
<td>Amlodipine (46)</td>
<td>1 (1,153)</td>
<td>0.80 (0.63–1.02)</td>
</tr>
<tr>
<td>Bucindorol (19)</td>
<td>1 (2,708)</td>
<td>0.88 (0.75–1.03)</td>
</tr>
<tr>
<td>Bisoprolol (47)</td>
<td>1 (2,647)</td>
<td>0.64 (0.51–0.79)</td>
</tr>
<tr>
<td>CRT (48)</td>
<td>2 (1,738)</td>
<td>0.69 (0.51–0.94)</td>
</tr>
<tr>
<td>Candesartan (49,50)</td>
<td>2 (4,576)</td>
<td>0.83 (0.73–0.95)</td>
</tr>
<tr>
<td>Captopril (51)</td>
<td>1 (2,231)</td>
<td>0.79 (0.64–0.96)</td>
</tr>
<tr>
<td>Carvedilol (52–54)</td>
<td>3 (5,342)</td>
<td>0.62 (0.47–0.81)</td>
</tr>
<tr>
<td>Digoxin (55)</td>
<td>1 (6,800)</td>
<td>0.99 (0.89–1.09)</td>
</tr>
<tr>
<td>Enalapril (56)</td>
<td>1 (2,569)</td>
<td>0.82 (0.70–0.97)</td>
</tr>
<tr>
<td>Enalapril-Prev (57)*</td>
<td>1 (4,228)</td>
<td>0.93 (0.79–1.10)</td>
</tr>
<tr>
<td>Enoximone (58)</td>
<td>1 (1,854)</td>
<td>0.95 (0.76–1.18)</td>
</tr>
<tr>
<td>Etanercept (4)</td>
<td>1 (2,356)</td>
<td>1.22 (0.93–1.61)</td>
</tr>
<tr>
<td>Felodipine (59)</td>
<td>1 (450)</td>
<td>1.07 (0.62–1.84)</td>
</tr>
<tr>
<td>Flosequinan (60)</td>
<td>1 (2,345)</td>
<td>1.42 (1.15–1.76)</td>
</tr>
<tr>
<td>Hydralazine-ISO (16,61)</td>
<td>2 (1,509)</td>
<td>0.70 (0.51–0.96)</td>
</tr>
<tr>
<td>Ibopamine (62)</td>
<td>1 (1,906)</td>
<td>1.27 (1.02–1.57)</td>
</tr>
<tr>
<td>Metoprolor CR (63)</td>
<td>1 (3,991)</td>
<td>0.67 (0.54–0.83)</td>
</tr>
<tr>
<td>Mibefradil (44)</td>
<td>1 (2,571)</td>
<td>1.14 (0.96–1.36)</td>
</tr>
<tr>
<td>Milrinone (64)</td>
<td>1 (1,088)</td>
<td>1.35 (1.03–1.76)</td>
</tr>
<tr>
<td>Moxonidine (65)</td>
<td>1 (1,934)</td>
<td>1.64 (1.05–2.57)</td>
</tr>
<tr>
<td>Prazosin (16)</td>
<td>1 (456)</td>
<td>1.26 (0.87–1.84)</td>
</tr>
<tr>
<td>Spironolactone (66)</td>
<td>1 (1,663)</td>
<td>0.62 (0.51–0.76)</td>
</tr>
<tr>
<td>Telvaptan (67)</td>
<td>1 (4,134)</td>
<td>0.98 (0.85–1.12)</td>
</tr>
<tr>
<td>VALsartan (18)</td>
<td>1 (5,010)</td>
<td>1.02 (0.89–1.18)</td>
</tr>
</tbody>
</table>

*Enalapril was examined separately when studied in patients with asymptomatic left ventricular dysfunction (LVD), as this was the only trial which exclusively examined patients with asymptomatic LVD. †When ≥1 trial existed for a specific therapy, meta-analysis was used to calculate the mortality odds ratio (OR).
term remodeling studies (6 studies in which the follow-up was <12 weeks) resulted in very similar correlation estimates between the mortality OR and the placebo-corrected change in ventricular remodeling reflect the probability of a categorical mortality outcome (favorable, neutral, adverse) for those therapies.

Several individual therapeutic agents with favorable effects on remodeling, including angiotensin-converting enzyme inhibitors (20,42) and beta-adrenergic blockers (85), are associated with favorable effects on clinical outcomes in HF trials. Conversely, agents such as omapatrilat (5,6) or ibopamine (62,125) with neutral or adverse effects on remodeling relative to a comparator have been found to be associated with neutral or adverse effects on clinical outcomes. In distinct trials, the vasopressin V2-receptor antagonist tolvaptan was shown to have a neutral effect on both ventricular remodeling and long-term clinical outcomes (45,67).

Based in part on such data, it is widely conceptualized that ventricular remodeling is biologically related to and involved in the progression of HF (140,141). The relationship is not consistently demonstrated, however. Bozkurt et al. (120) showed a dose-dependent reverse remodeling effect over 3 months with etanercept, a cytokine inhibitor. Three years later, a much larger long-term study failed to show any long-term clinical outcome benefit (4).

### Discussion

The results of the present study demonstrate a significant association between short-term trial-level therapeutic effects of a drug or device on parameters of LV remodeling and longer-term trial-level therapeutic effects on mortality in LVD. Furthermore, these drug/device-induced changes in ventricular remodeling reflect the probability of a categorical mortality outcome (favorable, neutral, adverse) for those therapies.

#### Table 2 Absolute Effect of Drug/Device on Change in EF Compared With Placebo

<table>
<thead>
<tr>
<th>Intervention (Ref. #)</th>
<th>No. of Studies (n [Range])</th>
<th>ΔEF (95% CI)†</th>
<th>Mean Follow-Up Weeks [Range]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amlodipine (70)</td>
<td>1 (362)</td>
<td>1.9 (1.8 to 2.0)</td>
<td>12</td>
</tr>
<tr>
<td>Bisoprolol (41)</td>
<td>1 (28)</td>
<td>12.0 (4.4 to 19.6)</td>
<td>52</td>
</tr>
<tr>
<td>Bucindolol (19-71-73)</td>
<td>4 (2,915 [19-2,708])</td>
<td>4.2 (3.7 to 4.7)</td>
<td>22 [12-52]</td>
</tr>
<tr>
<td>CRT (74-77)</td>
<td>4 (1,052)</td>
<td>2.7 (1.9 to 3.5)</td>
<td>21 [6-26]</td>
</tr>
<tr>
<td>Candesartan (78)</td>
<td>1 (305)</td>
<td>4.0 (0.5 to 7.5)</td>
<td>26</td>
</tr>
<tr>
<td>Captopril (79-84)</td>
<td>6 (543 [40-204])</td>
<td>3.3 (0.3 to 6.4)</td>
<td>21 [3-26]</td>
</tr>
<tr>
<td>Carvedilol (23,24,49,85-104)</td>
<td>22 (2,780 [15-415])</td>
<td>6.9 (5.8 to 8.0)</td>
<td>30 [13-52]</td>
</tr>
<tr>
<td>Digoxin (84,105-109)</td>
<td>6 (624 [13-196])</td>
<td>2.7 (1.2 to 4.1)</td>
<td>48.3 [12-208]</td>
</tr>
<tr>
<td>Enalapril (20,42,110-113)</td>
<td>6 (431 [12-301])</td>
<td>3.7 (1.5 to 5.9)</td>
<td>24 [4-52]</td>
</tr>
<tr>
<td>Enalapril-Prev (21)*</td>
<td>1 (108)</td>
<td>2.0 (-0.8 to 4.8)</td>
<td>52</td>
</tr>
<tr>
<td>Enoximone (114-119)</td>
<td>6 (203 [12-114])</td>
<td>3.4 (0.5 to 6.3)</td>
<td>8.7 [4-16]</td>
</tr>
<tr>
<td>Etanercept (120)</td>
<td>1 (47)</td>
<td>4.4 (3.7 to 5.1)</td>
<td>13</td>
</tr>
<tr>
<td>Felodipine (43,121,122)</td>
<td>3 (532 [20-260])</td>
<td>4.0 (1.2 to 6.7)</td>
<td>30 [12-52]</td>
</tr>
<tr>
<td>Flosequiman (123,124)</td>
<td>2 (210 [17-193])</td>
<td>-3.0 (-3.6 to -2.4)</td>
<td>10 [8-12]</td>
</tr>
<tr>
<td>Hydralazine-SDN (16,22)</td>
<td>2 (1,137 [459-678])</td>
<td>2.9 (0.8 to 5.0)</td>
<td>39 [26-52]</td>
</tr>
<tr>
<td>Iboamphetamine (125)</td>
<td>1 (18)</td>
<td>0.0 (-4.9 to 4.9)</td>
<td>5</td>
</tr>
<tr>
<td>Metoprolol CR (39,40,126,127)</td>
<td>4 (587 [41-426])</td>
<td>4.5 (1.8 to 7.1)</td>
<td>25.5 [24-26]</td>
</tr>
<tr>
<td>Milbemidrid (44)</td>
<td>1 (117)</td>
<td>0.5 (-2.8 to 3.8)</td>
<td>26</td>
</tr>
<tr>
<td>Milrinone (109)</td>
<td>1 (108)</td>
<td>2.2 (1.5 to 2.9)</td>
<td>53</td>
</tr>
<tr>
<td>Moxonidine (128)</td>
<td>1 (85)</td>
<td>4.0 (-0.5 to 8.5)</td>
<td>19</td>
</tr>
<tr>
<td>Prazosin (16,129-131)</td>
<td>4 (523 [22-456])</td>
<td>2.5 (0.6 to 4.4)</td>
<td>28.3 [9-52]</td>
</tr>
<tr>
<td>Spironolactone (132-134)</td>
<td>3 (185 [37-106])</td>
<td>3.0 (1.9 to 4.1)</td>
<td>25.7 [8-52]</td>
</tr>
<tr>
<td>Tolvaptan (45)</td>
<td>1 (240)</td>
<td>0.8 (-0.3 to 1.9)</td>
<td>54</td>
</tr>
<tr>
<td>Valsartan (38)</td>
<td>1 (5,010)</td>
<td>1.3 (0.7 to 1.9)</td>
<td>78</td>
</tr>
</tbody>
</table>

*Enalapril was examined separately when studied in patients with asymptomatic LVD, as this was the only trial that exclusively examined patients with asymptomatic LVD. †When 1 trial existed for a specific therapy, meta-analysis was used to calculate the absolute change (Δ) in ejection fraction (EF) compared with placebo (ΔEF = mean net difference in EF between intervention and placebo groups: (intervention EF – baseline) – (placebo EF – baseline), in EF % units).

Abbreviations as in Table 1.
The morbidity and mortality associated with HF result from a multifactorial and complex process unlikely to be wholly captured by change in a single volume measurement or serologic parameter (7,142). Although some events that impact clinical course are plausibly related to remodeling (such as hospitalization for HF or risk of sudden death), other mortal events, such as acute myocardial infarction, are likely less related. Thus, the expectation that a single surrogate marker could predict the effects of a drug or device on clinical outcomes with high precision is unrealistic. For this reason, we recognize that the effect of a drug or device on LV remodeling, or on any single parameter for that matter, is unlikely to achieve the level of precision discussed by Prentice (143) or by Fleming and DeMets.

### Table 3: Absolute Effect of Drug/Device on Change in EDV Compared With Placebo

<table>
<thead>
<tr>
<th>Intervention (Ref. #)</th>
<th>No. of Studies (n [Range])</th>
<th>ΔEDV (95% CI)†</th>
<th>Mean Follow-Up, Weeks [Range]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisoprolol (25,41)</td>
<td>2 (188 [28–160])</td>
<td>−25.5 (−108.1 to 3.2)</td>
<td>37 [22–52]</td>
</tr>
<tr>
<td>Bucindolol (72,73)</td>
<td>2 (188 [49–139])</td>
<td>−37.1 (−75.8 to 1.6)</td>
<td>12</td>
</tr>
<tr>
<td>CRT (74–78,135)</td>
<td>5 (1,086 [34–490])</td>
<td>−31.8 (−33.6 to −30.0)</td>
<td>19.4 [6–26]</td>
</tr>
<tr>
<td>Candesartan (79)</td>
<td>1 (305)</td>
<td>−8.2 (−232.0 to 215.6)</td>
<td>26</td>
</tr>
<tr>
<td>Captopril (81–83,85,136–138)</td>
<td>7 (668 [40–298])</td>
<td>−15.4 (−19.5 to −11.4)</td>
<td>44.4 [25–52]</td>
</tr>
<tr>
<td>Carvedilol (85,86,88–91,95,97,98,103,139)</td>
<td>11 (900 [21–415])</td>
<td>−26.7 (−40.5 to −13.0)</td>
<td>29.3 [13–52]</td>
</tr>
<tr>
<td>Digoxin (106,108)</td>
<td>2 (266 [88–178])</td>
<td>−9.9 (−39.7 to 20.0)</td>
<td>16 [12–20]</td>
</tr>
<tr>
<td>Enalapril (20,42,110)</td>
<td>3 (374 [17–301])</td>
<td>−11.1 (−20.8 to −1.4)</td>
<td>30 [12–52]</td>
</tr>
<tr>
<td>Enalapril-Prev (21)*</td>
<td>1 (108)</td>
<td>−5.0 (−20.0 to 10.0)</td>
<td>52</td>
</tr>
<tr>
<td>Enoximone (114,115)</td>
<td>2 (44 [20–24])</td>
<td>31.6 (−85.0 to 148.3)</td>
<td>10 [4–16]</td>
</tr>
<tr>
<td>Etanercept (120)</td>
<td>1 (47)</td>
<td>−18.0 (−22.7 to −13.3)</td>
<td>13</td>
</tr>
<tr>
<td>Felodipine (43,122)</td>
<td>2 (280 [20–260])</td>
<td>−52.7 (−161.8 to 56.4)</td>
<td>39 [26–52]</td>
</tr>
<tr>
<td>Hydralazine-ISON (22)</td>
<td>1 (459)</td>
<td>−5.6 (−19.8 to 8.5)</td>
<td>26</td>
</tr>
<tr>
<td>Ibopamine (125)</td>
<td>1 (19)</td>
<td>33.9 (−43.8 to 111.5)</td>
<td>5</td>
</tr>
<tr>
<td>Metoprolol CR (39,126,127)</td>
<td>3 (486 [41–426])</td>
<td>−27.6 (−63.9 to 8.8)</td>
<td>24.7 [24–26]</td>
</tr>
<tr>
<td>Prazosin (129,130)</td>
<td>2 (45 [22–23])</td>
<td>4.1 (−50.8 to 59.1)</td>
<td>17.5 [9–26]</td>
</tr>
<tr>
<td>Spironolactone (132,133)</td>
<td>2 (143 [37–106])</td>
<td>−26.9 (−42.3 to −11.5)</td>
<td>34.5 [17–52]</td>
</tr>
<tr>
<td>Tolvaptan (45)</td>
<td>1 (240)</td>
<td>−3.4 (−9.1 to 2.3)</td>
<td>54</td>
</tr>
<tr>
<td>Valsartan (38)</td>
<td>1 (5,010)</td>
<td>−0.0 (−1.5 to 1.4)</td>
<td>78</td>
</tr>
</tbody>
</table>

*Enalapril was examined separately when studied in patients with asymptomatic LVD, as this was the only trial that exclusively examined patients with asymptomatic LVD. †When >1 trial existed for a specific therapy, meta-analysis was used to calculate the absolute change (Δ) in end-diastolic volume (EDV) compared with placebo (ΔEDV = mean net difference in EDV between intervention and placebo groups; [intervention EDV − baseline] − [placebo EDV − baseline], in ml).

### Table 4: Absolute Effect of Drug/Device on Change in ESV Compared With Placebo

<table>
<thead>
<tr>
<th>Intervention (Ref. #)</th>
<th>No. of Studies (n [Range])</th>
<th>ΔESV (95% CI)†</th>
<th>Mean Follow-Up, Weeks [Range]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisoprolol (25,41)</td>
<td>2 (188 [28–160])</td>
<td>−63.0 (−111.1 to −14.9)</td>
<td>37 [22–52]</td>
</tr>
<tr>
<td>CRT (74–78,135)</td>
<td>5 (1,086 [34–490])</td>
<td>−25.8 (−28.5 to −23.2)</td>
<td>19.4 [6–26]</td>
</tr>
<tr>
<td>Candesartan (79)</td>
<td>1 (305)</td>
<td>−11.3 (−222.8 to 200.1)</td>
<td>26</td>
</tr>
<tr>
<td>Captopril (80–82,84,136–138)</td>
<td>7 (668 [40–298])</td>
<td>−15.7 (−21.9 to −9.6)</td>
<td>44.4 [25–52]</td>
</tr>
<tr>
<td>Carvedilol (85,86,88,89,95,97,98,103,139)</td>
<td>9 (852 [21–415])</td>
<td>−33.9 (−48.4 to −19.3)</td>
<td>32.2 [13–52]</td>
</tr>
<tr>
<td>Digoxin (108)</td>
<td>1 (178)</td>
<td>−19.5 (−40.1 to 1.0)</td>
<td>12</td>
</tr>
<tr>
<td>Enalapril (20,42)</td>
<td>2 (351 [50–301])</td>
<td>−19.6 (−46.2 to 7.0)</td>
<td>52 [52]</td>
</tr>
<tr>
<td>Enalapril-Prev (21)*</td>
<td>1 (108)</td>
<td>−5.0 (−18.6 to 8.6)</td>
<td>52</td>
</tr>
<tr>
<td>Etanercept (120)</td>
<td>1 (47)</td>
<td>−24.3 (−28.8 to −19.8)</td>
<td>13</td>
</tr>
<tr>
<td>Felodipine (43,122)</td>
<td>2 (280 [20–260])</td>
<td>−55.4 (−157.1 to 46.4)</td>
<td>39 [26–52]</td>
</tr>
<tr>
<td>Hydralazine-ISON (22)</td>
<td>1 (459)</td>
<td>8.5 (−18.4 to 0.8)</td>
<td>26</td>
</tr>
<tr>
<td>Ibopamine (125)</td>
<td>1 (18)</td>
<td>28.2 (−48.5 to 104.9)</td>
<td>5</td>
</tr>
<tr>
<td>Metoprolol CR (39,127)</td>
<td>2 (467 [41–426])</td>
<td>−30.8 (−73.4 to 11.7)</td>
<td>25 [24–26]</td>
</tr>
<tr>
<td>Prazosin (129,130)</td>
<td>2 (45 [22–23])</td>
<td>−4.3 (−62.2 to 53.5)</td>
<td>17.5 [9–26]</td>
</tr>
<tr>
<td>Spironolactone (133,134)</td>
<td>2 (143 [37–106])</td>
<td>−27.7 (−31.3 to −24.0)</td>
<td>34.5 [17–52]</td>
</tr>
<tr>
<td>Tolvaptan (45)</td>
<td>1 (240)</td>
<td>−5.4 (−12.2 to 1.4)</td>
<td>54</td>
</tr>
</tbody>
</table>

*Enalapril was examined separately when studied in patients with asymptomatic LVD, as this was the only trial that exclusively examined patients with asymptomatic LVD. †When >1 trial existed for a specific therapy, meta-analysis was used to calculate the absolute change in end-systolic volume (ESV) compared with placebo (ΔESV = mean net difference in ESV between intervention and placebo groups; [intervention ESV − baseline] − [placebo ESV − baseline], in ml).

Abbreviations as in Table 1.
as a valid surrogate marker for mortality in HF.

Rather, the data from this analysis suggest that drug/device effects on remodeling should be viewed as suggestive of the intervention’s potential effect on mortality. Given the demonstrated proportional relationship between drug/device effects on short-term ventricular remodeling and long-term mortality, it is reasonable to conclude that the effects of a drug or device on LV remodeling can be taken into consideration during the development process of novel therapeutic agents for HF, in creating a probability signal of the likelihood of a favorable, neutral, or adverse effect of the therapy being investigated on longer-term mortality outcomes.

Study limitations. There are important limitations to this analysis. Inherent in this retrospective analysis of prospective studies are both publication bias and selection bias. To
avoid within-trial selection biases, we used only randomized trials that fit rigorous inclusion criteria (recognizing that some potentially relevant works might be excluded from the analysis). Publication bias poses threats to all meta-epidemiological studies, including this study. It may be less of a problem for the mortality trials, because large RCTs may be published irrespective of their findings. Even if publication bias is a problem for remodeling trials, it is unclear whether it would affect the associations described, and to which direction. Nonpublication of neutral remodeling trials would have no systematic effect on our results for interventions with a neutral effect on mortality, but it would bias our results away from the null for interventions with adverse or favorable effects on mortality. Moreover, published studies of remodeling are analyses of “completers,” that is, patients who have both baseline and final data for

Figure 3  Quantitative Relationship Between Drug/Device Effects on EDV and Mortality, and Predicted Probability of Mortality by Drug/Device Effect on EDV

(A) Quantitative relationship between drug/device effects on end-diastolic volume (EDV) and mortality: each data point represents a placebo-corrected change in EDV from an individual remodeling trial plotted against the mortality OR for the specific therapy. Color-coded mortality effect based on data from mortality trials listed in Table 1: favorable (blue circles), neutral (black circles), or adverse (red circles). Definition of mortality effect for a given intervention is as described in Figure 2A. There was a significant correlation between effect sizes in the mortality trials and mean therapy-induced changes in EDV ($r = 0.44, p = 0.002$). Remodeling data derived from analysis of 50 randomized controlled trials (RCTs) of 19 interventions, including 10,855 total patients. (B) Predicted probability of a categorical mortality outcome based on drug/device effect on EDV: the lines represent the likelihood of a categorical mortality outcome based on an intervention’s trial-level effect on EDV compared with placebo (unadjusted, weighted, ordered logistic regression). Color-coded mortality effect based on data from mortality trials listed in Table 1. Definition of mortality effect for a given intervention is as described in Figure 2A. The graph suggests that if the mean change in EDV is $-10$ ml in the short-term studies, the probabilities that the long-term mortality studies will be significantly favorable (blue line), neutral (black line), and significantly unfavorable (red line) is approximately 56%, 43% and 1%, respectively.
It is difficult to predict the effect of the “noncompleters” on the analysis if their information was available. If anything, data from noncompleters would attenuate the average EF, EDV, or ESV changes toward the unfavorable direction. It is likely that this shift would be influential only if a substantial proportion of patients had died in the remodeling studies. The median proportion of deaths in the remodeling studies was 3.6%, (25th and 75th quartiles, 0% – 3.6%).
and 8.3%, respectively) among 79 trials that report this information. These numbers suggest that the effects of noncompleters on our analyses would not likely be dramatic.

Second, most data pertain to EF. We approximated changes in LV volume-related outcomes (EDV, ESV) from data on LV indexes and diameters in the majority of pertinent studies. As discussed in Methods, the concordance of the approximations and the derived data was not good in a small number of examples. All analyses were nonsignificant among the 14 trials that directly reported EDV and ESV measurements, but this could be attributed to substantial loss of statistical power. Therefore, results on LV volumes should be viewed with caution, and there are insufficient data to draw any conclusions regarding which of the parameters in this analysis may best correlate with mortality outcomes. Additionally, most of the trials evaluated remodeling using radionuclide ventriculography or 2-dimensional echocardiography. In clinical practice, 2-dimensional echocardiography, with its inherent limitations, is predominantly utilized. While we recognize that more contemporary techniques, such as 3-dimensional echocardiography, contrast echocardiography, and cardiac magnetic resonance imaging would more precisely measure volumes, these techniques have not as yet been widely deployed for use in multicenter trials, and were not used for the drugs and devices that were included in this analysis.

Third, we grouped results from trials spanning different therapeutic eras and were heterogeneous in their follow-up duration, ranging from 4 weeks to >1 year. This was expected, as there is no generally established time frame for the ideal assessment of an intervention’s effect on ventricular remodeling. External validation is not yet available for these results that represent the current published literature.

Fourth, our analyses are sensitive to confounding and to ecological fallacies. Ideally, we would associate changes in the remodeling outcomes and corresponding changes in the clinical outcomes measured in the same patients. If individual patient data are available, one can use structural equation modeling or path analysis methodologies. If sufficient summary statistics are available, one can use specially developed meta-analytic techniques (145). However, in most cases, including the current one, such data are not available. We believe that even ecological associations are of interest because they can help formulating hypotheses for further study.

In addition, we acknowledge that our analyses do not prove the validity of remodeling outcomes as surrogate outcomes for mortality in LVD, as defined by others (143,144), and indeed, we believe the data suggest that no serum biomarker or cardiac structural marker will do so, given the complexity of the HF syndrome. Rather, the data suggest a quantifiable association of a marker such as remodeling and a longer-term outcome such as mortality, which may be a useful signal in the therapeutic development process. We acknowledge the possibility that there is indeed a stronger relationship between remodeling and mortality, and that the limitations of our approach given the available data obfuscate our ability to discern such a relationship, particularly within an individual intervention. The modest correlation of the remodeling effect of interventions with the mortality effect seen in this analysis is influenced by both the underlying inherent relation between the therapeutic effect on remodeling on mortality, as well as the fact that different interventions will have various proportions of their effect mediated through a remodeling mechanism.

Finally, the analyses of changes in LVEF and volumes reported in the literature and summarized here do not allow a clear distinction between a simple functional change in EF or volumes and a true long-term structural change in the underlying LV architecture resulting from therapy (what would truly be considered remodeling). To assess the latter, a study design needs to include a “withdrawal” study, in which EF and volumes are reassessed after withdrawal of drug or device influence. Some published studies indeed report such results (20,21,45), although the vast majority of remodeling studies do not.

Conclusions

Hence, based on analysis of the current literature investigating patients with LVD, there is a significant correlation and a salient predictability signal between short-term trial-level therapeutic effects on LV remodeling and longer-term trial-level therapeutic effects on mortality. Our findings indicate that the effect of a drug or device on LV remodeling can be viewed as a probability signal of outcome effects of those therapies during the development process of novel therapeutic agents for the treatment of HF, rather than as a precise surrogate. Although no surrogate end point, alone, will ever fully substitute for a mortality assessment, our findings of an increased probability of a survival benefit for an intervention associated with improvements in remodeling parameters imply that the demonstration of favorable remodeling renders a survival signal more credible. Moreover, further development of this dataset may allow for quantitative modeling to predict a new agent’s likelihood of mortality benefit given the agent’s short-term effect on remodeling parameters.

Reprint requests and correspondence: Dr. James E. Udelson, Division of Cardiology, Tufts Medical Center, 750 Washington Street, Box 70, Boston, Massachusetts 02111. E-mail: judelson@tuftsmedicalcenter.org.

REFERENCES


27. Antonopoulos GV, Konstam MA, Udelen JE. Are drug-induced changes in left ventricular ejection fraction or volumes adequate surrogates for long-term natural history outcomes in heart failure? Circulation 1999;100:1296.


125. Rousseau MF, Konstam MA, Benedict CR, et al. Progression of left ventricular dysfunction secondary to coronary artery disease, sus-
tained neurohormonal activation and effects of ibopamine therapy
during long-term therapy with angiotensin-converting enzyme
126. de Milliano PA, de Groot AC, Tijssen JG, van Eck-Smit BL, Van
Zwieten PA, Lie KI. Beneficial effects of metoprolol on myocardial
sympathetic function: evidence from a randomized, placebo-
controlled study in patients with congestive heart failure. Am Heart J
2002;144:E3.
127. The RESOLVD Investigators. Effects of metoprolol CR in patients
with ischemic and dilated cardiomyopathy: the randomized evalua-
tion of strategies for left ventricular dysfunction pilot study. Circula-
128. Swedberg K, Bristow MR, Cohn JN, et al. Effects of sustained-
release moxonidine, an imidazoline agonist, on plasma norepineph-
rine in patients with chronic heart failure. Circulation 2002;
129. Markham RV Jr., Corbett JR, Gilmore A, Pettinger WA, Firth BG.
Efficacy of prazosin in the management of chronic congestive heart
failure: a 6-month randomized, double-blind, placebo-controlled
therapy of heart failure with prazosin: a randomized double blind
trial. Am J Cardiol 1980;45:337–44.
131. Higginbotham MB, Morris KG, Bramlet DA, Coleman RE, Cobb
FR. Long-term ambulatory therapy with prazosin versus placebo for
chronic heart failure: relation between clinical response and left
ventricular function at rest and during exercise. Am J Cardiol
effects of spironolactone on left ventricular function and exercise
tolerance in patients with chronic heart failure. J Am Coll Cardiol
plasma brain natriuretic peptide and left ventricular remodeling in
patients with congestive heart failure. J Am Coll Cardiol 2001;37:
1228–33.
Effects of adding spironolactone to an angiotensin-converting en-
zyme inhibitor in chronic congestive heart failure secondary to
135. Duncan A, Wait D, Gibson D, Daubert JC. Left ventricular
remodelling and haemodynamic effects of multisite biventricular
pacing in patients with left ventricular systolic dysfunction and
activation disturbances in sinus rhythm: sub-study of the MUSITC
(Multisite Stimulation in Cardiomyopathies) trial. Eur Heart J
2003;24:430–41.
136. Pfeffer MA, Lamas GA, Vaughan DE, Parist AF, Braunwald E.
Effect of captopril on progressive ventricular dilatation after anterior
137. van Gilst WH, Kingma JH, Peels KH, Dambrik JH, St John Sutton
M. Which patient benefits from early angiotensin-converting enzyme
inhibition after myocardial infarction? Results of one-year serial
echocardiographic follow-up from the Captopril and Thrombolysis
138. Sogaard P, Gotzsche CO, Ravkilde J, Thygesen K. Effects of
captopril on ischemia and dysfunction of the left ventricle after
on left ventricular remodelling in chronic stable heart failure: a
140. Cohn JN, Ferrari R, Sharpe N. Cardiac remodeling—concepts and
clinical implications: a consensus paper from an international forum
141. Sutton MG, Sharpe N. Left ventricular remodeling after myocardial
infarction: pathophysiology and therapy. Circulation 2000;101:
2981–8.
143. Prentice RL. Surrogate endpoints in clinical trials: definition and
144. Fleming TR, DeMets DL. Surrogate endpoints in clinical trials: are
145. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of

Key Words: heart failure ● remodeling ● clinical trials.