



Re-Examination of the BEST Trial Using Composite Outcomes, Including Emergency Department Visits

Li Shen, MChB,^a Pardeep S. Jhund, MChB, PhD,^a Ulrik M. Mogensen, MD, PhD,^{a,b} Lars Køber, MD, DMSc,^b Brian Claggett, PhD,^c Jennifer K. Rogers, PhD,^d John J.V. McMurray, MD^a

ABSTRACT

OBJECTIVES The influence of choice of endpoint on trial size, duration, and interpretation of results was examined in patients with heart failure who were enrolled in BEST (Beta-blocker Evaluation of Survival Trial).

BACKGROUND The choice of endpoints in heart failure trials has evolved over the past 3 decades.

METHODS In the BEST trial, we used Cox regression analysis to examine the effect of bucindolol on the current standard composite of cardiovascular death or heart failure hospitalization (CVD/HFH) compared with the original primary mortality endpoint and the expanded composite that included emergency department (ED) visits. We also undertook an analysis of recurrent events primarily using the Lin, Wei, Ying, and Yang model.

RESULTS Overall, 448 (33%) patients on placebo and 411 (30%) patients on bucindolol died (hazard ratio [HR]: 0.90; 95% confidence interval [CI]: 0.78 to 1.02; $p = 0.11$). A total of 730 (54%) patients experienced CVD/HFH on placebo and 624 (46%) on bucindolol (HR: 0.80; 95% CI: 0.72 to 0.89; $p < 0.001$). Adding ED visits increased these numbers to 768 (57%) and 668 (49%), respectively (HR: 0.81; 95% CI: 0.73 to 0.90; $p < 0.001$). A total of 568 (42%) patients on placebo experienced HFH compared with 476 (35%) patients on bucindolol (HR: 0.78; 95% CI: 0.69 to 0.89; $p < 0.001$), with a total of 1,333 and 1,124 admissions, respectively. With the same statistical assumptions, using the composite endpoint instead of all-cause mortality would have reduced the trial size by 40% and follow-up duration by 69%. The rate ratio for recurrent events (CVD/HFH) was 0.83 (95% CI: 0.73 to 0.94; $p = 0.003$).

CONCLUSIONS Choice of endpoint has major implications for trial size and duration, as well as interpretation of results. The value of broader composite endpoints and inclusion of recurrent events needs further investigation. (Beta Blocker Evaluation in Survival Trial [BEST]; [NCT00000560](https://clinicaltrials.gov/ct2/show/study/NCT00000560)) (J Am Coll Cardiol HF 2017;5:591-9) © 2017 by the American College of Cardiology Foundation.

The choice of endpoints in heart failure (HF) trials has evolved over the past 3 decades. Initially, death from any cause was commonly used as the primary endpoint; however, with incremental improvements in therapy, it has become more common to use mortality–morbidity composite outcomes (1–3). In part, these reflect improving survival in HF and the resultant feasibility and affordability of conducting mortality trials.

However, incorporation of hospital admissions for HF in composites also recognizes the importance of these nonfatal events to the overall burden of HF and their economic significance (4–6). Recently, cardiovascular (CV) mortality, rather than all-cause mortality, has been incorporated into composite outcomes. This recognizes the likely absence of the effect of novel treatments for HF on noncardiovascular death and the growing proportion of deaths

From the ^aBHF Cardiovascular Research Centre, University of Glasgow, Glasgow, United Kingdom; ^bRigshospitalet Copenhagen University Hospital, Copenhagen, Denmark; ^cDivision of Cardiovascular Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts; and the ^dUniversity of Oxford, Oxford, United Kingdom. Dr. Shen is supported by a post doctoral research grant from the China Scholarship Council, China. Dr. Køber has received honoraria from Sanofi and Novartis for speaking. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ABBREVIATIONS AND ACRONYMS

CI	= confidence interval
CV	= cardiovascular
CVD	= cardiovascular death
ED	= emergency department
HF	= heart failure
HFH	= heart failure hospitalization
HR	= hazard ratio
LVEF	= left ventricular ejection fraction
LWYY	= Lin, Wei, Ying, and Yang
NYHA	= New York Heart Association
WLW	= Wei, Lin, and Weissfeld

attributable to noncardiovascular causes, because of the cumulative benefits of effective treatments on CV mortality (7-9). Similarly, with improving survival and chronicity of HF, it has been suggested that analysis of all events, including repeat events, better reflects the overall burden of the condition than the conventional time-to-first-event analysis (10-14). Recently, clinical practice has evolved, particularly in the United States, to attempt to manage worsening episodes of HF without formal admission to hospital. This potentially means that heart failure hospitalization (HFH) may no longer reflect the true extent of treatment failure. Consequently, it has been suggested that these nonhospitalized episodes should be included in composite outcomes (5,15,16). However, there are few data on the frequency of these and whether they respond to study treatment in the same way as hospital admission.

SEE PAGE 600

We used the BEST (Beta-blocker Evaluation of Survival Trial) to examine the implications of this evolution in trial endpoints in HF with a reduced ejection fraction (17,18). BEST is of particular interest because information of emergency department (ED) visits and HFHs was collected systematically during the trial.

METHODS

STUDY DESIGN AND PATIENTS. BEST was a randomized double-blind trial of bucindolol in patients with HF, funded by the National Heart, Lung, and Blood Institute and the Department of Veterans Affairs (17,18). The BEST protocol and results have been published. In brief, 2,708 patients with HF with a left ventricular ejection fraction (LVEF) $\leq 35\%$ and New York Heart Association (NYHA) functional class III or IV symptoms were enrolled in the United States and Canada from 1995 to 1998, and randomly assigned to receive bucindolol or placebo. The primary endpoint was death from any cause. The secondary endpoints included cardiovascular death (CVD) and HFH. The cause of death was adjudicated blindly by the central endpoint committee. The de-identified public-use copy of the BEST database provided by the National Heart, Lung, and Blood Institute, which included all but 1 participant, was used for the present analysis.

OUTCOMES. HFHs and ED visits for HF were reported by investigators. Specifically, the investigator was asked on the hospitalization or ED visit form to

state whether the visit was due to worsening HF ("yes" or "no"), for which investigators were instructed to select "yes" only if the visit was due to decompensated HF. We defined an isolated ED visit for HF as one that occurred without a subsequent HFH within 30 days, and if patients were hospitalized within 30 days after an ED visit, they were classified as having a HFH. The outcomes of interest in this analysis included: the composite of time to first HFH or CVD; the expanded composite of time to first CVD, HFH, or ED visit for HF; all HFHs (including repeats); and a composite of all HFHs and CVD (each CVD was counted as an additional event except when a patient died during a HF admission).

STATISTICAL ANALYSES. Time-to-first-event analyses. The baseline characteristics of patients who had a first isolated ED visit for HF, HFH, or CVD, or none of these events, were compared using analysis of variance for continuous variables and the chi-square test for categorical variables. HF duration was not normally distributed and thus was compared using the Kruskal-Wallis test.

The association between a first nonfatal event (ED visit for HF or HFH) and subsequent mortality was examined using time-updated Cox regression analysis with patients with neither event as the reference group. The association was adjusted for treatment assignment and baseline covariates, including age, sex, race, systolic blood pressure, heart rate, body mass index, LVEF, NYHA functional class, ischemic etiology, hypertension, diabetes, myocardial infarction, atrial fibrillation, previous implantable cardioverter-defibrillator, and serum creatinine. The treatment effect on the composite, and on the expanded composite and its components was examined using Cox regression analysis.

Assuming all-cause mortality, the composite endpoint, or the expanded composite as the endpoint, we examined the time taken to accrue a certain number of the assumed events. We also examined the sample size required to detect a 20% reduction in the assumed endpoint with bucindolol therapy using the log-rank test, assuming a 2-sided significance level of 5%, statistical power of 85%, equal allocation, 3-year uniform accrual period, a minimum follow-up of 1 year, and a maximum follow-up of 4 years.

RECURRENT EVENTS ANALYSIS. Recurrent events are commonly analyzed using count data methods (e.g., negative binomial regression) and time-to-event data methods (e.g., Andersen-Gill, Wei, Lin, and Weissfeld [WLW], and Lin, Wei, Ying, and Yang [LWYY] models), all of which are extensions of Cox proportional hazards regression (10-13). There is

debate about which of these approaches is best to use, and some considerations about this debate are outlined in the [Online Appendix](#). As in the present study, the event rate and treatment effect were not constant during follow-up, which violates the assumption of negative binomial regression; therefore, the LWYY model was used as the primary method, and the negative binomial and WLW regressions were used as sensitivity analyses.

We calculated the HFH rate by treatment group by dividing the total number of HFHs by the total number of follow-up years in each group. The cumulative rates of HFHs over time by treatment group were plotted using the nonparametric Ghosh and Lin method, which accounted for the competing risk of death. Treatment effect on all HFHs and on the composite of all HFHs and CVD was analyzed primarily using the LWYY model and also using the negative binomial and WLW regressions. Because of the inconstant treatment effect on HF hospitalization over time, sensitivity analyses were performed by assessing the treatment effects within 6 months and beyond 6 months since randomization. To account for the association between HFH and subsequent mortality and the competing risk of mortality on HFHs, the joint frailty model was used to analyze recurrent HFHs and time to CVD simultaneously.

A 2-sided p value <0.05 was considered significant. The recurrent event analysis was undertaken using R version 3.2.3 (R Foundation, Vienna, Austria). All other analyses were performed using version 14 (Stata, College Station, Texas).

RESULTS

Of the 2,707 patients analyzed, 1,353 were randomized to placebo and 1,354 to bucindolol. The median duration of follow-up was 2.0 years.

DEATHS. Overall, 448 patients (33%) assigned to placebo and 411 (30%) assigned to bucindolol died with a hazard ratio (HR) in the bucindolol group of 0.90 (95% confidence interval [CI]: 0.78 to 1.02; p = 0.11). The number in each treatment group who died from a CV cause was 388 (29%) and 342 (25%), respectively (HR: 0.86; 95% CI: 0.75 to 1.00; p = 0.045).

HOSPITAL ADMISSIONS FOR HF. Overall, 568 (42%) patients assigned to placebo and 476 (35%) patients assigned to bucindolol had a HFH (HR: 0.78; 95% CI: 0.69 to 0.89; p < 0.001). There was a total of 1,333 admissions in the placebo group and 1,124 in the bucindolol group ([Table 1](#)).

ED VISITS FOR HF. A total of 334 placebo-treated patients (25%) had an ED visit for HF; there were

TABLE 1 Number of Heart Failure Hospitalizations and Emergency Department Visits for HF by Treatment Group in BEST

	Placebo (n = 1,353)	Bucindolol (n = 1,354)
Total yrs of follow-up	2,694	2,755
No. of deaths	448	411
No. of CV deaths	388	342
HF hospitalization		
Patients with ≥1 hospitalization	568	476
Patients with ≥2 hospitalizations	323	246
No. of HF hospitalizations	1,333	1,124
Patients with number of hospitalizations		
1	245	230
2	152	97
3	74	61
4	32	27
5	27	24
6	17	13
7	7	9
8	4	4
9	2	4
10	3	3
11	0	1
12	3	1
13	0	1
14	0	1
16	2	0
ED visit for HF		
Patients with ≥1 ED visit	334	281
Patients with ≥2 ED visits	122	97
No. of ED visits	586	510
Isolated ED visit for HF		
Patients with ≥1 isolated ED visit	161	138
Patients with ≥2 isolated ED visits	31	25
No. of isolated ED visits	211	176

Values are n.
 CV = cardiovascular; ED = emergency department; HF = heart failure.

281 (21%) patients in the bucindolol group (HR: 0.81; 95% CI: 0.69 to 0.95; p = 0.01). Of these, 161 (11.9% of all patients) and 138 (10.2%) patients, respectively, were not admitted to hospital (48% and 49% of patients, respectively, who presented to the ED were not admitted) (HR: 0.84; 95% CI: 0.67 to 1.06; p = 0.14). Overall, there were 586 ED visits for HF in the placebo group and 510 in the bucindolol group. Of these, 211 (36% of visits) and 176 (35% of visits), respectively, did not result in a proximate hospital admission ([Table 1](#)).

CHARACTERISTICS OF PATIENTS WITH AN ADVERSE OUTCOME. The baseline characteristics of patients who experienced a CVD, HFH, or ED visit for HF (or none of these) are shown in [Table 2](#). Overall, patients who died had more characteristics associated with worse outcome (e.g., older age, lower blood pressure, estimated glomerular filtration rate, and

TABLE 2 Baseline Characteristics of Patients With First ED Visit for HF, HF Hospitalization, or Neither, or Experiencing CV Death					
	No Relevant Event	ED Visit for HF	HF Hospitalization	CV Death	p Value
n (%)	1,271 (47.0)	199 (7.4)	986 (36.4)	251 (9.3)	
Age (yrs)	59.1 ± 12.4	59.6 ± 12.2	61.2 ± 12.3	62.8 ± 11.9	<0.001
Male	982 (77.3)	155 (77.9)	776 (78.7)	201 (80.1)	0.727
Race					0.055
White	911 (71.7)	133 (66.8)	662 (67.1)	189 (75.3)	
Black	274 (21.6)	48 (24.1)	256 (26.0)	49 (19.5)	
Hispanic	67 (5.3)	11 (5.5)	56 (5.7)	9 (3.6)	
Other	19 (1.5)	7 (3.5)	12 (1.2)	4 (1.6)	
Blood pressure (mm Hg)					
Systolic	119.5 ± 17.9	116.7 ± 18.5	114.5 ± 17.8	115.7 ± 17.9	<0.001
Diastolic	72.3 ± 11.2	71.8 ± 11.8	69.9 ± 10.9	68.8 ± 11.1	<0.001
Heart rate (beats/min)	81.4 ± 13.0	81.2 ± 13.9	82.1 ± 13.4	81.1 ± 12.8	0.566
BMI (kg/m ²)	28.3 ± 6.0	29.1 ± 6.4	27.5 ± 5.7	27.1 ± 5.4	<0.001
HF duration* (months)	30 (10-64)	39 (10-72)	41 (16-75)	46 (13-78)	<0.001
Left ventricular ejection fraction (%)	24.4 ± 7.0	23.0 ± 7.1	21.7 ± 7.3	21.6 ± 7.1	<0.001
NYHA functional class					<0.001
III	1,208 (95.0)	184 (92.5)	869 (88.1)	220 (87.7)	
IV	63 (5.0)	15 (7.5)	117 (11.9)	31 (12.4)	
Ischemic etiology	672 (52.9)	111 (55.8)	623 (63.2)	181 (72.1)	<0.001
Medical history					
Myocardial infarction	485 (38.2)	75 (37.7)	452 (45.8)	132 (52.6)	<0.001
Angina	618 (48.6)	107 (53.8)	535 (54.3)	140 (55.8)	0.024
CABG	331 (26.0)	60 (30.2)	303 (30.7)	88 (35.1)	0.010
PCI	189 (14.9)	35 (17.6)	161 (16.3)	38 (15.1)	0.672
Hypertension	720 (56.6)	126 (63.3)	598 (60.6)	151 (60.2)	0.129
Diabetes	393 (30.9)	84 (42.2)	380 (38.5)	107 (42.6)	<0.001
Atrial fibrillation	263 (20.7)	39 (19.6)	283 (28.7)	68 (27.1)	<0.001
Treatment					
Randomized treatment	686 (54.0)	97 (48.7)	449 (45.5)	122 (48.6)	0.001
Digitalis	1,141 (89.8)	186 (93.5)	933 (94.6)	239 (95.2)	<0.001
Diuretic	1,156 (91.0)	190 (95.5)	946 (95.9)	241 (96.0)	<0.001
ACE inhibitor	1,168 (91.9)	183 (92.0)	886 (89.9)	232 (92.4)	0.308
ACE inhibitor or ARB	1,241 (97.6)	195 (98.0)	942 (95.5)	239 (95.2)	0.016
Spironolactone	35 (2.8)	4 (2.0)	43 (4.4)	10 (4.0)	0.121
ICD	36 (2.8)	4 (2.0)	46 (4.7)	4 (1.6)	0.020
Pacemaker	95 (7.5)	19 (9.5)	93 (9.4)	24 (9.6)	0.328
Laboratory measures					
Serum creatinine (mg/dl)	1.16 ± 0.36	1.26 ± 0.43	1.32 ± 0.44	1.31 ± 0.40	<0.001
eGFR (ml/min/1.73 m ²)	74.2 ± 23.8	69.4 ± 23.8	65.8 ± 25.5	64.4 ± 22.1	<0.001
eGFR <60 ml/min/1.73 m ²	347 (28.1)	70 (35.5)	427 (44.4)	118 (48.0)	<0.001

Values are n (%) or mean ± SD. *HF duration is presented as median with interquartile range.
ACE = angiotensin converting enzyme; ARB = angiotensin II receptor antagonist; BMI = body mass index; CABG = coronary artery bypass grafting; eGFR = estimated glomerular filtration rate; ICD = implantable cardioverter-defibrillator; NYHA = New York Heart Association; PCI = percutaneous coronary intervention; other abbreviations as in Table 1.

LVEF, ischemic etiology, NYHA functional class IV) and those who had no events had the least of these characteristics. Patients with HFHs and ED visits were in-between these two extremes, although patients with ED visits appeared less sick, overall, compared with those who were hospitalized.

ASSOCIATION BETWEEN HF WORSENING AND SUBSEQUENT MORTALITY.

Compared with patients

who did not experience an ED visit (or HFH), those with an ED visit for HF were subsequently twice as likely to die during follow-up (HR: 2.05; 95% CI: 1.47 to 2.84; $p < 0.001$), even after adjustment for other prognostic variables (HR: 1.90; 1.37 to 2.65; $p < 0.001$). In similar analyses, patients hospitalized for worsening HF were 4 times as likely to die (unadjusted HR: 4.65; 95% CI: 4.02 to 5.37; adjusted HR: 3.72; 95% CI: 3.20 to 4.33; both $p < 0.001$).

COMPOSITE CLINICAL OUTCOMES. The number of patients who experienced the composite of first HFH or CVD was 730 (54%) patients and 624 (46%) patients in the placebo and bucindolol groups, respectively (HR: 0.80; 95% CI: 0.72 to 0.89; $p < 0.001$). Adding ED visits for HF increased the numbers of affected patients to 768 (57%) and 668 (49%), respectively (HR: 0.81; 95% CI: 0.73 to 0.90; $p < 0.001$) (Figure 1).

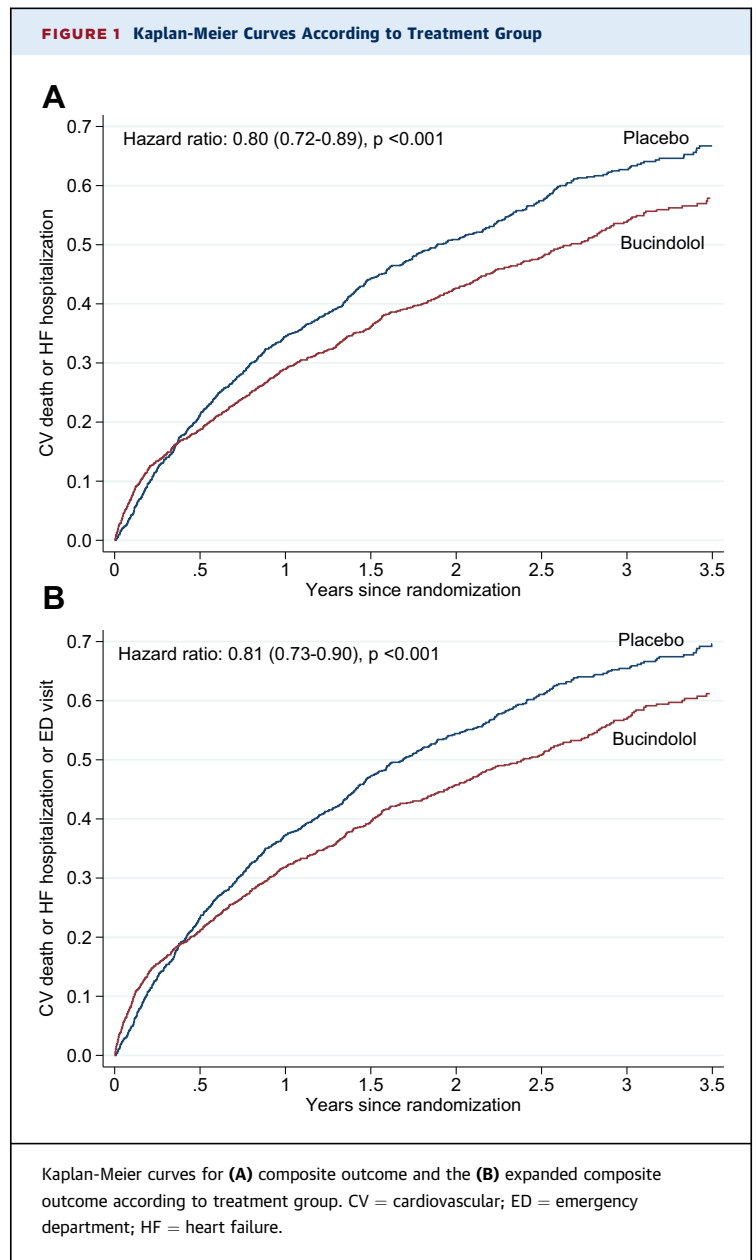
IMPLICATIONS FOR TRIAL SIZE AND DURATION. The number of days taken to accrue 500 patients with a death from any cause was 515; for the composite of HFH or CVD, this number was 162 days, and for the expanded composite including ED visits, it was 136 days. There was a substantial decrease in the sample size using the composite outcomes. For example, with a power of 85% to detect 20% reduction in the bucindolol group at a significance level of 5%, the sample size was 2,454 for death from any cause, 1,524 for the composite, and 1,432 for the composite that included ED visits.

RECURRENT HFHs. There were a total of 1,333 HFHs in the placebo group and 1,124 HFHs in the bucindolol group, including 765 (57.4% of all admissions) and 648 (57.7%) repeated admissions, respectively. The frequencies of HFHs by treatment group are presented in Table 1. More than 25% of all HF admissions occurred within 6 months of randomization, and percentages were 25.2% ($n = 336$) in the placebo group and 29.3% ($n = 329$) in the bucindolol group (Online Figure 1).

The HFH rates were 49.5 and 40.8 per 100 patient-years in the placebo and bucindolol groups, respectively. Compared with the placebo group, the cumulative rate in the bucindolol group was lower after 6 months, although before 6 months, it was slightly higher (i.e., the cumulative event curves crossed over at ~6 months). The corresponding cumulative rate ratio (bucindolol vs. placebo) appeared to remain constant at approximately 0.83 after 6 months (Figure 2).

Patients with at least 1 HFH were more likely to have baseline characteristics associated with worse outcomes (Online Table 1).

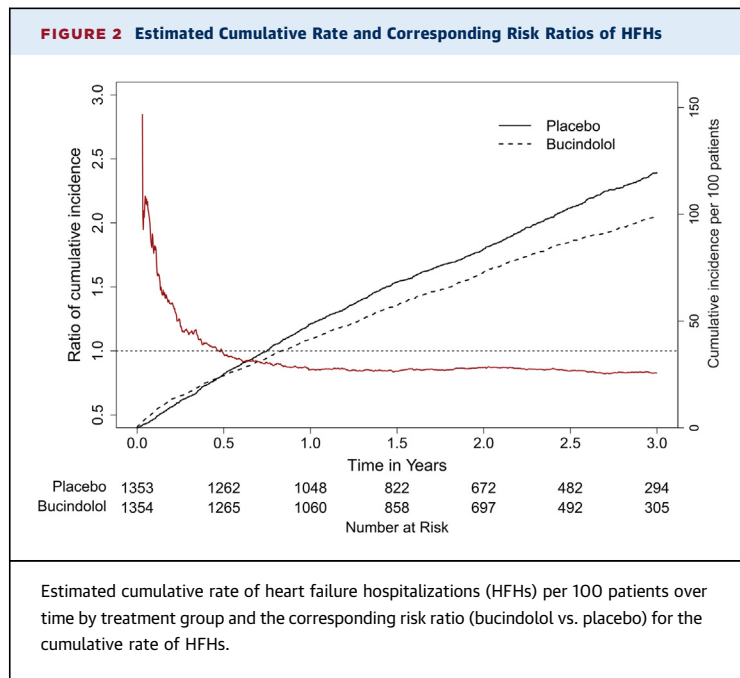
For all HFHs, the LWYY regression model gave an overall HR for bucindolol of 0.82 (95% CI: 0.71 to 0.95; $p = 0.008$). A similar estimate was observed from the WLW model (HR: 0.80; 95% CI: 0.68 to 0.94; $p = 0.005$), whereas a smaller and nonsignificant effect was obtained from negative binomial and joint frailty models (both gave a rate ratio of 0.89, with 95% CI: 0.77 to 1.04) (Table 3). However, when separate estimations were made for the first 6 months and the remainder of follow-up, the results were



consistent across different methods. Based on the LWYY model, the estimate was 0.98 (95% CI: 0.80 to 1.21; $p = 0.88$) within 6 months and 0.77 (95% CI: 0.65 to 0.91; $p = 0.002$) after 6 months. Nearly identical estimates were observed for the composite of all HFHs and CVD from the corresponding regression models (Table 3).

DISCUSSION

The purpose of this study was to illustrate the implications of the choice of primary endpoint in clinical



trials in HF with a reduced ejection fraction and how this choice evolved (and continues to evolve) in recent years. Perhaps the most striking conclusion was that, had the primary endpoint most commonly used in recent HF trials been used in BEST, the trial would have clearly been positive, instead of neutral or negative as it is historically regarded. This difference reflected 2 things. First, the much larger number of events in the composite outcome (1,354 composite events vs. 859 deaths) and the fact that 129 deaths were noncardiovascular. Although a larger number of events, per se, did not increase statistical power,

HFHs were events likely to be favorably influenced by an effective therapy, and therefore, did increase power. Conversely, a beta-blocker was unlikely to decrease the risk of noncardiovascular death, meaning that the 15% of deaths that were not CV effectively diluted the benefit of bucindolol on the original primary endpoint (by adding “noise”). As a result, switching from an all-cause mortality endpoint to the composite of CVD or HFH would have had a dramatic impact on sample size in BEST—assuming the same treatment effect size, power, and significance level (20%, 85%, and 5%, respectively)—the sample size would have been reduced by nearly 40% (from $n = 2,454$ to 1,524), and the time taken to accrue a requisite number of endpoints (e.g., $n = 500$) would have been reduced by an even greater amount (515 to 162 days, a 69% reduction).

Although we found bucindolol reduced the composite of CVD or HFH, a broader composite including ED visits, and recurrent events, it did not reduce all-cause mortality, which was demonstrated with 3 other beta-blockers (19). The reason for this remained uncertain, although the specific pharmacological properties of bucindolol, the racial mix of the population studied in BEST, and interactions between the two were implicated (20,21).

Despite a benefit of bucindolol on the composite of CVD or HFH, an early increase in HFH was observed among patients treated with bucindolol. HF worsening is a recognized risk early after initiation of beta-blocker treatment and is believed to be minimized by starting with a low-dose treatment. This finding was also seen in the MERIT-HF (Metoprolol CR/XL Randomized Intervention Trial in Congestive Heart Failure), although it was apparently not found in the

TABLE 3 Treatment Effects (Bucindolol vs. Placebo) on All HF hospitalizations (First and Recurrent) and the Composite of All HF Hospitalizations and Cardiovascular Death Using Different Methods to Analyze Recurrent Events

	Entire Follow-Up		Within 6 months		Beyond 6 months	
	HR/RR (95% CI)	p Value	HR/RR (95% CI)	p Value	HR/RR (95% CI)	p Value
Time to first HF hospitalization	0.78 (0.69–0.89)	<0.001	0.88 (0.73–1.06)	0.190	0.72 (0.61–0.84)	<0.001
All HF hospitalizations						
LWYY	0.82 (0.71–0.95)	0.008	0.98 (0.80–1.21)	0.880	0.77 (0.65–0.91)	0.002
WLW	0.80 (0.68–0.94)	0.005	0.99 (0.80–1.23)	0.940	0.74 (0.62–0.89)	0.001
Negative binomial	0.89 (0.77–1.04)	0.137	1.00 (0.82–1.24)	0.970	0.80 (0.68–0.95)	0.010
Joint frailty*	0.89 (0.77–1.04)	0.128	1.02 (0.82–1.27)	0.840	0.80 (0.67–0.95)	0.010
All HF hospitalizations and CV death						
LWYY	0.83 (0.73–0.94)	0.003	0.97 (0.80–1.18)	0.770	0.78 (0.67–0.90)	0.001
WLW	0.80 (0.69–0.92)	0.002	0.98 (0.80–1.19)	0.840	0.75 (0.64–0.88)	<0.001
Negative binomial	0.89 (0.78–1.01)	0.066	1.00 (0.82–1.21)	0.990	0.80 (0.69–0.93)	<0.001

*The corresponding hazard ratio (HR) for CV death in the joint frailty model was 0.93 (95% confidence interval [CI]: 0.77 to 1.13; $p = 0.47$) in the entire follow-up, 1.02 (95% CI: 0.72 to 1.44; $p = 0.91$) within 6 months, and 0.86 (95% CI: 0.69 to 1.07; $p = 0.17$) beyond 6 months, respectively.

LWYY = Lin, Wei, Ying, and Yang model; RR = rate ratio; WLW = Wei, Lin, and Weissfeld model; other abbreviations as in Table 1.

COPERNICUS (Carvedilol Prospective Randomized Cumulative Survival) trial, which also enrolled patients with more severe HF, as was done in BEST (the relevant analysis was not reported for CIBIS-2 (Cardiac Insufficiency Bisoprolol Study-2) (22-24). However, the dose uptitration rate in BEST was more rapid (weekly) than in the other trials (2 weeks in COPERNICUS and MERIT-HF; weekly to 5 mg in CIBIS-2, and then 4 weekly to 10 mg). Another possible reason could be the potent sympatholytic effect of bucindolol. A further analysis of BEST showed that a decrease in plasma norepinephrine levels after 3 months of treatment was associated with higher risk of death or HFH in the bucindolol group (25).

Because of changing practice, it was suggested that composite outcomes be further expanded to include episodes of HF worsening that did not lead to formal hospital admission (5,15,16,26,27). BEST was unusual in systematically documenting ED visits. Although not as numerous as HFHs, ED visits were common. However, most were associated with a hospital admission shortly thereafter. Consequently, in a time-to-first-event analysis, isolated ED visits added relatively few unique events (5%). Nevertheless, these were enough to shorten the time to accrual of a target number of events (as used in an event-driven trial) by approximately 15%. There might be concerns about inclusion of ED visits in the composite outcome (5,15,16,26,27). First, these events might not reflect worsening of HF in the same way as hospitalization, because the events might be less severe or because patient evaluation during an ED visit might be less comprehensive than during a hospital admission. Diagnosis might also be less certain. For these reasons, ED visits might also be less responsive to the experimental treatment (especially if misdiagnosed). Nevertheless, scrutiny of the characteristic of patients with ED visits showed they had features associated with worse outcomes, although these were less marked than in patients who were hospitalized or died. In keeping with this, patients with an ED visit in BEST were subsequently twice as likely to die compared with those without an ED visit (or HFH), which confirmed the findings of another more recent trial and some epidemiological data (16,27); patients hospitalized with worsening HF were 4 times as likely to die. The effect of bucindolol on ED visits was similar to that on CVD and on HFH. Consequently, including ED visits in the composite outcome would not only have reduced the study size (from 1,524 to 1,432 in the previously outlined scenario) and shortened the time to accrual of a target number of events (e.g., from 162 to 132 days for 500 events), but would also have slightly narrowed the

95% CIs around the point estimate for the effect of bucindolol.

As survival has increased, HF has become a more chronic condition with recurrent nonfatal hospitalizations an increasingly important reflection of the overall burden of the disease on patients and health care systems alike. This has led to the suggestion that analysis of all events, including repeated hospital admissions, may provide a better evaluation of the effect of treatment than time-to-first-event analysis, which has been the conventional approach used to estimate treatment effect in clinical trials (5,11-15). A variety of statistical approaches can be used to do such analyses, and there has been discussion about which of these is best to use. We found the two most commonly advocated approaches (i.e., negative binomial and joint frailty models) showed somewhat less favorable treatment effects than the WLW and LWYY models (the principal method in this study). This might result from the violation of 2 important assumptions of negative binomial regression in BEST, that is, the constant event rate and the constant treatment effect over time. This was also the case for the joint frailty model, which is, in effect, a combination of negative binomial regression for recurrent HFHs and Cox regression for time to CVD. The rate of HFHs was relatively high early after randomization (i.e., over the first 6 months) and lower thereafter. As mentioned previously, bucindolol treatment led to an early increase in risk of HF hospitalization followed by a later decrease. When estimations were made separately within 6 months and beyond 6 months, fairly consistent results were observed using the different modeling approaches.

Interestingly, the proportional reduction in risk estimated using all of these methods was smaller than obtained in conventional time-to-first-event analysis. The reason for this observation was uncertain but might reflect the early increase in hospitalization after initiation of bucindolol before the longer term reduction in recurrent events with this treatment became evident. If correct, and whatever the reason, these findings highlighted the need to better understand the effect of therapies on recurrent events and how analyses of these might be used in future clinical trials.

STUDY LIMITATIONS. First, this was a post hoc analysis. Second, it was argued that the actions of bucindolol might be unique among the beta-blockers tested in large outcome trials, although the benefits observed in BEST were generally in keeping with those seen in the other trials (20,28). Third, there was

a potential violation of the proportional hazards assumption for the Cox models in the composite and the expanded composite outcome analyses, because of the crossover in the Kaplan-Meier curves, although the Schoenfeld residuals test was not significant for either (both p value >0.05). Lastly, we only used investigator-reported HFHs and ED visits in our analyses; however, a previous analysis showed a similar treatment effect of bucindolol on first hospitalizations for HF when adjudicated events were used instead (29).

CONCLUSIONS

The choice of endpoint had major implications for trial size and duration, as well as interpretation of results. The use of broader composite endpoints that included nonhospitalized manifestations of HF worsening might further reduce sample size and trial length. However, the role of additional manifestations of worsening other than ED visits needs further study. Similarly, the potential role of analysis of recurrent events as a trial endpoint needs further investigation. This type of analysis might not give the same estimate of treatment effect as time-to-first-event analysis, although the level of agreement might differ for different treatment. However, this

finding did raise the interesting question: which approach, time-to-first-event analysis or analysis of all (first and recurrent) events, gives the more clinically relevant answer?

ADDRESS FOR CORRESPONDENCE: Prof. John J.V. McMurray, British Heart Foundation Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, United Kingdom. E-mail: john.mcmurray@glasgow.ac.uk.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: The choice of primary endpoints has major influence on trial size and duration, and on the interpretation of results. The use of broader composite endpoints including ED visits for HF worsening may further reduce trial size and length.

TRANSLATIONAL OUTLOOK: Further studies are required to determine the values of broader composite endpoints and inclusion of recurrent events, and to standardize the approach for the analysis of recurrent events if included as endpoints.

REFERENCES

1. The CONSENSUS Trial Study Group. Effects of enalapril on mortality in severe congestive heart failure. Results of the Cooperative North Scandinavian Enalapril Survival Study (CONSENSUS). *N Engl J Med* 1987;316:1429-35.
2. SOLVD Investigators, Yusuf S, Pitt B, Davis CE, Hood WB, Cohn JN. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med* 1991;325:293-302.
3. Pitt B, Zannad F, Remme WJ, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. *N Engl J Med* 1999;341:709-17.
4. Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in end point selection for heart failure trials: composite end points. *J Card Fail* 2005;11:567-75.
5. Anker SD, Schroeder S, Atar D, et al. Traditional and new composite endpoints in heart failure clinical trials: facilitating comprehensive efficacy assessments and improving trial efficiency. *Eur J Heart Fail* 2016;18:482-9.
6. Stewart S, Jenkins A, Buchan S, McGuire A, Capewell S, McMurray JJ. The current cost of heart failure to the National Health Service in the UK. *Eur J Heart Fail* 2002;4:361-71.
7. Rush CJ, Campbell RT, Jhund PS, et al. Falling cardiovascular mortality in heart failure with reduced ejection fraction and implications for clinical trials. *J Am Coll Cardiol HF* 2015;3:603-14.
8. Jhund PS, Macintyre K, Simpson CR, et al. Long-term trends in first hospitalization for heart failure and subsequent survival between 1986 and 2003: a population study of 5.1 million people. *Circulation* 2009;119:515-23.
9. Schaufelberger M, Swedberg K, Koster M, Rosen M, Rosengren A. Decreasing one-year mortality and hospitalization rates for heart failure in Sweden; data from the Swedish Hospital Discharge Registry 1988 to 2000. *Eur Heart J* 2004;25:300-7.
10. Rogers JK, Jhund PS, Perez AC, et al. Effect of rosuvastatin on repeat heart failure hospitalizations: the CORONA Trial (Controlled Rosuvastatin Multinational Trial in Heart Failure). *J Am Coll Cardiol* 2014;2:289-97.
11. Rogers JK, Pocock SJ, McMurray JJ, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. *Eur J Heart Fail* 2014;16:33-40.
12. Rogers JK, McMurray JJ, Pocock SJ, et al. Eplerenone in patients with systolic heart failure and mild symptoms: analysis of repeat hospitalizations. *Circulation* 2012;126:2317-23.
13. Anker SD, McMurray JJ. Time to move on from 'time-to-first': should all events be included in the analysis of clinical trials? *Eur Heart J* 2012;33:2764-5.
14. Goldenberg I, Hall WJ, Beck CA, et al. Reduction of the risk of recurring heart failure events with cardiac resynchronization therapy: MADIT-CRT (Multicenter Automatic Defibrillator Implantation Trial With Cardiac Resynchronization Therapy). *J Am Coll Cardiol* 2011;58:729-37.
15. Skali H, Dwyer EM, Goldstein R, et al. Prognosis and response to therapy of first inpatient and outpatient heart failure event in a heart failure clinical trial: MADIT-CRT. *Eur J Heart Fail* 2014;16:560-5.
16. Okumura N, Jhund PS, Gong J, et al. Importance of clinical worsening of heart failure treated in the outpatient setting: evidence from the Prospective Comparison of ARNI With ACEI to Determine Impact on Global Mortality and Morbidity in Heart Failure Trial (PARADIGM-HF). *Circulation* 2016;133:2254-62.
17. Design of the Beta-Blocker Evaluation Survival Trial (BEST). The BEST Steering Committee. *Am J Cardiol* 1995;75:1220-3.
18. Beta-Blocker Evaluation of Survival Trial Investigators, Eichhorn EJ, Domanski MJ, Krause-Steinrauf H, Bristow MR, Lavori PW. A trial of the beta-blocker bucindolol in patients with advanced chronic heart failure. *N Engl J Med* 2001;344:1659-67.

19. Domanski MJ, Krause-Steinrauf H, Massie BM, et al. A comparative analysis of the results from 4 trials of beta-blocker therapy for heart failure: BEST, CIBIS-II, MERIT-HF, and COPERNICUS. *J Card Fail* 2003;9:354-63.
20. O'Connor CM, Fiuzat M, Carson PE, et al. Combinatorial pharmacogenetic interactions of bucindolol and beta1, alpha2C adrenergic receptor polymorphisms. *PLoS One* 2012;7:e44324.
21. Taylor MR, Sun AY, Davis G, Fiuzat M, Liggett SB, Bristow MR. Race, common genetic variation, and therapeutic response disparities in heart failure. *J Am Coll Cardiol HF* 2014;2:561-72.
22. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet* 1999;353:9-13.
23. Hjalmarson A, Goldstein S, Fagerberg B, et al. Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in Congestive Heart Failure (MERIT-HF). *MERIT-HF Study Group. JAMA* 2000;283:1295-302.
24. Packer M, Fowler MB, Roecker EB, et al. Effect of carvedilol on the morbidity of patients with severe chronic heart failure: results of the carvedilol prospective randomized cumulative survival (COPERNICUS) study. *Circulation* 2002;106:2194-9.
25. Bristow MR, Krause-Steinrauf H, Nuzzo R, et al. Effect of baseline or changes in adrenergic activity on clinical outcomes in the beta-blocker evaluation of survival trial. *Circulation* 2004;110:1437-42.
26. Lee DS, Schull MJ, Alter DA, et al. Early deaths in patients with heart failure discharged from the emergency department: a population-based analysis. *Circ Heart Fail* 2010;3:228-35.
27. Ezekowitz JA, Kaul P, Bakal JA, Quan H, McAlister FA. Trends in heart failure care: has the incident diagnosis of heart failure shifted from the hospital to the emergency department and outpatient clinics? *Eur J Heart Fail* 2011;13:142-7.
28. Liggett SB, Mialet-Perez J, Thaneemit-Chen S, et al. A polymorphism within a conserved beta(1)-adrenergic receptor motif alters cardiac function and beta-blocker response in human heart failure. *Proc Natl Acad Sci U S A* 2006;103:11288-93.
29. Carson P, Fiuzat M, O'Connor C, et al. Determination of hospitalization type by investigator case report form or adjudication committee in a large heart failure clinical trial (beta-Blocker Evaluation of Survival Trial [BEST]). *Am Heart J* 2010;160:649-54.

KEY WORDS BEST, endpoint, heart failure, recurrent events

APPENDIX For supplemental text, a figure, and a table, please see the online version of this paper.