CASE 11

The Data Monitoring Experience in the Candesartan in Heart Failure Assessment of Reduction in Mortality and Morbidity Program*

Stuart Pocock
Duolao Wang
Lars Wilhelmsen
Charles H. Hennekens

ABSTRACT

The Candesartan in Heart Failure Assessment of Reduction in Mortality and morbidity (CHARM) program was designed as three separate randomized trials comparing candesartan with placebo in patients with chronic heart failure (CHF) who (1) were intolerant to angiotensin converting enzyme (ACE)-inhibitor and had left ventricular ejection fraction (LVEF) ≤ 40%, (2)) were on ACE-inhibitor and had LVEF ≤ 40% or (3)) had LVEF > 40%. CHARM provides an interesting example of the challenges faced by a Data and Safety Monitoring Committee (DSMC).

While the primary efficacy endpoint for each component trial was cardiovascular (CV) death or hospitalization for CHF, the primary outcome for the overall program was all-cause mortality. The DSMC received monthly safety reports and also met every six months (seven times in all) to review interim reports. Statistical stopping guidelines were predefined for all-cause mortality in the overall program. The overarching principle of the DSMC was proof beyond a reasonable doubt that would be likely to influence clinical practice.

There were significant treatment differences in all-cause mortality at several interim analyses, and the statistical stopping guideline was reached on one occasion. The DSMC consistently recommended that the program

* This paper first published in the *Am Heart J* 2005; 149:939–943.

continue as planned. The final published results for all-cause death over a median 3.1 years were a 9% reduction in hazard (95% CI 0% to 17%, p = 0.055), whereas for CV death or hospitalization for CHF there was a 16% reduction in hazard (95% CI 9% to 23%, p < 0.0001). Subsequent exploratory analyses suggest that the hazard reduction in CV death was more marked in the first year after randomization, and that, if real, this apparent treatment–time interaction offers a plausible explanation for why the interim mortality data showed statistically more extreme findings than the overall final results.

The DSMC experience in the CHARM program illustrates the importance of continuing a trial to its scheduled completion unless there is proof beyond reasonable doubt that would influence clinical practice rather than strict reliance on a statistical stopping guideline.

INTRODUCTION AND BACKGROUND

Angiotensin-receptor blockers such as candesartan offer the potential to improve clinical outcomes in heart failure patients as alternatives or adjuncts to those seen with angiotensin-converting enzyme (ACE) inhibitors. Accordingly the CHARM program[1–4] was designed as three independent randomized double blind trials comparing candesartan with placebo in three populations of patients with symptomatic heart failure:

1. *CHARM—Alternative* patients (N = 2,028) had a left ventricular ejection fraction (LVEF) ≤ 40% and were not on ACE inhibitor because of previous intolerance.[2]
2. *CHARM—Added* patients (N = 2,548) also had LVEF ≤40% and were being treated with an ACE inhibitor.[3]
3. *CHARM—Preserved* patients (N = 3,023) had LVEF >40%.[4]

The primary endpoint for each trial was CV death or hospitalization for worsening CHF and each required sample size was based on power calculations for this endpoint. The overall program was designed to evaluate all-cause mortality in the broad spectrum of symptomatic heart failure patients, with the overall sample size (N = 6,500) equal to the sum of all three trials.[1] With an estimated overall annual mortality in the placebo group of 8% the program had over 85% power to detect a 14% reduction in mortality at two-sided 5% significance based on a logrank test.

All three trials were done at the same 618 sites in 26 countries. The CHARM program exceeded its recruitment goal of 6,500 by enrolling 7,599 patients between March 1999 to March 2001, who were followed for a minimum of two years. Hence, all follow-up was concluded on March 31, 2003, resulting in a median duration of 3.14 years. The final results were

published in the *Lancet* on September 6, 2003.[1-4] For the overall CHARM program , CV death or hospitalization for worsening CHF had a 16% reduction on candesartan versus placebo, 95% CI 9% to 23%, $p < 0.0001$. For all-cause mortality there was a 9% reduction, 95% CI 0% to 17%, $p = 0.055$.

DATA MONITORING EXPERIENCE

The Data Safety and Monitoring Committee (DSMC) had three members, two physicians, Charles Hennekens (chair) and Lars Wilhelmsen, and a statistician, Stuart Pocock. In collaboration with the CHARM Executive Committee, a charter was drawn up, defining the terms of reference, operating procedures as well as guidelines for early termination, which included statistical stopping boundaries. The overarching principle for early termination was proof beyond a reasonable doubt that would be likely to influence clinical practice.

It was agreed that the DSMC would receive a monthly safety report primarily containing data on all serious adverse events and deaths to date. In addition, the DSMC would meet twice a year to evaluate a fuller interim report containing more extensive follow-up data, especially as regards deaths, primary and secondary clinical outcomes, and serious adverse events. Such safety reports and interim reports would present results for the overall program, and also separately for each component trial. The Endpoint Committee verdicts on causes of death and non-fatal major clinical events were used when available, but for events pending Endpoint Committee validation the investigator's classification was used. All six-monthly interim reports and monthly safety reports were produced by a data analyst, Duolao Wang, who was independent of the trial sponsor, Astra Zeneca. Results were presented in a blinded manner, i.e., with coded treatment groups A and B, with the option of unblinding at any stage, i.e., identifying whether candesartan was A or B, if the DSMC thought this was appropriate.[5] Such unblinding in fact occurred at the second interim analysis.

Following each monthly safety report any DSMC member could identify any safety concerns or call for a teleconference or meeting if warranted. No such concerns arose, so each time the DSMC statistician faxed and mailed confirmation to the Executive and Sponsor that the study should continue as planned.

For each six-monthly DSMC meeting there was a closed session attended only by the DSMC members and the independent data analyst. These were the only individuals to see and to discuss any interim results by coded treatment group. At least two DSMC members were always present face-to-face for such meetings, but on two occasions a third member joined by teleconference.

For all but the first DSMC meeting, there was an open session also attended by members of the CHARM Executive Committee and sponsor representatives. Such open sessions were primarily to share information on the study progress and organization. From the fourth interim report onward, a blinded interim report was produced for the Executive Committee containing only the data for both treatment groups combined. The existence of this open session was also helpful should the DSMC have needed to make any recommendations regarding cessation or modification of either the overall program or any specific component trial(s). In fact, no such recommendations needed to be made.

Guidelines for Early Termination

The principle adopted by the DSMC for early termination required a totality of evidence that provided proof beyond a reasonable doubt that would be likely to influence clinical practice. The emerging data would also have to fulfill predefined statistical stopping guidelines.

In the DSMC charter, which was jointly agreed by the DSMC, Executive Committee and sponsor there were no statistical stopping guidelines for the primary efficacy outcome of each trial, i.e., CV death or CHF hospitalization. It was agreed that pre-defined intentions for stopping the program early should focus on all cause mortality.

The Haybittle–Peto rule[6] was employed at each interim analysis, requiring two-sided $p < 0.001$ for the overall program treatment difference in mortality in favor of candesartan using a logrank test stratified by trial. However, two modifications were pre-defined:

1. For each interim analysis occurring within 18 months of the date of first patient's being randomized in the CHARM program, the rule was made more stringent requiring two-sided $p < 0.0001$.
2. Stopping a specific trial required the same trial-specific p-value criteria as above, and also statistical evidence of heterogeneity among trials as regards estimated hazard ratios for mortality of sufficient strength to merit termination of one trial only. In fact, no such statistical heterogeneity arose.

In order to stop for safety (i.e., mortality greater on candesartan) the same general principle applied, except one required $p < 0.001$ for any analysis within 18 months and $p < 0.01$ for any subsequent analysis.

Interim Mortality Results

About three weeks before each six-monthly meeting of the DSMC, a data file was transferred from the sponsor's data management department to the

independent data analyst. He then merged the data with the treatment code to produce the interim report that was then couriered to the DSMC members a few days before the meeting.

Table 1 lists for each interim analysis the numbers of deaths by treatment group, both overall and for each constituent trial, and the overall logrank test P-value. Figure 1 plots the consequent hazard ratio and 95% CI at each analysis. The corresponding results for the final published data are also given.

By the *second interim analysis* in March 2000 there was a substantial difference in mortality overall: 76 deaths on candesartan versus 123 on placebo, with hazard ratio 0.63, 95% CI 0.49 to 0.80, p = 0.0007. The DSMB unblinded themselves as to which treatment was which at this point. The formal stopping boundary p < 0.0001 had not been crossed. A total of 5,800 patients had been randomized since patient entry began one year earlier. There were more deaths in CHARM-Added since patient recruitment was more rapid than in CHARM-Alternative, N = 2,548 and 1,212 respectively. CHARM-Preserved had many fewer deaths because recruitment was somewhat slower than in CHARM-Added (N = 2,040) and its population had lower mortality rates.

The situation was broadly similar at the *third interim analysis* in July 2000, though with 67% more deaths. The magnitude of treatment effect was slightly reduced: hazard ratio 0.66, 95% CI 0.53 to 0.82. But with a larger number of deaths, statistical significance was slightly enhanced at p = 0.0002, still just short of the stopping boundary of p < 0.0001.

At the *fourth interim analysis* in March 2001, there were almost twice as many deaths compared to six months earlier. The overall treatment effect

Table 1 CHARM Mortality Results At Each Iinterim Analysis and At Study Close-Out*

	CHARM-Alternative		CHARM-Added		CHARM-Preserved		Overall Program		
Analysis Date	**C**	**P**	**C**	**P**	**C**	**P**	**C**	**P**	**p-value**
							8	4	0.3
9 Aug 1999	3	0	5	4	0	0	76	123	0.0007
27 March 2000	20	38	45	69	11	16	133	198	0.0002
27 July 2000	39	60	76	113	18	25	260	339	0.0006
1 March 2001	66	100	140	168	54	71	387	474	0.0010
9 Aug 2001	117	148	186	219	84	107	556	631	0.009
22 Feb 2002	166	198	258	285	132	148	682	756	0.015
1 Aug 2002	210	236	298	336	174	184			
Final Report**							886	945	0.055

* Each line gives the number of deaths by treatment group for each constituent trial and overall, plus the overall logrank p-value, stratified by trial (C = candesartan, P = placebo).

** Final report on September 6, 2003, based on follow-up to March 31, 2003.

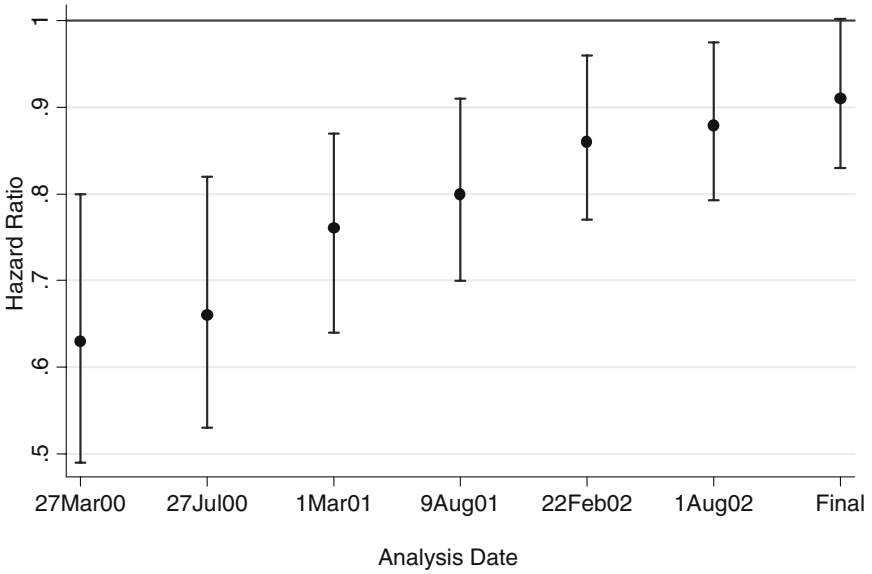

Figure 1 Hazard ratio and 95% for all-cause mortality (candesartan vs. placebo) at each interim analysis and at study close-out.

was further attenuated, hazard ratio 0.76 95% CI 0.64 to 0.87, but being based on more deaths statistical significance was maintained at p = 0.0006. This was well past 18 months from the start of recruitment and hence the stopping boundary p < 0.001 had been crossed. The more rapid recruitment in CHARM-Added (final N = 2,548 completed over a year earlier) meant that it had almost twice as many deaths as in CHARM-Alternative (N = 1,989 with one month of recruitment still to go) while CHARM Preserved had fewer deaths in its lower risk population (final N = 3,023 completed six months earlier). As was the case at all other analyses, there was no statistical heterogeneity in hazard ratios between trials, interaction test p = 0.45. However, it was noted that the treatment difference in mortality only achieved even a conventional level of significance in CHARM-Alternative (66 vs. 100 deaths, p = 0.006) compared with CHARM-Added (140 vs. 168 deaths, p = 0.07) and CHARM-Preserved (54 vs. 71 deaths, P = 0.14).

The DSMC recommended that the program continue without alteration. Thus, as on previous occasions the DSMC requested that the Executive Committee and sponsor ensure that data be as complete as possible for future interim analyses, particular assurance being sought that the Endpoint Committee adjudicate all causes of deaths and major morbid events that had arisen and were available to them.

The DSMC unanimously agreed at this fourth interim analysis that the CHARM program should continue for the following reasons:

1. While the overall mortality result in favor of candesartan reached the statistical stopping guideline, the mortality differences in two of the three component trials did not achieve even a conventional level ($p < 0.05$) level of statistical significance.
2. Data on the primary efficacy endpoint, CV death and CHF hospitalization, were incomplete at this point with many reported endpoints awaiting adjudication by the Endpoint Committee.
3. The average length of patient follow-up was relatively short and one major goal was to evaluate candersartan's effect over two or more years' treatment.
4. There was no previous trial evidence regarding a survival benefit of candesartan, or indeed other angiotensin-receptor blockers, in patients with CHF. In fact, one earlier small pilot trial RESOLVD[2] had shown possible but inconclusive higher mortality on candesartan (with or without enalapril), compared with enalapril alone.
5. The DSMC was mindful of the likelihood that trials that stop early for efficacy are liable to exaggerate the true treatment effect with the danger that people may infer that the observed result is "too good to be true.[8] Aware that from such a potentially "random high" there may well be some "regression to the truth" of a more modest estimated mortality reduction, the DSMC voted unanimously to continue for at least a further six months.

There did not appear to be proof beyond a reasonable doubt of treatment efficacy that would be likely to influence clinical practice.

At the *fifth interim analysis* there was a further attenuation of the mortality hazard ratio now 0.80, 95% CI 0.70 to 0.91 with stratified logrank $p = 0.00103$, so the DSMC felt once again that early stopping was not warranted. There were in fact two more interim analyses, each with less statistically convincing evidence of a mortality difference $p = 0.009$ and $p = 0.014$, respectively, so that it became increasingly straightforward for the DSMC to recommend continuation of CHARM.

Final Results of CHARM

Patient follow-up continued for a further seven months after the last planned interim analysis. Published results were available 5 months later as follows: the numbers of deaths were 886 on candesartan versus 945 on placebo, hazard ratio 0.91, 95% CI 0.83 to 1.00, $p = 0.055$. The predefined secondary analysis adjusting for 33 baseline covariates had hazard ratio 0.90 $p = 0.032$.

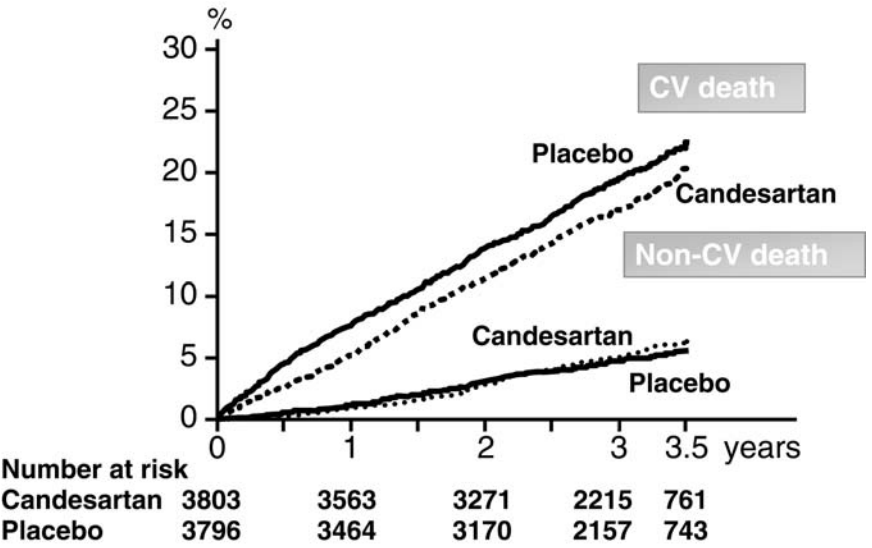

Figure 2 Kaplan–Meier curves for cardiovascular and non-cardiovascular deaths.

This treatment difference could be entirely attributed to cardiovascular deaths 691 versus 769, hazard ratio 0.88, 95% CI 0.79 to 0.97, p = 0.012 as shown in Figure 2. Subgroup analyses revealed no relevant interactions between treatment and baseline features, and there was no evidence of heterogeneity among trials.

Trial continuation to its intended conclusion enabled clear results for the primary efficacy endpoint, CV death or CHF hospitalization. Over a mean 3.14 years follow-up there were 1,150 (30.2%) vs. 1,310 (34.5%) cases, hazard ratio 0.84 95% CI 0.77 to 0.91, p < 0.0001. There was no statistical heterogeneity among trials, interaction p = 0.33, though this efficacy appeared somewhat less pronounced in patients with preserved LV systolic function.

The investigators concluded that "candesartan was generally well tolerated and significantly reduced cardiovascular deaths and hospitalizations for heart failure. The clinical evidence we report . . . offers the opportunity to further reduce cardiovascular mortality and morbidity in this expanding segment of our aging population."[1]

LESSONS LEARNED

The DSMC experience in the CHARM program illustrates the crucial importance of continuing a trial to its scheduled termination unless there emerges evidence of proof beyond a reasonable doubt that would influence clinical practice. Indeed, early termination of CHARM based solely on a

statistical guideline would have been misleading. During March 2000 to August 2002 there were six interim analyses, followed by the final analysis in 2003. For these seven successive analyses the difference in the numbers of deaths (candesartan vs. placebo) were 47, 65, 79, 87, 73, 74, and 59 respectively. Thus, the early mortality difference persisted but was not increased by further follow-up.

The final data indicate that mortality benefit was confined to CV deaths, as one would expect. Closer inspection of Figure 2 reveals that the treatment difference in CV deaths was substantial by one year of follow-up, 199 versus 285 deaths on candesartan and placebo, respectively, an absolute difference of 2.29% mortality. Beyond one year the numbers of subsequent deaths in candesartan and placebo groups are very similar: 492 and 484, respectively, and the estimated absolute treatment difference in CV deaths at 3 years is 2.31%. This indicates that the early benefit in CV mortality reduction attributed to candesartan was maintained but not enhanced by further follow-up. It is worth noting that a similar pattern emerged in the SOLVD trial[9] comparing enalapril and placebo in patients with chronic heart failure: the mortality reduction due to enalapril occurred within 18 months of randomization, with no additional benefit over a further mean two years of follow-up.

This post hoc exploratory finding, if real, offers a plausible explanation as to why the early interim results, based exclusively on short-term follow-up gave the greatest reduction in hazard.

1. The experience of the DSMC in the CHARM program emphasizes the importance of judging early mortality differences in the context of the totality of evidence and not relying exclusively on a statistical stopping guideline when a trial is designed to determine the overall longer-term benefits (if they exist) of a treatment for a chronic condition that is intended to be given for several years.
2. The CHARM experience illustrates the complexity of simultaneously monitoring these inter-related trials in one overall program. In particular, it is difficult to pre-specify a statistical stopping guideline that will correct all contingencies that may arise.

ACKNOWLEDGEMENTS

We are grateful to members of the CHARM Executive Committee for their helpful comments.

REFERENCES

1. Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJV, Michelson EL,et al. 2003. Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall Programme. *Lancet* 362:759–766.

2. McMurray JJV, Östergren J, Swedberg K, Granger C B, Held P, Michelson EL, et al. 2003. Effects of candesartan in patients with chronic heart failure and reduced left-ventricular systolic function taking angiotensin-converting-enzyme inhibitors: The CHARM-Added Trial. *Lancet* 362:767–771.
3. Granger CB, McMurray JJV, Yusuf S, Held P, Michelson EL, Olofsson B, et al. 2003. Effects of candesartan in patients with chronic heart failure and reduced left-ventricular systolic function intolerant to angiotensin-converting-enzyme inhibitors: The CHARM-Alternative Trial. *Lancet* 362:772–776.
4. Yusuf S, Pfeffer MA, Swedberg K, Granger CB, Held P, McMurray JJV, et al. 2003. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved Trial. *Lancet* 362:777–781.
5. Pocock S, Furberg CD. 2001. Procedures of Data and Safety Monitoring Committees. *Am Heart J* 141:289–294.
6. Peto R, Pike MC, Armitage P, et al. 1976. Design and analysis of randomised clinical trials requiring prolonged observation of each patient: I. Introduction and design. *Br J Cancer* 34:585–612.
7. McKelvie RS, Rouleau JL, White M, Afzal R, Young JB, Maggioni AP, et al. 2003. Comparative impact of enalapril, candesartan or metoprolol alone or in combination on ventricular remodelling in patients with congestive heart failure. *Eur Heart J* 24:1727–1734.
8. Pocock SJ, White I. 1999. Trials stopped early: Too good to be true. *Lancet* 353:943–944.
9. SOLVD Investigators. 1991. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med* 325:293–302.

SECTION 3

General Harm