The Management of Heart Failure with Preserved Ejection Fraction (HFpEF)

Andrew J Stewart Coats MD PhD FRCP 1,2 and Louise G Shewan PhD 1,2

1. Monash University, Melbourne, Australia
2. University of Warwick, Coventry, UK

Abstract

Heart failure is a clinical syndrome. Given that as many patients have a preserved ejection fraction as they do a reduced one, we have dramatically underrepresented HFpEF patients in trials. We have only performed 4 trials in HFpEF specifically, and two trials that recruited both HFpEF and HFrEF. When we consider the similarity in outcomes and neurohormonal activation between HFrEF and HFpEF, the huge amount of consistent trial data that we have for the benefits of ACE inhibitors, ARBs, beta-blockers and aldosterone antagonists in HFrEF, and the much more limited number of trials of similar agents showing near statistically significant benefits in HFpEF (even though none of the trials is absolutely definitive) the time has come to use Bayesian thinking; HFrEF is more like HFpEF than anything else, drug classes that work in HFrEF are highly suggestive of being beneficial in the few trials we have, and we should start treating our HFpEF patients with these effective drug classes. We haven’t to date and that is why the fate of HFpEF patients, unlike their counterparts with HFrEF, has not been improved by the last two decades of treatment success in heart failure.

Key words: Guidelines, Heart Failure, HFpEF, Clinical trials

Heart failure as a clinical syndrome

Heart failure always has been a clinical diagnosis. Unlike cancer, or even myocardial infarction, there is no pathological or biochemical test that is either sufficient or necessary to diagnose heart failure. As a result its identification, and therefore treatment, depends subjectively on the willingness or ability of a physician or medical team to call a particular case heart failure. This is of no concern where there is a fair degree of consensus about whether an individual case is or is not heart failure. One can imagine a younger man with a massive myocardial infarction who survives this initial insult and goes home later to be readmitted regularly with pulmonary oedema and on echocardiography is shown to have a dilated, poorly contracting left ventricle. This patient is easily recognised as fitting one of the clinical patterns of what we have for decades called the clinical spectrum of heart failure. He has what we used to call systolic heart failure, and what we now term Heart Failure with reduced Ejection fraction (HFrEF). Luckily for this patient he fits into the type of patient who was recruited into a number of landmark clinical trials conducted over the period from the late 1980’s to the late 2000’s when most of our modern accepted heart failure therapies were proven to reduce morbidity and mortality in major randomised placebo-controlled outcome trials. I say lucky, because another heart failure patient, just as likely to be diagnosed with this condition in the 1970’s is not so lucky now because she (this patient is more likely to be a female patient) is not the typical patient who was recruited into such trials and therefore the proof that modern therapies work for her are not so well established. Such a patient may have a long history of hypertension, a small sized heart chamber, and may have developed left ventricular hypertrophy (and a degree of basal septal hypertrophy that on occasion may generate a gradient or partial obstruction to left ventricular outflow). She may also have a heightened amount of myocardial fibrosis, her diastolic function may be impaired and she may be at risk of atrial arrhythmias and sub-endocardial myocardial ischemia due to vasomotor disturbance and endothelial dysfunction in her coronary vasculature. Some of these processes working alone or more likely in concert may precipitate an episode of marked shortness of breath or peripheral oedema and may necessitate an admission to hospital where she will in all likelihood be diagnosed as having heart failure and be treated with intravenous diuretics with reasonable symptomatic improvement. Our forebears called her a heart failure patient, and she certainly satisfies the modern clinical guideline definitions of heart failure, yet this type of patient was not recruited into to many of the landmark clinical trials of heart failure: think of CONSENSUS, SOLVD, Copernicus, Rales, Merit-HF, CIBIS-II, Ephesus, etc. As a result we don’t know if she will respond to the treatments we offer our first patient and she is largely left untreated. Does this matter? Yes, enormously.

The macro (population study) view versus the micro (individual physician) view

Figure 1 shows patients admitted to European hospitals with a diagnosis of heart failure. As we can see high ejection fractions are just as likely as low, especially in females where they form the majority. Figures 2a and 2 b show the 15 year change in mortality of two sub-types of heart failure, the more common variety similar to our male patient above, with HFrEF and our second patient’s type of heart failure with relatively preserved left ventricular ejection fraction, now more commonly called Heart Failure with Preserved Ejection Fraction (or HFpEF). Two features are notable. The mortality of HFpEF is virtually indistinguishable from that of HFrEF and only for HFpEF has there been any improvement over this 15-year period. This period coincided with one of the most significant advances in the therapy of cardiovascular disease, the revolution in our treatment of chronic heart failure (CHF). In contrast to this macro view, derived from all consecutive patients hospitalized with decompensated heart failure at Mayo Clinic Hospitals in Olmsted County, Minnesota, from 1987 through 2001 without further specialised testing. More recent reports similarly show outcomes as poor for HFpEF as for HFrEF. 1 In contrast the composite micro view looks at patients specifically investigated and chosen to enter clinical trials. Some reports of this nature suggest that survival is significantly better for HFpEF than HFrEF.
when comparing clinical trials, such as DIG-REF with DIGpEF or CHARM preserved versus the two other CHARM studies, some suggest prognostic and pathophysiological factors may be distinct but of course these comparisons are biased by the fact recruitment is much slower for HFpEF trials as many more patients are excluded for co-morbidities and they are older. Thus the population mortality rate of HFpEF is reduced in trials of HFpEF because so many of the higher risk HFpEF patients are excluded to find a somehow “purer” form of HFpEF. Two explanations exist, either the treatment benefit of ACE inhibitors, beta-blockers and aldosterone antagonists are restricted to HFrEF or the fact that HFpEF patients were not included in the landmark trials that proved the benefit of these therapies mean that physicians are similarly not using these agents in HFpEF patients. Of course both factors may be operating to some extent. How did we get into this situation? Early trials such as the DIG trial of digoxin recruited heart failure patients of both HFpEF and HFrEF sub-types (sometimes called the DIGREF trial and the DIG HFpEF trial) and the effects were similar for the types. Later trials, in the interest of increasing event rates, over-recruited HFrEF and many restricted entry to patients with a LVEF less than 45%, 40%, 35% or even lower (25% for Copernicus). This was done to increase mortality rates, but had the effect of leaving HFpEF patients unstudied and hence many years later untreated. We should be able to play catch-up by recruiting patients with HFpEF to later trials. But this proves very difficult.

The bias against identifying HFpEF

We have seen that the major trials have largely been restricted to HFrEF patients. Surely we should now be able to recruit patients to purely HFpEF trials. The problem is that when we try, recruitment proves to be very slow, with clinical trial cardiologists finding a myriad of reasons why a patient is not suitable for inclusion; age, co-morbidities, lack of certainty that the cause of symptoms really is heart failure, so that recruitment is very slow, many trials last for a long time, there are many drop-ins to therapy for the control group and drop-outs for the active group. In the PEP-CHF trial for example by the end of the study, 35% of patients assigned to perindopril and 37% assigned to placebo were taking open-label ACE-inhibitors, reducing the power of the trial quite dramatically. Whereas in epidemiological studies and hospital series the prevalence of HFpEF is high and is usually as frequent or more frequent than HFrEF, in recruitment to clinical trials in heart failure most observers find that recruitment for HFpEF is significantly more difficult to achieve and much slower, as though once doctors look more carefully they find reasons for excluding HFpEF from clinical trials; could the answer be that the macro-view of heart failure diagnosed by symptoms or hospital statistics indicate a high prevalence, whereas taking a micro-view and asking physicians to identify individuals where they look carefully at echocardiography and other investigations, they find a myriad of other causes for the patient’s symptoms and use these to call the condition something other than heart failure? In that case is the macro-view wrong or is it simply that we are asking too much of disease classification that groups a variety of patient types together to then expect the same clinical trial certainty when we split them into smaller and smaller sub-sets when we come to performing clinical trials. It can be argued that no heart failure sub-set is a clinically defined and homogeneous group, even HFrEF includes ischaemic and non-ischaemic, metabolic and infiltrative, aneurysmal left ventricles, hypertrophied ventricles and any manner of pathophysiological sub-types, yet we are happy to group them together and treat them as the same just because we did so once when we designed a trial and the trials largely confirmed benefit and so we never subsequently looked to see if the benefit differed between the sub-types. If we had only ever recruited heart failure patients into the major trials irrespective of LVEF we would have found treatments worked and then we would have recommended therapy irrespective of LVEF. We have created this problem for ourselves. If we ask: is there is any evidence in those trials that did recruit heart failure patients with a range of LVEF’s that there is a distinctly different pattern of response for the subset with HFpEF, then the answer there is no. Both the DIG trial and The SENIORS trial had both types and the headline effect was similar and there was no evidence.
of a diametrically different response; i.e. we caused this problem and we are to blame for why we have not improved the outcome of half the heart failure population in the world we ignored when conducted trials with them excluded and as a result we now don’t know what to recommend for them.

The effect of trials that excluded HFpEF

We have seen that the major trials have largely been restricted to HFrEF patients. Surely we should now be able to recruit patients to purely HFpEF trials and use these trials to assess what is effective therapy. The problem is one of statistical power and of priority for funding. Take the example of one of the most successful drugs in the treatment of heart failure, carvedilol. Carvedilol is now off patent in most developed countries so further company sponsorship of large expensive trials is unlikely. The sponsors did however pay for three trials, the US Carvedilol program\textsuperscript{15}, Copernicus and COMET\textsuperscript{16}. One might have expected that out of three trials we could have covered the heart failure spectrum, but this did not happen, all three were restricted to those with HFrEF, all three recruited younger patients and one was more a marketing trial (comparing one beta blocker carvedilol with another metoprolol) rather than addressing gaps in our knowledge of how best to treat heart failure. We cannot, therefore, depend on sponsors doing the best for the whole community of heart failure sufferers, they focus where their drug will look best and avoid the more difficult or uncertain areas. If we had recruited patients with heart failure irrespective of LVEF in the Copernicus trial of carvedilol and pre-defined LVEF as a stratification variable and the principal sub-group analysis we would have had sufficient information by now to evaluate carvedilol’s benefits in heart failure as clinically identified and we would have had testing available to see if LVEF was a significant predictor of response to carvedilol and whether indeed there was evidence of a similar, greater, lesser or no benefit in HFpEF as now defined. By excluding patients with LVEF’s above 25\% in Copernicus we are missing crucial data, and we now need another trial, which has never been performed and in all likelihood never will be because the drug is off patent. I admit to some responsibility as I was on the steering committee of Copernicus, yet the pressure to optimise the result of the trial by going for the sweet spot where treatment effect was felt to be easiest to show (younger patients with severe disease and few comorbidities) we have done the more typical heart failure community a disservice by choosing an odd subset of patients to test rather than the typical older more complicated patient. Also by having to omit those in whom a treatment effect is easiest to see, any later trial, if it is ever done, is made more likely to be inconclusive because precisely those patients are excluded at the outset.

The trials that did recruit HFpEF patients

There are two types of trial to recruit HFpEF patients, those that did what I argue should be standard: recruit heart failure and stratify and analyse for subsets, and those that are playing catch up by recruiting solely within a HFrEF cohort. There have been two of the former type with mortality and morbidity predefined end-points: DIG and SENIORS and three of the latter type, I-Preserve\textsuperscript{17}, PEP-CHF\textsuperscript{18} and TOPCAT (see later). One trial CHARM-Preserved\textsuperscript{19} is a type of hybrid. The CHARM programme of Candesartan has its cake and eats it too. It is called three trials that were combined together with a single powered endpoint and recruited and analysed together. It thus could be thought of as a single trial with a HFpEF subset (CHARM-preserved\textsuperscript{20}) or three trials, one of which CHARM-preserved, is in HFpEF.

What do the results show? Now this is where the pre-test probability comes in. There is a form of analysis called Bayesian that states your proof of certainty versus chance depends of the pre-test likelihood, i.e. you do not asses every question in the same way or with the same necessary level of statistical certainty, if you have a lot of background to suggest a treatment may work you have a different level of proof compared to your assessment of something where you have no prior assumption. In other words don’t ignore the prior related data.

For the DIG trial the main trial was positive for hospitalisation and neutral for mortality. Digoxin is recommended to prevent hospitalisations in HFrEF. The results of HFpEF were identical in all essences to that of HFrEF, yet its recommendation and uptake is much more muted. For SENIORS the overall trial was positive and the subset with preserved LVEF did just as well, there was no statistically significant interaction between LVEF and treatment effect yet guidelines fail to recommend nebivolol other than for lower EF even thought this is based on an analysis of a subset of the pre-specified question and they maintain that because the HFpEF subset was not independently significant you cannot recommend nebivolol for this cohort, even though we know that the correct statistical analysis is to assume any subset behaves as the whole cohort unless you have a reason or a statistical suggestion that it does not. If you test a cohort of patients and reach your primary end-point and there is no evidence of a significantly different effect in any subset then the results of the subset are best estimated by the overall trials results. Nebivolol should therefore be recommended for elderly heart failure patient irrespective of LVEF. None of the guidelines follow this logic, and in doing so are themselves illogical. CHARM I have mentioned above could be considered in the same way. Indeed the CHARM-preserved even if taken in isolation and ignoring Bayesian logic, is very close to being positive. 3023 HFpEF patients were randomised to candesartan (n=1514, target dose 32 mg once daily) or placebo (n=1509). The primary outcome of cardiovascular death or heart failure hospitalisation was very nearly significantly reduced, (unadjusted hazard ratio 0.89 [95\% CI 0.77 – 1.03], p=0.118; covariate adjusted 0.86 [0.74 – 1.0], p=0.051). Cardiovascular death did not differ but heart failure hospitalisations were significantly reduced. Whether we consider candesartan effective in HFpEF depends solely on how we interpret the CHARM programme. If we consider it as a combined programme the overall trial was positive and the HFpEF subset behaved similarly. If we consider it as a separate trial then it just failed to reach significance. Both ways ignore the prior evidence that Angiotensin receptor blockers (ARB’s) benefit heart failure.

The trials that restricted their recruitment to HFpEF patients were not testing entirely unproven treatments; they were all testing drugs or drug classes that had been shown to be effective in the very closely related condition HFrEF. I say closely related because the only differentiating factor between two patients may be a tiny difference in a highly variable measurement on one day, that of LVEF one being above 45\% and one below. Of the three trials, one chose the ARB, Irbesartan. ARB’s are considered effective as a class in CHF, and as alternatives to ACE inhibitors. One chose the ACE inhibitor perindopril that yet again these are considered as a class to be effective in CHF, and one chose the aldosterone antagonist spironolactone, that as an agent and as a member of a class is considered to be effective in CHF. So the a-priori hypothesis to be tested had a positive pre-test probability by Bayesian analysis. This was not factored into the analysis.

PEP-CHF was a randomized double-blind trial, comparing placebo with perindopril, 4 mg/day in patients aged >70 years with a diagnosis of heart failure, and echocardiographic evidence of diastolic dysfunction and excluding substantial LV systolic
dysfunction or valve disease. The primary endpoint was a composite of all-cause mortality and unplanned heart failure related hospitalisation. 850 patients were randomised and followed-up for an average of 2.1 years. The power of the study to show a difference in the primary endpoint was reported to be only 35% (because of poor recruitment and lower than expected event rates) showing only a one third chance of showing an effect event if a real effect were present. Overall, 107 patients assigned to placebo and 100 assigned to perindopril reached the primary endpoint (HR 0.919: 95% CI 0.700–1.208; P = 0.545). By 1 year, before the extent of loss of adherence to randomised drug groups had become so catastrophically high as mentioned earlier, the reductions in the primary outcome (HR 0.692: 95% CI 0.474–1.010; P = 0.055) and hospitalisation for heart failure (HR 0.628: 95% CI 0.408-0.966; P = 0.033) were observed and functional class (P = 0.030) and 6-min corridor walk distance (P = 0.011) had improved in those assigned to perindopril.

I-Preserve similarly was a randomised double-blind placebo-controlled trial in HFpEF, but in this case was much larger. 4128 patients 60 years or older and LVEF > 45% were randomised for an average of 49.5 months to 300 mg of irbesartan or placebo. The primary endpoint was death or CV hospitalization. The primary outcome occurred in 742 patients in the irbesartan group and 763 in the placebo group, giving primary event rates of 100.4 and 105.4 per 1000 patient-years respectively (HR, 0.95; 95% confidence interval [CI], 0.86 to 1.05; P = 0.35). The mortality rates were similar. This result seems disappointing but it is directionally and in scale not dissimilar to the result of VAL-HeFT of Valsartan in HFrEF where in 5010 patients 160 mg of valsartan reduced the primary mortality/morbidity endpoint, by 13.2 percent (relative risk, 0.87; 95% confidence interval, 0.77 to 0.97; P =0.009), with no difference in mortality. Also in I-PRESERVE there was a high rate of discontinuation of study treatment (34% by the end of the study) and a high rate of concomitant use of ACE inhibitors, spironolactone, and beta-blockers.

The most recent trial, TOPCAT35, has not yet been published in full manuscript form but the main results were presented at the American Heart Association meeting November 201333. It builds upon earlier smaller trials, investigating another HFpEF proven treatment in HFrEF patients36. TOPCAT randomized 3445 patients 50 years or older and LVEF > 45% to spironolactone to 30 to 45 mg/day or placebo. The trial was not quite positive, the primary composite endpoint was reduced from 20.4% to 18.6% (HR 0.89 (95% CI’s 0.77–1.04) p=0.138 and heart failure hospitalisations reduced from 14.2% to 12.0% (HR0.83, 95%CI’s 0.69-0.99). Yet again a negative trial but in its pattern not dissimilar to VAL-HEFT. Much discussion centred on a difference in outcome between western countries and eastern countries: in four Western-hemisphere nations (51% of the trial) the Hazard ratio was 0.82 (95% CI 0.69–0.98), compared to 1.10 (95% CI 0.79–1.51) for the eastern European countries. This was, we must remember, only a post-hoc analysis but it does resonate with a similar trend in I-Preserve. The country issue is a distraction, but it may be a surrogate for something more important; what matters is whether there was a reason based on identifiable patient characteristics that the trial was not positive. There is concern expressed that HFpEF has no confirmatory lab test and patients may be recruited with little evidence they have a serious condition. In what was both a pre-specified analysis and using a variable that was actually stratified for at randomisation (ensuring the likelihood of good balance between placebo and active) in those patients who qualified for TOPCAT on the basis of an elevated natriuretic peptide level (BNP ≥100 pg/mL or NT-proBNP ≥360 pg/mL) there was a highly significant 35% reduction in the primary endpoint. In the elevated NP group there were 78 primary events in 490 patients (15.9%) compared to 116 events in 491 placebo patients (23.6%). HR 0.65, 95CI’s 0.49-0.87 p=0.00335, entirely consistent with what has been seen in HFrEF with spironolactone or eplerenone. There were more low NP patients in the east. The less intensively investigated and managed patients in the eastern countries may simply include too many patients without “real” CHF.

It is hard to avoid the conclusion we should investigate36,7,28 and treat HFpEF as rigorously as their HFrEF counterparts.

What can we conclude?

We have a problem of our own making. We chose the easy route to study only HFrEF patients when there were many drugs under testing and large trial budgets. Now the patents are expired we are playing catch up with slow and inadequately powered trials. We are ignoring pre-test likelihood that they work and artificially demanding an excessive level of proof. All the trials reviewed above using agents that work in the very similar condition of HFrEF are numerically beneficial, none even trend to harm in their primary endpoint. The results are close to being significant for many useful end-points. Based on the common sense test of what you would take yourself or recommend to a patient with HFrEF they have a problem of our own making. We should investigate26, 27, 28 and treat HFpEF as rigorously as their HFrEF counterparts.

Table 1: Diagnosis of Heart Failure (ESC-May 2012)

- An abnormality of cardiac structure or function leading to failure of the heart to deliver oxygen at a rate commensurate with the requirements of the metabolising tissues, despite normal filling pressures (or only at the expense of the increased filling pressures)
- Or more simply typical symptoms and signs resulting from an abnormality of cardiac structure or function.

References

1. Senni M1, Gavazzi A1, Oliva F2, Mortara AS, Urso R4, Pozzoli M5, Metra M6, Lucci D4, Gonzelin LA, Cirincione V7, Montagna L8, Di Lenarda A9, Magniion AP10, Tavazzi L5; On the behalf of the IN HF Outcome Investigators. In-hospital and 1-year outcomes of acute heart failure patients according to presentation (de novo vs. worsening) and ejection fraction. Results from IN-HF Outcome Registry. Int J Cardiol. 2014 Feb 22. pii: S0167-5273(14)00371-4. doi: 10.1016/j.ijcard.2014.02.018. [Epub ahead of print]
2. Campbell RT, Jhund PS, Castagno D, Hawkins NM, Petrie MC, McMurray JJ. What have we learned about patients with heart failure and preserved ejection fraction from DIG-PEF, CHARM-preserved, and I-PRESERVE? [Review] Journal of the American College of Cardiology. 60(23):2349-56
4. Maeder MT, Kaye DM.


13. The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. European Heart Journal (2012) 33, 1787–1847 doi:10.1093/eurheartj/ehs104
