# Research in acute decompensated heart failure: challenges and opportunities

## Clin. Invest. (2011) 1(10), 1361–1373

Acute heart failure continues to increase in incidence and prevalence, with minimal improvement in mortality. We have access to vast amounts of patient information (both research databases and administrative/clinical records) from which to derive associations and generate hypotheses. However, despite solid groundwork, rationale hypotheses and firm preliminary studies, the heart failure literature has not witnessed substantial successes when it comes to clinical trials. This article discusses recent acute heart failure trials and advances suggestions as to why these trials have not been able to demonstrate significant differences. In the area of registry-based studies, the article discusses potentially useful data management techniques that ameliorate two of the common criticisms of retrospective studies – missing data and the correlation versus causation conundrum.

Keywords: acute heart failure • multiple imputation • propensity score

The burden of heart failure continues to increase, with a population prevalence of nearly six million patients in the USA and an annual incidence that exceeds 500,000 new cases per year [1]. Mortality remains high after the index diagnosis, with a 1 year mortality rate of 20% and a 5-year mortality of 50% [1], survival rates that are worse than many malignancies [2]. Heart failure is the single most costly diagnosis for the Center for Medicare/Medicaid Services in the USA. Acutely decompensated heart failure, defined as the rapid worsening or accumulation of heart failure symptoms [3], is especially problematic, as patients experiencing an episode of acute heart failure usually require inpatient admission, suffer prolonged hospital stays, face high inpatient mortality (4–7%) [4,5], and are quite likely to have a recurrence within 30 days (26.9% readmission rate) [6].

The recent literature base regarding the clinical management of decompensated heart failure management has been predominantly comprised of two types of studies – analyses of massive retrospective databases collected from inpatient stays, and interventional clinical trials studying a drug, device or management strategy in a prospective fashion. Both approaches, although designed to answer substantially different questions, can yield important information and insights into heart failure management. However, both study design types have pitfalls that can bias the results of the analysis, or even curtail the probability of finding any reason to exclude the null hypothesis. In this paper, I will explore statistical methods that can increase the validity of conclusions drawn from observational studies, including the use of propensity scores to adjust for imbalances between subgroups within the sample, and multiple imputation techniques to account for missing data elements. In addition, I will examine several recent clinical trials of acute heart failure therapeutics that have failed to establish their primary hypothesis, with an eye to common research strategies undertaken by these study groups that may have led to the systematic failure of these studies.

## Brian C Hiestand<sup>1</sup>

<sup>1</sup>Department of Emergency Medicine, Medical Center Boulevard, Wake Forest University, Winston-Salem, NC, 27157, USA E-mail: bhiestan@wfubmc.edu

In terms of the outpatient management of chronic heart failure, there have been several successful clinical trials that have led to an increase in the number of tools validated for the outpatient stabilization and treatment of patients with chronic heart failure. Studies such as COMET, CHARM, EPHESUS, the various MADIT trials, and CARE-HF have demonstrated the potential for  $\beta$ -blockade [7], angiotensin-receptor blockade [8], aldosterone antagonism [9], prophylactic defibrillator placement [10,11] and cardiac resynchronization therapy [12,13] to have a generally beneficial impact on hospitalization-free survival in patients with heart failure. These studies have led to solid, evidence-based guidelines for the management of chronic heart failure. However, these studies, and the guidelines they inform, reference a stable outpatient heart failure population.

Such relative stability is, of course, antithetical to the concept of decompensated heart failure. The patient with acute heart failure is buffeted by several compensatory physiologic mechanisms overzealously performing their roles in an attempt to increase cardiac output. The aldosterone-angiotensin axis works to retain sodium and water so as to increase preload, forcing the ventricles further along the Starling curve to generate improved contractility out of a dysfunctional pump. Adrenergic tone increases, generating an increase in stroke volume and heart rate, but increasing afterload and contributing to increased arrhythmogenicity. Pulmonary congestion leads to dyspnea, the most frequent presenting complaint with acute heart failure. With this storm of derangements buffeting the patient's cardiopulmonary physiology, it is small wonder that the admission rate for acute heart failure presenting to the emergency department (ED) approaches 80%.

## Initial presentation: heterogeneity

A useful paradigm for examining acute heart failure, initially proposed by Nohria [14], asks the physician to simultaneously consider the patient's volume status and perfusion status (Figure 1). Although modestly simplistic, and with the acknowledgement that patients present along a continuum as opposed to presenting within neat categories, this thought exercise allows the physician to consider the clinical needs of the patient as dependent upon the pathology leading to the decompensation. In terms of volume, patients can be wet (volume overloaded) or dry (euvolemic, or occasionally hypovolemic due to excessive diuresis). From a perfusion standpoint, patients can be cold (hypotensive with poor systemic circulation) or warm (normo- or hypertensive with sufficient circulation to the periphery). The 'warm and dry' patient is generally not found to be in acute heart failure - the presence of adequate perfusion and absence of signs of volume overload should prompt the physician to search for other causes for the patient's symptoms other than acute heart failure.

The 'warm and wet' presentation describes the majority of patients presenting for acute care due to heart failure [4,5]. These patients maintain sufficient peripheral perfusion, but will frequently present with substantial worsening of their dyspnea due to an increase in total lung water. In addition, diastolic dysfunction and impaired ventricular filling may play a significant role. These patients may not necessarily be volume-overloaded - rather, due to high afterload the problem lies with volume distribution into the lungs, even though the body as a whole may be euvolemic. Note that hypertension can be a relative term - a patient with an ejection fraction of 15% may be overwhelmed by a systolic blood pressure of 150 mmHg. These patients benefit the most from acute reduction of their afterload so that flow can move forward from the left ventricle and decongest the pulmonary tree.

Conversely, a patient with well compensated and well tolerated afterload may still present with decompensated heart failure due to pure volume overload. Weight gain, peripheral edema and a subacute worsening of symptoms are hallmarks of this presentation. These individuals will benefit from a regimen that emphasizes volume reduction via diuresis. As well as screening for worsening renal function as a contribution to the fluid accumulation, modestly increased doses of their home diuretic, the addition of thiazide diuretics to achieve a synergistic effect, or even ultrafiltration, may be indicated for these patients.

As the failing ventricle loses its ability to compensate, even with the pathologic increase of renin-angiotensin activity and sympathetic tone, poor peripheral perfusion and symptomatic hypotension will result. Altered mental status, cool, clammy extremities, and other symptoms of shock may be noted. In the setting of hypoperfusion, inotropic agents (dobutamine and milrinone) are indicated. For severe hypotension and shock, vasopressors (dopamine and norepinephrine) may be required even though afterload will inevitably increase. Even in the volume overloaded patient ('cold and wet') diuretics should be used cautiously; paradoxically, the over-diuresed patient in shock ('cold and dry') may benefit from judicious fluid administration. Bedside ultrasound may be useful to guide the assessment of fluid status by examining inferior vena caval dimensions and ventricular filling [15,16].

The European Society of Cardiology (ESC) released a set of guidelines for the diagnosis and treatment of heart failure in 2008 that incorporates a classification scheme acknowledging the diversity of presentations of decompensated heart failure [17]. However, the ESC

# Research in acute decompensated heart failure Clinical Trial Perspective

paradigm is less of a guide for a potential therapeutic approach as it is a recognition of different etiologies and overlapping symptoms that can characterize acute heart failure. Six broad modal domains are presented as potential descriptors: hypertensive acute heart failure, acutely decompensated chronic heart failure, pulmonary edema, right-sided heart failure, heart failure and acute coronary syndrome, and cardiogenic shock. While this classification scheme lacks the specificity and therapeutic guidance provided by the perfusion/congestion framework previously discussed, the ESC model certainly highlights the lack of a single pathophysiologic process that constitutes 'acute heart failure'.

Given the heterogeneity of patient presentations with acute heart failure [4,18], it would stand to reason that

therapeutic interventions would need to be targeted for the patient profile most likely to benefit from them. For example, it has been clearly demonstrated that hypotension is a substantial risk factor for inpatient mortality [4,19] and that patients requiring inotropic/vasopressor support fare worse than those patients requiring afterload reduction [20]. A research protocol that does not acknowledge the heterogeneity of acute heart failure and fails to target appropriate patient 'phenotypes' based on the mechanism of action of the intervention is more likely to discover substantial variance in the response to therapy by patient subgroups, and have a difficult time of it when attempting to reject the null hypothesis.

A three-stage model of the course of therapy of acute heart failure has been proposed, along with potential therapeutic targets and goals for each phase [21,22]. Stage A consists of the first 48 h or so, when the goal of treatment is hemodynamic stabilization and symptom control. This is largely successful, as up to 70% of patients report improvement in dyspnea symptoms in the first few hours of therapy [23]. Increasingly, the task of acute stabilization and, therefore, stage A management, falls to the ED. As stated previously, nearly 80% of all hospitalizations for acute heart failure come through the ED [24]. In addition, as hospital crowding and ED boarding result in prolonged lengths of stay, more care is being provided in the ED as opposed to being deferred to the inpatient floor. It has been demonstrated that, among patients requiring intravenous vasoactive medications (including vasodilators, vasopressors and inotropes), early initiation of treatment is





associated with substantially improved patient hemodynamics, symptoms, length of stay, ICU utilization and in-hospital mortality [25]. As treatment is effected and patients stabilize, the heterogeneity of the different phenotypes will begin to settle out as the patients clinically improve and approach the asymptote of their baseline status. Logically, the more the patient improves clinically, the more difficult it will be to demonstrate a substantial treatment effect due to an investigative therapy or strategy if it is implemented after most of the patient's immediate symptoms have improved.

This leads to stage B, where the therapeutic goal switches from 'stabilization' to 'optimization'. Many patients in this stage who still have symptoms will consist of that subset of patients who are refractory to initial management. A protocol that allows delayed enrollment but also requires ongoing symptoms (e.g., dyspnea at rest) will preferentially sample from this subpopulation. Interventions during this phase of treatment include those that could decrease length of stay, ICU length of stay and in-hospital mortality, and may set the stage for initiating therapies such as  $\beta$ -blockers and angiotensin converting enzyme (ACE) inhibitors that have known benefit on long-term outcomes. Those same long-term, post-discharge outcomes and interventions represent stage C - the outpatient management of the heart failure patient.

Unfortunately, many interventional trials of agents that should have the majority of their effect in stage A (stabilization) are conducted in a stage B population. Two considerations come to mind – first, these patients that are still symptomatic after initial stabilization may suffer from more severe disease and their asymptotic maximal improvement, even with optimal therapy, may simply not get that much better. Therefore, effect size from an intervention in this subgroup may be small, and therefore hard to detect. Second, what if there is a narrow therapeutic index with the experimental treatment? For patients with refractory symptoms and profound decompensation, a treatment with a number of deleterious side effects may have an acceptable risk: benefit ratio, as there would be scant else to offer them. I will take this analogy to the extreme by considering the use of ventricular assist devices (VADs) as destination therapy in end-stage heart failure - the risk of stroke, hemorrhage and infection is counterbalanced by the absence of alternatives, and the literature has demonstrated an improvement in hospitalization-free survival and improvement in quality of life with VAD use. However, no one is advocating VAD implantation in the ED for routine acute heart failure. Again, an extreme (if not slightly absurd) analogy - but what if the risk was not hemorrhagic stroke or overwhelming sepsis resulting from a VAD, but renal insufficiency with large doses of diuretics? Or ventricular dysrhythmia with inotropes? A treatment strategy with a risk:reward ratio that is acceptable in patients with refractory symptoms may prove to be unacceptably hazardous in a population that could yet respond to other interventions. I would ask the reader to keep this concept in mind as I review some of the recent clinical trials in heart failure.

## **Recent clinical trials**

The EVEREST program investigated the effect of tolvaptan, a vasopressin-receptor antagonist, on a variety of clinical outcomes in patients admitted for acute heart failure [26,27]. The proposed mechanism for vasopressin antagonism in heart failure is to increase free water excretion at the level of the distal nephron, with a concomitant decrease in edema and improvement in hyponatremia. Enrollment could occur up to 48 h after presentation, and there were no restrictions on therapies used prior to, or during, the study. Patients were required to have persistent peripheral edema, jugular venous distension, or dyspnea (at least two of the three) at enrollment, and the trial was limited to those with a known ejection fraction <40%. The study program consisted of two phases - a short-term outcomes study examining inpatient and immediate response to tolvaptan [26], and a 6 month follow-up study [27]. In the short-term outcomes arm of the study, despite an early improvement in clinical parameters (Jugular venous distension, peripheral edema, rales and dyspnea) with tolvaptan, there was no difference in the primary end point of overall well-being, as assessed by a visual analog scale (VAS) at discharge or day 7. Inpatient adverse events and mortality were similar between the tolvaptan and placebo arms. In the longer term outcomes companion trial, where patients were maintained on tolvaptan or placebo for at least 60 days, there was no evidence of mortality or heart failure-related morbidity improvement with tolvaptan at 6 months.

In VERITAS, researchers examined the effects of tezosentan, a short-acting endothelin-receptor antagonist, on heart failure outcomes in a multinational population of patients hospitalized with acute heart failure [28]. Tezosentan's mechanism of action is via vasodilation that decreases systemic vascular resistance and increases cardiac output. Patients could be enrolled up to 24 h after presentation, and were required to still have refractory dyspnea at rest despite the administration of at least one dose of an intravenous diuretic. Although hypotension was an exclusion criteria (SBP <100 mmHg or <120 mmHg if on a vasodilator infusion), there was no requirement for hypertension to be present for enrollment. The median interval from enrollment to study drug (tezosentan vs placebo) administration was 11 h - no description of the range of this time interval (e.g., interquartile range, standard deviation) was provided. The primary outcome in the study was change in dyspnea from baseline, measured using a VAS. In the 2161 subjects randomized to placebo or tezosentan, dyspnea rapidly declined from baseline in the first 3 h and continued to decline substantially by 6 h in both groups. No difference between placebo and tezosentan was noted for dyspnea resolution, 7 or 30 day mortality, or hospital length of stay. Due to perceived futility this trial was stopped early.

In the PROTECT study, patients hospitalized with acute heart failure were randomized to placebo or rolofylline, an adenosine A,-receptor antagonist designed to preserve renal glomerular blood flow, enhance urine sodium excretion and enhance diuretic efficacy [29]. Patients had to have ongoing dyspnea at rest despite diuretic therapy, and could be enrolled up to 24 h after presentation. Due to the mechanism of action of rolofylline, enrollment was restricted to patients with an estimated creatinine clearance of 20-80 ml/min. The primary end point was improvement in dyspnea at 24 and 48 h, with a number of secondary and safety end points collected. Over the course of the trial, 2033 patients were randomized and followed out to 6 months. No mention was made of the interval between presentation and initiation of study drug. Although patients receiving rolofylline lost marginally more weight by day 4 (3.0 vs 2.6 kg), there was no difference in clinical improvement (odds ratio for improvement vs no change vs worsening heart failure = 0.92; 95% CI: 0.78-1.09). Mortality and heart failure readmission were similar at 7 days and up to 180 days after enrollment.

The DOSE trial was designed to examine both furosemide dose and method of administration in patients admitted with acute heart failure [30]. Patients underwent a dual randomization strategy, first to dose (total daily furosemide dose vs 2.5-times the patient's daily furosemide dose) and then to continuous intravenous infusion versus twice-daily dosing. Patients could be enrolled up to 24 h after presentation, there was no restriction on the amount of diuretic patients could receive prior to enrollment, and patients requiring intravenous vasodilators were excluded. Patients with baseline renal insufficiency (serum creatinine >3.0 mg/dl) were also excluded. The primary efficacy end point was the serial change in overall wellbeing as measured on a VAS (0-100 mm), with a secondary efficacy end point examining change in dyspnea on a similar scale. The primary safety end point was the overall change in creatinine, with a secondary safety end point of the proportion of individuals with a change in creatinine >0.3 mg/dl. Given that diuretics are routinely given as standard of care, there was no placebo arm to the trial.

A total of 308 subjects were enrolled, assigning each combination of characteristics (e.g., low dose/bolus) to approximately 150 patients. The median time from presentation to enrollment was 14.6 h. Neither comparison group (low vs high dose, bolus vs continuous) met statistical significance for the primary end point of improvement in overall wellbeing, although the high-dose regimen did achieve a marginally greater reduction in dyspnea than the low-dose group, as well as a greater weight loss and urine output. These findings, however, did not translate into a reduction in length of stay or days free from hospitalization at follow-up. Although the mean serum creatinine change did not vary between groups, the highdose regimen did see a higher proportion of patients with a serum creatinine change >0.3 mg/dl (23% in the high-dose group vs 14% in the low-dose group).

The ASCEND–HF study enrolled subjects hospitalized with acute heart failure, randomized in a doubleblind fashion to placebo or nesiritide bolus and infusion for at least 24 h and up to 7 days [31]. The primary short-term outcome was assessment of dyspnea at 6 and 24 h after initiation of treatment (on a seven-point Likert scale), and the primary long-term outcome was 30 day death or heart failure readmission. Secondary end points, including assessment of wellbeing, safety metrics (including renal function and hypotension) and 6 month outcomes were also recorded. Enrollment was allowed up to 48 h after hospitalization, and subjects could have received up to 24 h of treatment with intravenous diuretics, inotropes or vasodilators.

At the conclusion of the study, 7141 subjects in 30 countries had been randomized. Approximately 95% of the subjects in both arms received diuretics prior to

enrollment, 4% received inotropes and approximately 15% had received vasodilators. The median time to enrollment was 15.5 h. The study failed to meet prespecified statistical significance for improvement in dyspnea at 6 and 24 h with nesiritide. In addition, there was no improvement in mortality, heart failure readmissions or days alive out of the hospital at 30 days with nesiritide use. There was, however, no worsening of renal function associated with nesiritide use, although the incidence of both symptomatic and asymptomatic hypotension was greater in the nesiritide arm. Interestingly, there was a statistically significant trend in improved dyspnea reduction noted based on time to treatment - the sooner treatment was started after hospitalization, the more likely nesiritide was to improve dyspnea at 6 h. This, I believe, is the key finding from the ASCEND-HF study, and provides a proof-of-concept for the rationale that studying acute heart failure interventions needs to occur while the disease is still acute, and not once stabilized.

## Clinical trials summary

It is apparent that multiple studies of potential heart failure therapeutic drugs, all of which demonstrated promise in preliminary investigations, have attempted to study stage A drugs in stage B populations, with unimpressive results (Table 1). This paper is certainly not the first to recognize this unfortunate trend - multiple manuscripts have been released on the subject [21,22,32]. In August 2009, the National Heart, Lung and Blood Institute convened a meeting of leading US cardiology and emergency medicine specialists in acute heart failure research to discuss the current state of acute heart failure research, and opportunities for improving the research process for this disease state [32]. Multiple recommendations ensued, but the primary message coming from the consensus group was that, in order to effectively study acute heart failure, heart failure trials should explicitly be designed to incorporate the ED setting for conducting the trial. Identification, recruitment and initiation of study interventions in the ED setting should be incorporated in research protocols. While the heterogeneity of presentations in the ED setting will likely be greater, selection of patients by phenotype characteristics easily available in the immediate evaluation (blood pressure, oxygenation, physical exam characteristics, natriuretic peptides and cardiac troponin elevation) can better match research subjects with the interventions that are likely to positively affect their particular pathophysiologic profile.

## **Database studies**

Although considered the reference standard for evaluating new therapeutic interventions, the prospective, double-blind, placebo-controlled, randomized trial

# Clinical Trial Perspective Hiestand

Table 1. Recent clinical trials in acute heart failure.					
Study	Comparison	Target population	Enrollment window (h)	End points	Ref.
EVEREST	Tolvaptan vs placebo	EF <40% NYHA III–IV Presence of congestion	48	Global clinical status/change in bodyweight at 7 days All-cause mortality Time to first HF hospitalization/cardiac death	[27,26]
VERITAS	Tezosentan vs placebo	SBP >100 mmHg Need for diuretic	24	Change in dyspnea over 24 h Death or worsening heart failure at 7 days Death or major cardiovascular event at 30 days	[28]
PROTECT	Rolofylline vs placebo	Persistent dyspnea Elevated BNP Impaired CrCl	24	Composite clinical improvement at 24 and 48 h in the absence of adverse events All-cause mortality or readmission at 60 days	[29]
DOSE	High vs low-dose furosemide	Presence of congestion SBP >90 mmHg No vasodilator therapy	24	Global symptom score at 72 h Change in creatinine at 72 h Death, rehospitalization or ED visit at 60 days	[30]
ASCEND– HF	Nesiritide vs placebo	Dyspnea and objective evidence of heart failure No hypotension No uncontrolled hypertension	48	Change in dyspnea at 6 and 24 h All-cause mortality at 30 days	[31]
BNP: B-type natriuretic peptide; CrCI: Creatinine clearance; ED: Emergency department; EF: Ejection fraction; HF: Heart failure; NYHA: New York Heart Association; SRP: Systolic blood pressure					

is not the only means of increasing our knowledge. Indeed, there are some research questions where placebo interventions, randomization or blinding is unethical or not feasible. Cohort studies and case-control studies are valuable research designs that can inform epidemiologic data, generate hypotheses, and describe associations between predictive factors and outcomes. The ADHERE and OPTIMIZE HF registries, immense prospective registries examining decompensated heart failure in hospitalized patients, and IMPROVE HF in outpatients, have provided numerous advances in our knowledge of heart failure [5,24,25,33-41]. It has also been established that patient characteristics described in registries are more representative of the 'real world' patients than those in clinical trials. Patients in clinical trials tend to be younger, predominantly male and with few active co-morbidities, whereas patients in heart failure registries tend to be older, female and suffering from multiple medical conditions [5]. Comparisons with Medicare claims data suggest that registry patients are substantially more reflective of the general population [42,43].

Observational data, such as that which the investigator had no control over interventions chosen, may come from a number of sources. Databases collected with an intent to perform research generally provide the richest source of information, but other sources of aggregate data may also be useful, such as administrative or billing data [44]. However, research utilizing observational data is constrained by the fact that there is no control over actions taken, medications utilized and other such predictors and outcomes of interest. Therefore, with standard data analysis techniques we are provided the ability to describe associations, but we venture out onto thin ice when we try to extrapolate cause and effect from conventional analyses.

The use of a randomized controlled trial has advantages when one is attempting to discern a direct cause and effect relationship between an intervention and an outcome. The purpose of randomization is to ensure balance between study arms of both known and unknown causes of bias and confounding. Subsequently, observed differences in outcomes between randomized groups are assumed to be due to the treatment assignment. When treatment is not randomized, however, the factors that affect the decision to use a particular treatment can also affect the observed outcome. Therefore, ignoring the conditions that led to the use of a specific treatment will run the risk of creating a biased estimate of the effect of the treatment on the patient outcome.

For example, we may want to examine the use of norepinephrine in patients in a retrospective heart failure registry and its relationship to inpatient mortality. For each encounter, the physician deciding to use norepinephrine would make this decision based on heart rate, blood pressure, code status and other clinical factors present on examination. We can reasonably expect that some combination of these same clinical factors would also be associated with patient mortality. During analysis, we can quantify how likely it was that a patient with a certain pattern of characteristics would have received norepinephrine. There are a number of statistical maneuvers available that can address this issue, including Heckman two-step estimators, treatment effect models and instrumental variable models [45]. However, many of these techniques require assumptions regarding the relationship between the predictor variables, the choice of treatment assignment, and the outcome under study that may not hold up in real-world situations.

An alternative technique that accommodates the interplay between predictors, treatment and outcomes is the calculation of propensity scores [46]. A propensity score is, in its simplest conception, the probability of assignment to a treatment based on the observed covariates. Using the above example, the propensity score for norepinephrine use would be the probability that a patient received norepinephrine, given their systolic blood pressure, heart rate, code status and other clinical factors. The propensity score is therefore a summary statistic that replaces the collection of covariates that confound the decision to proceed with one treatment instead of the other, and the score itself is then used as if it were the only covariate affecting the relationship between treatment assignment and outcome [47]. Therefore, the cause and effect relationships can be more accurately described. To continue the example, a regression model of the likelihood of inpatient death in patients that were given norepinephrine versus nitroglycerin, after adjustment of the model for the propensity score, will give a much cleaner estimate of the true relationship between the treatment and the outcome, assuming that the propensity score is properly calculated and specified. It is critical to note that propensity scores can only be generated from those covariates that have been observed, and more importantly, recorded. It is this fact that will maintain the superiority of randomized trials, where feasible, in the ability to delineate causal relationships. Theoretically, randomization will distribute both known and unknown confounding covariates equally between groups when a large enough sample is utilized. By definition, the unknown covariates cannot be controlled for in registry work. Sensitivity analyses can be conducted to determine to what degree the observed treatment effect varies when one manipulates the odds of treatment group assignment. I would refer readers interested in a detailed discussion of the mechanics of sensitivity analysis in the setting of propensity score modeling to Guo and Fraser [45].

The propensity score is generally calculated as a logistic regression function, given that most treatment assignments are binary in nature (received norepinephrine/did not receive norepinephrine). The process of modeling the estimated propensity score is subject to the usual considerations for logistic regression – issues of multicollinearity and overly influential covariate patterns must be resolved with propensity scores as well. The selection of predictor variables should have scientific plausibility as well. Higher order (polynomials, such as squared, cubic or fractional polynomials) variables and interaction terms may be pertinent to providing the best predictive value of the propensity score [45]. Of note, the outcome variable (death, in our example) does not enter into the prediction equation for the propensity score. In addition to these usual steps one takes in order to assure a properly executed logistic regression, however, propensity score calculation requires an additional element – that of 'balancing'.

Balancing refers to the property of the propensity score that, when treated and untreated subjects are matched on propensity scores, the distributions of predictor covariates are now approximately equal between treatment groups. Thus, when two subjects with the same value of a properly specified and well-balanced propensity score are matched, one who received treatment A and one who received treatment B, it is as if the treatment assignment had been randomized, with all the implications of bias reduction (at least, upon those data elements that are actually observed and included within the propensity score calculation) that accompany that assumption [45,48]. If a propensity score cannot create balance between groups on a particular covariate, the database in use may not be able to adequately address the association between the predictor and treatment assignment. For example, if a critical measure such as systolic blood pressure was unable to be balanced in our nitroglycerin versus dobutamine analysis, we would have to strongly consider that we would not be able to answer the question of mortality dependent on treatment based on our dataset. However, if it were a less directly consequential covariate (presence of a CRTpacemaker, for example) we might exclude that covariate from the analysis and chalk it up to material we can use in the discussion section of the ensuing manuscript. Once a propensity score is balanced, we can state that the difference in predictor covariates is due to random error, as opposed to systematic bias, and the estimation of the treatment effect approaches that which would be seen in a randomized trial [45,46,48]. It cannot be overemphasized, however, that this only applies to random error with the observed covariates that were included within the properly specified and balanced propensity score.

Once derived, there are multiple ways that a propensity score can be utilized. The propensity score can be used as covariate in a model, such as linear or logistic regression. In this setting, the propensity score acts as a composite covariate describing the probability of being assigned to the treatment group; the outcome of the model is now interpreted as estimating the effect of the treatment after adjusting for the probability of receiving the treatment [49]. This approach offers a potential advantage in accommodating a large number of predictor variables and higher order terms in generating the propensity score – as the researcher is trying to determine the best possible predictive model for the treatment assignment, there is minimal concern regarding generating an 'over-fit' propensity score [50].

Alternatively, the propensity score can be used to match pairs of treated versus controls based on likelihood of receiving treatment versus not being treated. A substantial advantage of propensity score matching is pure simplicity in the avoidance of data loss. For example, in any reasonably expansive dataset, it is simple to match cases and controls (or treatment groups) on one variable, such as sex. Adding a second variable, such as age, to the matching scheme is usually feasible. But matching each patient receiving nitroglycerin to a patient receiving dobutamine, matching on age, sex, systolic blood pressure, oxygen saturation, code status, preexisting β-blocker use, preexisting ACE-inhibitor use, and location of admission (direct admission vs ED) will result in many subjects going without a match, and thus a resultant loss of observations in the dataset. By reducing all the predictor covariates into a single propensity score, matching becomes substantially simplified. Note that there are different matching strategies available individual versus stratified matching, and a number of techniques for determining what qualifies as the 'best' match - nearest neighbor, caliper matching and so forth. A detailed discussion of the various strategies is beyond the scope of this document, but I would again refer interested readers to the excellent text by Guo and Fraser [45], which provides a very accessible discussion on the gamut of issues dealing with propensity scores.

As stated before, the use of propensity scores is just one of several techniques in use to address the interplay between predictors, treatment assignment and outcome, although it is one of the more versatile. There are of course limitations to the technique – it is not a universal panacea. Just as with a randomized trial when the sample size is small, the baseline characteristics may not be equitably distributed between treatment groups due to chance alone. The analogous situation with a propensity score approach in a small cohort is a score in which critically important covariates simply cannot be balanced – a larger dataset will be less likely to be subject to such a situation [46,47]

Multiple studies in acute heart failure have utilized propensity score adjustment to correct for treatment assignment biases in exploring registry-based data. Abraham *et al.* utilized the ADHERE registry (patients

admitted with the primary diagnosis of acute heart failure) to evaluate the inpatient mortality risk of patients receiving intravenous vasoactive medications (nitroglycerin, nesiritide, dobutamine and milrinone) [20]. As there were multiple treatment options considered, pairwise comparisons (e.g., dobutamine vs milrinone, nitroglycerin vs dobutamine) were established using separate propensity scores for each comparison. The propensity scores were generated using a form of regression known as classification and regression tree analysis (CART), and a logistic regression model utilizing the propensity score as an independent covariate, along with gender and known mortality risk factors. Approximately 18,000 patients received vasoactive medications in the analysis (out of over 65,000 patients in the registry at the time of the study). Across all pairwise comparisons, propensity score-adjusted analyses trended towards the null, demonstrating that observed differences in treatment effects were not as extreme as would be suggested by unadjusted analyses.

Mebazaa *et al.* performed a similar analysis out of the ALARM-HF registry, but with a primary focus on comparing diuretic therapy versus diuretic plus vasodilator therapy in terms of inpatient mortality [51]. As opposed to the Abraham study, this study utilized matching on propensity scores to perform a case–matched control analysis. As a result of their analysis, this group noted an improvement in inpatient mortality (adjusted OR: 0.71; 95% CI: 0.51–0.98) with the addition of vasodilator therapy to a diuretic regimen. The use of propensity scoring again refined the treatment effect towards the null – the unadjusted OR for mortality with the use of vasodilators was 0.55 (95% CI: 0.47–0.71), indicating a substantial selection bias on the part of the treating physicians when initiating vasodilators.

#### Missing data

Another potential source of bias in analyzing registry data are missing data. While this is less of a problem with prospectively collected registries explicitly intended for research, even the best designed data plans involve human beings, and therefore the probability of human error, somewhere in the chain. The likelihood of missing data increases when data are gleaned from medical records or administrative data, where the data elements that the researcher is interested in evaluating may not have been recorded by the clinician or retained by the data recorder. All too frequently, when confronted with missing data, the researcher chooses to use complete case analysis, where only those observations with no missing data are included in the analysis. This results in the loss of data, as potentially informative subjects are excluded. As well, if there is an association between the missing data elements and confounders, treatment, or

outcomes, systematic bias could be introduced as individuals from discrete subpopulations are at increased risk for being excluded. One example of such systematic bias due to missing data is the concept of workup bias-patients who are less ill, and more likely to have normal test results, are less likely to undergo diagnostic tests. In the acute heart failure arena, patients hospitalized with less severe exacerbations could be less likely to undergo echocardiography during their inpatient stay. Let us assume, for the sake of this example, that sicker patients are likely to have worse ejection fractions. An analysis of ejection fraction performed without taking the missing measurements in the healthier patients into consideration would therefore likely overestimate the degree of systolic dysfunction present in the cohort.

When dealing with the issue of missing data, it is important to determine if there is any pattern as to how the data are missing. Data will fall into one of three patterns-missing completely at random, missing at random and missing not at random. This is an critical issue, as different patterns of missing data warrant different analytical responses. In data that are missing completely at random, there is no relationship between the absence of the element, the value of the missing element, and any of the other variables observed on that subject. This is generally an unusual occurrence. When data are missing at random, there is a relationship between the other variables in the dataset and the probability of the element being missing. When data are missing not at random, the actual value of the missing data element determines the probability of the element being missing. For example, on an employment application, one might see the question, 'Do you use drugs?' One could reasonably opine that there would be a relationship between the true answer being 'Yes', and a missing response.

When data are missing completely at random (again, an unusual situation but one frequently assumed), one could proceed with complete case analysis and not suffer from biased estimators [52,53]; however, researchers will still be left with a reduced sample size, with all the ensuing consequences. For the other missing data patterns, however, pursuing complete case analysis will likely produced biased results, and therefore methods for finding replacements for the missing observations such as imputing, have been developed. There are a variety of options available for specific situations, such as randomly sampling existing observations, replacing all missing values with the mean value of the data element in the data base, weighted estimations, maximum likelihood estimators and single imputation methods [52,54,55]. Each of these has advantages – usually simplicity in execution, but sacrifice in the ability to produce a richer estimate of the true distribution of the element of interest. However, one of the more flexible and simpler

methods is the use of multiple imputation as originally described by Rubin [56]. Multiple imputation starts with the concept that, if we know the values of the covariates surrounding a missing data element, and we can determine the relationship of those known covariates to the values of that element that are present for other subjects, we can estimate what the likely value of the missing element is. Multiple imputation repeats this process several times, generating multiple versions of the database that will have different values, randomly drawn from the probable values for the missing observations. Each imputation is then analyzed as if it were the complete dataset, and the results combined [55,57]. Multiple imputation techniques can be used in a large variety of analyses, including survey, clustered and survival data, with different variable types (dichotomous, continuous and multinomial). Multiple imputation can be implemented very simply when data are missing at random or completely at random; when data are missing not at random, an additional step in modeling predictors of the observation being missing enhances the validity of the subsequent multiple imputation [49].

It is this incorporation of a probability distribution and uncertainty that provides increased face validity to the concept of multiple imputation, as opposed to single imputation or direct replacement methods (such as when the missing observation is replaced with the mean of all the same observations that were present). We do not actually know, with 100% certainty, what the value of the missing observation is. We can establish what the values are likely to be, but there is an element of uncertainty that simply cannot be removed by any imputation process - this uncertainty is incorporated by the use of multiple iterations of the dataset and the eventual synthesis of the estimations into one set of results. In other words, we are likely to be at least a little bit wrong each time we sample an imputed value, but we also expect to be close to, and distributed around, the true value that is missing. Single imputation techniques are simply going to have that first likelihood of error, and not be able to provide a correction via sequential draws from the probability distribution that encompasses the missing value being estimated.

The techniques of multiple imputation and propensity score adjustment can definitely be combined in sequence, and in fact a propensity score will lose power and validity when derived from a dataset with substantial amounts of missing data. After the missing data are imputed, propensity scores can then be calculated and the covariate-treatment-outcome relationship analyzed [58]. Of note, as opposed to the approach for generating a propensity score, the outcome of interest should generally be included among the covariates used to estimate the distribution of the missing data elements. There is generally no good reason not to assume, and frequently strong face validity in assuming, that the outcome of the subject had a bearing on what data was, and was not, successfully captured. By creating reasonable estimates of missing data, and using the now full dataset to understand the relationships between preexisting characteristics and treatments utilized, registries can be successfully used to draw clearer conclusions regarding the true effects of strategies we currently employ to treat patients with acute heart failure.

## Conclusion & future perspective

Despite solid groundwork, rationale hypotheses and firm preliminary studies, the recent heart failure literature has not borne witness to substantial successes when it comes to clinical trials. The primary weakness of the current approach appears to be a mismatch in intervention and study population – interventions for unstable patients are being studied in patients who have been stabilized. As seen the ASCEND–HF trial, earlier enrollment and initiation of treatment is associated with improved outcomes in trial subjects. In order for future heart failure trials to succeed, researchers from both cardiology and emergency medicine will need to collaborate to be able to derive scientific knowledge from the truly unstable patient – the patient who currently receives the same treatment now (diuresis, nitrates and inotropes) as in the 1970s. As heart failure continues to increase in prevalence, continuing to design research as we have been doing will fail to meet the needs of our patients.

As clinical data are increasingly electronic and linked to billing and administrative data, we have access to greater and greater amounts of information from which to derive associations and generate hypotheses. Large research databases have the potential, with appropriate statistical correction and technique, to provide a clearer estimate of treatment effects, once covariates leading to treatment group assignment are adjusted for. Appropriate database analysis will allow us to best determine how best to apply the interventions and techniques we have available to us that are already in widespread use - tools that may not be currently leveraged in their most efficacious manner. Appropriate use of data registries will become more prevalent as the amount of electronic patient data explodes. With luck, the mining will also become 'cleaner' as computer power increases exponentially, rendering the computational burden required for propensity score generation and multiple imputation moot. We may even see some

## **Executive summary**

- Heart failure continues to increase in incidence and prevalence, with minimal improvement in mortality. Acute heart failure is a time of substantial clinical risk for the patient, with high rates of short-term morbidity and mortality. Several recent large clinical trials have simply failed to demonstrate clinical improvements in acute heart failure patients with various agents that performed well in preliminary trials. Issues to consider:
  - We use the term 'heart failure' as if it was a single clinical entity; in reality, this is at best a collective description of variety of pathologies that share a common root.
  - Initial stabilization of the acute heart failure patient is frequently successful, and frequently begins in the emergency department. Selecting patients that are refractory to stabilization will produce a systematic bias in the results of the study.
  - Likewise, studying a drug or an intervention that is designed to treat a decompensated patient after they have already been treated and stabilized will also bias the experiment towards the null hypothesis.
  - Successful studies for decompensated heart failure patients should identify, recruit and initiate treatment in patients while they are still decompensated, that is, successful strategies will need to incorporate the emergency department, the source of admission for 80% of acute heart failure patients in the USA.
- Large amounts of patient information are available to us, both in databases intended for research as well as within electronic clinical and administrative databases. However, traditional retrospective cohort studies have several weaknesses that have limited their role to describing epidemiology, or at best generating hypotheses. However, statistical techniques can substantially enhance the validity and information density of such data.
  - In retrospective studies, treatment assignment was not random patients received one agent or another based on a variety of factors, such as clinical, personal preference, physician preference and so on.
  - The same factors that led to treatment assignment frequently confound the relationship between treatment received and patient outcome.
  - Propensity scores can be generated that predict the probability that the patient received one treatment versus another (or vs no treatment). Propensity scores can then be utilized in a variety of ways (matching, stratification and regression weighting) to minimize the confounding of the covariates and maximize the signal-to-noise ratio in relating the treatment received to the outcome of interest.
  - Missing data are an omnipresent issue. Complete case analysis frequently biases the results of the analysis.
  - Multiple imputation is a technique (one of several, but flexible and accessible via most software packages) that can 'fill in' the missing data based on the relationships between those data elements that are not missing in other patient observations.

questions answered so well from dataset analysis that a randomized clinical trial would not be ethically appropriate to conduct, with patients knowingly exposed to a potentially inferior treatment. These two fields – missing data and causality analysis – are rapidly evolving areas of statistical research as well, and new advances are sure to emerge to increase the clarity of the signals we are getting from our data and our patients.

The role of prospective, randomized trials will not vanish, however. Development of novel approaches, drugs, and interventions are still best evaluated in this fashion – indeed, database work can hardly be expected to evaluate new interventions. Both properly analyzed database studies, to refine the tools and approaches already in use, and well-designed clinical trials focused on evaluating novel interventions in the appropriate populations and conducted in the setting of acute heart failure, will be required to advance the care of our patients.

## Financial & competing interests disclosure

BC Hiestand has received research funding from Medtronic Inc., Biosite Inc., Inovise Medical Inc., Heartscape International, Nanosphere Inc., Mitsubishi Medience, Sanofi-Aventis, The Medicines Company and a medical writing consulting relationship with new-Mentor Inc. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

## Bibliography

Papers of special note have been highlighted as: •• of considerable interest

- Lloyd-Jones D, Adams RJ, Brown TM *et al.* Heart disease and stroke statistics – 2010 update: a report from the American Heart Association. *Circulation* 121(7), e46–e215 (2010).
- 2 Stewart S, Ekman I, Ekman T et al. Population impact of heart failure and the most common forms of cancer: a study of 1,162,309 hospital cases in Sweden (1988 to 2004). Circ. Cardiovasc. Qual. Outcomes. 3(6), 573–580 (2010).
- 3 Gheorghiade M, Pang PS. Acute heart failure syndromes. J. Am. Coll. Cardiol. 53(7), 557–573 (2009).
- 4 Gheorghiade M, Abraham WT, Albert NM et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. JAMA 296(18), 2217–2226 (2006).
- 5 Adams KF Jr, Fonarow GC, Emerman CL et al. Characteristics and outcomes of patients hospitalized for heart failure in the United States: rationale, design, and preliminary observations from the first 100,000 cases in the Acute Decompensated Heart Failure National Registry (ADHERE). Am. Heart J. 149(2), 209–216 (2005).
- 6 Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N. Engl. J. Med. 360(14), 1418–1428 (2009).
- 7 Poole-Wilson PA, Swedberg K, Cleland JG et al. Comparison of carvedilol and metoprolol on clinical outcomes in patients with chronic heart failure in the Carvedilol

Or Metoprolol European Trial (COMET): randomised controlled trial. *Lancet* 362(9377), 7–13 (2003).

- 8 Pfeffer MA, Swedberg K, Granger CB *et al.* Effects of candesartan on mortality and morbidity in patients with chronic heart failure: the CHARM-Overall programme. *Lancet* 362(9386), 759–766 (2003).
- 9 Pitt B, Remme W, Zannad F et al. Eplerenone, a selective aldosterone blocker, in patients with left ventricular dysfunction after myocardial infarction. N. Engl. J. Med. 348(14), 1309–1321 (2003).
- 10 Moss AJ, Zareba W, Hall WJ *et al.* Prophylactic implantation of a defibrillator in patients with myocardial infarction and reduced ejection fraction. *N. Engl. J. Med.* 346(12), 877–883 (2002).
- 11 Moss AJ, Hall WJ, Cannom DS *et al.* Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. Multicenter automatic defibrillator implantation trial investigators. *N. Engl. J. Med.* 335(26), 1933–1940 (1996).
- 12 Moss AJ, Hall WJ, Cannom DS *et al.* Cardiac-resynchronization therapy for the prevention of heart-failure events. *N. Engl. J. Med.* 361(14), 1329–1338 (2009).
- Cleland JG, Daubert JC, Erdmann E *et al.* The effect of cardiac resynchronization on morbidity and mortality in heart failure. *N. Engl. J. Med.* 352(15), 1539–1549 (2005).
- 14 Nohria A, Tsang SW, Fang JC *et al.* Clinical assessment identifies hemodynamic profiles that predict outcomes in patients admitted with heart failure. *J. Am. Coll. Cardiol.* 41(10), 1797–1804 (2003).

- 15 Randazzo MR, Snoey ER, Levitt MA *et al.* Accuracy of emergency physician assessment of left ventricular ejection fraction and central venous pressure using echocardiography. *Acad. Emerg. Med.* 10(9), 973–977 (2003).
- 16 Nguyen VT, Ho JE, Ho CY *et al.* Handheld echocardiography offers rapid assessment of clinical volume status. *Am. Heart J.* 156(3), 537–542 (2008).
- 17 Dickstein K, Cohen-Solal A, Filippatos G et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2008: the task force for the diagnosis and treatment of acute and chronic heart failure 2008 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association of the ESC (HFA) and endorsed by the European Society of Intensive Care Medicine (ESICM). Eur. Heart J. 29(19), 2388–2442 (2008).
- 18 Weintraub NL, Collins SP, Pang PS et al. Acute heart failure syndromes: Emergency Department presentation, treatment, and disposition: current approaches and future aims: a scientific statement from the American Heart Association. *Circulation* 122(19), 1975–1996 (2010).
- Excellent summary of the challenges faced in acute heart failure research, with a call for a multicenter research network of collaborating cardiologists and emergency physicians.
- 19 Fonarow GC, Adams KF, Jr, Abraham WT et al. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA 293(5), 572–580 (2005).

# Clinical Trial Perspective Hiestand

- 20 Abraham WT, Adams KF, Fonarow GC et al. In-hospital mortality in patients with acute decompensated heart failure requiring intravenous vasoactive medications: an analysis from the Acute Decompensated Heart Failure National Registry (ADHERE). J. Am. Coll. Cardiol. 46(1), 57–64 (2005).
- 21 Collins SP, Levy PD, Lindsell CJ *et al.* The rationale for an acute heart failure syndromes clinical trials network. *J. Card. Fail.* 15(6), 467–474 (2009).
- 22 Gheorghiade M, Zannad F, Sopko G *et al.* Acute heart failure syndromes: current state and framework for future research. *Circulation* 112(25), 3958–3968 (2005).
- 23 Mebazaa A, Pang PS, Tavares M *et al.* The impact of early standard therapy on dyspnoea in patients with acute heart failure: the URGENT-dyspnoea study. *Eur. Heart J.* 31(7), 832–841 (2010).
- 24 Fonarow GC, Corday E. Overview of acutely decompensated congestive heart failure (ADHF): a report from the ADHERE registry. *Heart Fail. Rev.* 9(3), 179–185 (2004).
- 25 Peacock WF, Emerman C, Costanzo MR et al. Early vasoactive drugs improve heart failure outcomes. Congest. Heart Fail. 15(6), 256–264 (2009).
- 26 Gheorghiade M, Konstam MA, Burnett JC Jr et al. Short-term clinical effects of tolvaptan, an oral vasopressin antagonist, in patients hospitalized for heart failure: the EVEREST Clinical Status Trials. JAMA 297(12), 1332–1343 (2007).
- 27 Konstam MA, Gheorghiade M, Burnett JC et al. Effects of oral tolvaptan in patients hospitalized for worsening heart failure: the EVEREST outcome trial. JAMA 297(12), 1319–1331 (2007).
- 28 McMurray JJ, Teerlink JR, Cotter G et al. Effects of tezosentan on symptoms and clinical outcomes in patients with acute heart failure: the VERITAS randomized controlled trials. JAMA 298(17), 2009–2019 (2007).
- 29 Massie BM, O'Connor CM, Metra M et al. Rolofylline, an adenosine A1-receptor antagonist, in acute heart failure. N. Engl. J. Med. 363(15), 1419–1428 (2010).
- 30 Felker GM, Lee KL, Bull DA *et al.* Diuretic strategies in patients with acute decompensated heart failure. *N. Engl. J. Med.* 364(9), 797–805 (2011).
- 31 O'Connor CM, Starling RC, Hernandez AF et al. Effect of nesiritide in patients with acute decompensated heart failure. N. Engl. J. Med. 365(1), 32–43 (2011).

- 32 Peacock WF, Braunwald E, Abraham W et al. National heart, lung, and blood institute working group on Emergency Department management of acute heart failure: research challenges and opportunities. J. Am. Coll. Cardiol. 56(5), 343–351 (2010).
- Summary of the proceedings of an NHLBI-sponsored working group, spurred by the recent failures of clinical acute heart failure trials. A central theme was the need to incorporate the emergency department as a research setting for acute heart failure patients.
- 33 Kociol RD, Hammill BG, Fonarow GC et al. Generalizability and longitudinal outcomes of a national heart failure clinical registry: comparison of Acute Decompensated Heart Failure National Registry (ADHERE) and non-ADHERE Medicare beneficiaries. Am. Heart J. 160(5), 885–892 (2010).
- 34 Mehra MR, Yancy CW, Albert NM et al. Evidence of clinical practice heterogeneity in the use of implantable cardioverterdefibrillators in heart failure and postmyocardial infarction left ventricular dysfunction: findings from IMPROVE HF. *Heart Rhythm* 6(12), 1727–1734 (2009).
- 35 Diercks DB, Fonarow GC, Kirk JD et al. Illicit stimulant use in a United States heart failure population presenting to the emergency department (from the Acute Decompensated Heart Failure National Registry Emergency Module). Am. J. Cardiol. 102(9), 1216–1219 (2008).
- 36 Peacock WF, Costanzo MR, De Marco T et al. Impact of intravenous loop diuretics on outcomes of patients hospitalized with acute decompensated heart failure: insights from the ADHERE registry. Cardiology 113(1), 12–19 (2009).
- 37 Peacock WF, De Marco T, Fonarow GC et al. Cardiac troponin and outcome in acute heart failure. N. Engl. J. Med. 358(20), 2117–2126 (2008).
- 38 Levy PD, Nandyal D, Welch RD et al. Does aspirin use adversely influence intermediateterm postdischarge outcomes for hospitalized patients who are treated with angiotensin-converting enzyme inhibitors or angiotensin receptor blockers? Findings from Organized Program to Facilitate Life-Saving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). Am. Heart J. 159(2), 222–230 (2010).
- 39 Fonarow GC, Abraham WT, Albert NM et al. Age- and gender-related differences in quality of care and outcomes of patients

hospitalized with heart failure (from OPTIMIZE-HF). *Am. J. Cardiol.* 104(1), 107–115 (2009).

- 40 Hernandez AF, Hammill BG, O'Connor CM et al. Clinical effectiveness of β-blockers in heart failure: findings from the OPTIMIZE-HF (Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure) Registry. J. Am. Coll. Cardiol. 53(2), 184–192 (2009).
- 41 O'Connor CM, Abraham WT, Albert NM et al. Predictors of mortality after discharge in patients hospitalized with heart failure: an analysis from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). Am. Heart J. 156(4), 662–673 (2008).
- 42 Curtis LH, Greiner MA, Hammill BG et al. Representativeness of a national heart failure quality-of-care registry: comparison of OPTIMIZE-HF and non-OPTIMIZE-HF Medicare patients. Circ. Cardiovasc. Qual. Outcomes. 2(4), 377–384 (2009).
- 43 Heidenreich PA, Fonarow GC. Are registry hospitals different? A comparison of patients admitted to hospitals of a commercial heart failure registry with those from national and community cohorts. *Am. Heart J.* 152(5), 935–939 (2006).
- 44 Bufalino VJ, Masoudi FA, Stranne SK *et al.* The American Heart Association's recommendations for expanding the applications of existing and future clinical registries: a policy statement from the American Heart Association. *Circulation* 123(19), 2167–2179 (2011).
- 45 Guo S, Fraser MW. Propensity Score Analysis: Statistical Methods and Applications. Guo S, Fraser MW (Eds). Sage Publications, CA, USA (2010).
- Excellent textbook regarding the use of propensity scores – not as statistically dense as some works, but may be a bit beyond a statistical novice.
- 46 Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55 (1983).
- Seminal article by Rubin defining the role and value of propensity score analysis.
- 47 Rubin DB. Estimating causal effects from large datasets using propensity scores. Ann. Intern. Med. 127(8 Pt 2), 757–763 (1997).
- 48 Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat. Med.* 26(1), 20–36 (2007).

# Research in acute decompensated heart failure Clinical Trial Perspective

- 49 Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu. Rev. Public Health* 21, 121–145 (2000).
- 50 D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a nonrandomized control group. *Stat. Med.* 17(19), 2265–2281 (1998).
- 51 Mebazaa A, Parissis J, Porcher R et al. Short-term survival by treatment among patients hospitalized with acute heart failure: the global ALARM-HF registry using propensity scoring methods. Intensive Care Med. 37(2), 290–301 (2011).
- 52 Schafer JL. Multiple imputation: a primer. *Stat. Methods Med. Res.* 8(1), 3–15 (1999).

- •• A user friendly introduction to the underpinnings of multiple imputation.
- 53 White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat. Med.* 29(28), 2920–2931 (2010).
- 54 Ambler G, Omar RZ, Royston P. A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. *Stat. Methods Med. Res.* 16(3), 277–298 (2007).
- 55 Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev. Sci.* 8(3), 206–213 (2007).

- 56 Rubin DB. Multiple Imputation for Nonresponse in Surveys. Rubin DB (Ed.). Wiley, NY, USA (1987).
- 57 Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat. Methods Med. Res.* 8(1), 17–36 (1999).
- 58 Little RJA, Rubin DB. Statistical Analysis with Missing Data. Little RJA, Rubin DB (Eds). Wiley, NY, USA (2002).