Title: Can We Value Research Investments? A Systematic Review of Heart Failure Clinical Trial Endpoints 1996-2015

Short running title: A Review of Heart Failure Trial Endpoints

Authors: Meng Li, ScM[1]; Jordan Banks, MPA[2]; Kelley Branch, MD, MS[3]; David Veenstra, PharmD, PhD[1]; Joshua Carlson, PharmD, PhD[1]; Jeffrey Probstfield, MD[3]; Anirban Basu, PhD[1]

Affiliations: [1]The CHOICE Institute, School of Pharmacy, University of Washington; [2]Department of Health Services, School of Public Health, University of Washington; [3]Division of Cardiology, School of Medicine, University of Washington

Total word count: 2846

Financial support: This research is supported by grant R01HL126804. This research did not receive financial support from industry.

Address for correspondence:
Anirban Basu, PhD
1959 NE Pacific St, Box 357630
Magnuson Health Sciences Center, Room H-375Q
Seattle, WA, 98195-7630
Tel: 206.616.2986
Fax: 206.543.3964

ABSTRACT

Objectives The purpose of this study is to systematically review the endpoints used in heart failure clinical trials, and examine if evidence exists to extrapolate surrogate (non-mortality-related) endpoints to more comprehensive endpoints such as survival or quality-adjusted survival.

Background Quantitative estimates of the value of information for future clinical trials can be compared to inform research prioritization, as long as they are expressed in similar units such as comprehensive outcomes.

Methods We reviewed all primary endpoints of phase II-IV heart failure trials registered on ClinicalTrials.gov between 1996 and 2015 and their trends over time. We conducted a PubMed search for statistical associations between these endpoints and long-term survival.

Results 688 eligible trials reported a total of 992 endpoints. Mortality was used in 14.7% of the trials either as an individual end point (5.1%) or as part of a composite endpoint (9.6%). Frequently used surrogate endpoints include exercise capacity (10.8%), left ventricular ejection fraction (6.8%), and brain natriuretic peptides (6.1%). Use of surrogate outcomes appears to be trending up, especially in Phase IV trials. There exists at least one estimate of the statistical association between long-term survival and each of the secondary outcomes. The most (23 estimates) exist for left ventricular ejection fraction.

Conclusions The use of surrogate endpoints in heart failure trials was extensive and rising. The quantity of evidence linking endpoints to long-term survival varied significantly. Future work should review the quality of these links so that so as to incorporate uncertainty in projecting comprehensive outcomes in value calculations.

Key Words: Value of information, clinical trial endpoints, surrogate endpoints, comprehensive endpoints, statistical associations

Abbreviations

VOI = Value of Information

AACT = Aggregate Analysis of ClinicalTrials.gov

MI = Myocardial Infarction

NYHA = New York Heart Association

LVEF = Left Ventricular Ejection Fraction

BMI = Body Mass Index

INTRODUCTION

Limited biomedical research funding makes the optimal allocation of research funding a difficult but important task. Value of Information (VOI) analysis is a coherent, quantitative approach to research prioritization that has received increasing attention in recent years, particularly within the context of comparative effectiveness research.(1-3) This approach involves estimation of the economic value of the information generated through new research using methods from decision analysis and economic theory. It relies on the fact that new information can help make better clinical decisions and therefore produce value based on patient and societal welfare, as compared to clinical decisions being made in the absence of such information.(4) It can be viewed as an extension of meta-analysis, which produces estimates for the current level of uncertainty about the relative effectiveness of interventions. Only if it is determined that this level of uncertainty is unacceptable would investment in research costs to resolve this uncertainty is worthwhile.

Conduct of VOI involves estimating the likelihood of a level of the "true" relative effectiveness in uncertainty is resolved, where that level informs appropriate clinical practice and, therefore, patient outcomes. Since the goal is prioritization, one must be able to compare these VOI estimates for different clinical studies in order to compare the return on investment for these clinical studies. Such comparisons can help funders like the National Institutes of Health to visualize and optimize their portfolio of research investments.(5) However, a fundamental challenge in carrying out such comparison across VOI analysis is to express value in one common metric across all research studies. Metrics that would naturally serve this purpose are comprehensive patient outcomes measures that would translate to life expectancy or quality-adjusted life expectancy. Clinical trials that directly characterize uncertainty around the effects of interventions on these comprehensive endpoints make it possible to directly perform VOI analysis on these metrics and lend themselves comparable to others. For studies that employ a surrogate endpoint, estimation of comparable VOI would involve developing a simulation model to project the effects on these surrogate endpoints to effects on comprehensive outcomes.

The use of such surrogate outcomes in clinical research is ubiquitous and for good reasons. Effects on comprehensive endpoints may take many years and a very large sample size to emerge, especially in the case of patients with chronic conditions where there is not a significant acute mortality. In those cases, with constrained resources, many clinical trials use intermediate or surrogate endpoints as their primary endpoint, with the goal of reaching an answer with a reasonable amount of time and budget. However, to make sure that we are investing in research that likely to produce commensurate value in return, one must be able to link these intermediate outcomes to long-term comprehensive measures of health. Tenuous links between surrogate endpoints and comprehensive outcomes should attenuate the value of research investments.

This consideration in clinical trial design is particularly relevant in cardiovascular diseases where intermediate endpoints are more extensively used than in many other therapeutic areas. Downing et al. reviewed the design of pivotal efficacy trials providing the basis for approval of novel therapeutic agents by the US Food and Drug Administration between 2005 and 2012 and found that nearly 85% of pivotal trials in cardiovascular disease, diabetes mellitus, and hyperlipidemia used an intermediate endpoint as their primary endpoint, well above the average 49% across all therapeutic areas.(6) In fact, there is a large body of clinical trials outside of pivotal trials where intermediate outcomes are also used. In such cases, to assess the VOI of future clinical trials, choosing a validated intermediate endpoint that has established links to comprehensive endpoints in the trial is critical.(7)

In this paper, as part of a larger study to establishing the expected value of clinical trials funded by the National Heart, Lung and Blood Institute, we examined historical trends in the use of comprehensive and surrogate endpoints in heart failure trials and the degree of evidence that links common intermediate outcomes and life expectancy. The objective of this study was, therefore, two-fold: (1) to systematically review endpoints used in phases II-IV interventional clinical trials of drugs and devices in heart failure

over the last twenty years, and (2) to examine if the frequently used surrogate endpoints in heart failure clinical trials were linked to long-term survival in recent literature.

METHODS

To systematically examine the use of endpoints in heart failure clinical trials over time, we searched the Aggregate Analysis of ClinicalTrials.gov (AACT) database from 1996 to September 2015 for all completed, terminated, and ongoing Phase II to IV clinical trials in heart failure. The AACT database aggregates and restructures publicly available ClinicalTrials.gov data to facilitate analysis of clinical trials registered on the website. While not all trials on human subjects are registered with ClinicalTrials.gov, it is inclusive of all studies required by the Food and Drug Administration Amendments Act (FDAAA) 801 Requirements to register. Since drugs and devices form the majority of treatment for heart failure patients, we focused on these interventional trials. No studies were excluded based on their status, location, treatment, randomization, or other design characteristics. We extracted all primary endpoints from eligible clinical trials as well as the start year of each trial. If a clinical trial used a composite endpoint, we further broke that composite endpoint into individual endpoints.

The comprehensive set of endpoints that we collected through our systematic review were categorized into three levels: *Level 1*, direct measures of survival or quality of life; *Level 2*, clinical endpoints; and *Level 3*, surrogate endpoints. *Level 1* endpoints in this study included mortality either as an individual endpoint or as part of a composite endpoint and quality of life. *Level 2* included endpoints that are clinically meaningful and would normally involve a diagnosis or assessment by a healthcare provider, such as myocardial infarction (MI), stroke, or arrhythmia. Interactions with the health care system such as emergency department visit and hospital admission were also categorized as *Level 2*. *Level 3* endpoints include biological markers for the condition such as exercise capacity, maximum oxygen consumption, New York Heart Association (NYHA) status, left ventricular ejection fraction (LVEF), brain natriuretic peptide, body mass index (BMI), pulmonary capillary wedge pressure, mitral regurgitation, blood

pressure, left ventricular end systolic volume, urine output, serum creatinine, or heart rate, etc. A composite endpoint was considered as *Level 1* if it consisted of at least one individual endpoint from *Level 1*.

We summarized the frequencies of primary endpoints used in the identified heart failure clinical trials and reported endpoints that were included in at least 1% of all trials. We explored the use of endpoints and their trends over time by phases and by levels.

The second part of the study involved a review of long-term observational studies on heart failure patients, defined as having a follow-up period of at least five years, in PubMed. The purpose of the review was to examine if statistical associations existed between long-term survival (greater than or equal to five years) and frequently used clinical or surrogate (*Level 2* and *Level 3*) endpoints in phase II-IV heart failure clinical trials. We reviewed endpoints that were used in at least 5% of phase II-IV heart failure clinical trials. We limited the PubMed search to studies published in the past 20 years and written in English. Other trial eligibility criteria included: (1) being conducted in heart failure patients, (2) having examined statistical associations between heart failure clinical or surrogate endpoints and survival; and (3) having a mean follow-up of at least five years. We reviewed eligible studies and reported the number of statistical associations between each clinical or surrogate endpoint and long-term survival.

RESULTS

Evolution of Heart Failure Clinical Trial Endpoints

From 1996 to 2015, there were 688 phase II-IV heart failure clinical trials registered on ClinicalTrials.gov, reporting a total of 992 primary endpoints (Table 1). Among the 688 trials, 214 (31%) were phase II, 62 (9%) were phase II/III, 204 (30%) were phase III, and 208 (30%) were phase IV. The number of primary endpoints per trial ranged from 1 to 18, though in most trials only one primary endpoint was reported. A composite endpoint that includes mortality is the most frequently used primary

endpoint among phase II-IV heart failure trials [101/688 (14.7%) trials]. However, only 35 (5.1%) studies included mortality as an individual primary endpoint. Thirty (4.4%) studies reported quality of life as one of their primary endpoints. The most frequently used *Level 2* endpoints were hospital admission [47 (6.8%)], MI [28 (4.1%)] and dyspnea [25 (6%)]. Frequently used *Level 3* endpoints included exercise capacity [74 (10.8%)], LVEF [47 (6.8%)], brain natriuretic peptides [42 (6.1%)] and maximal oxygen consumption [39 (5.7%)]. (Table 1)

Eighty out of 101 trials that reported using a composite endpoint including mortality were Phase III or Phase IV studies (Table 2). Sixty-eight (67.3%) out of 101 had a composite primary endpoint that included hospital admission alongside mortality. The type of hospital admission varied significantly across clinical trials, including both cardiovascular and non-cardiovascular hospital admission, first hospital admission and readmission. Other individual endpoints frequently used in a composite endpoint including mortality in heart failure clinical trials include stroke (8 studies, 8% of all), MI (6 studies, 6%), arrhythmia (4 studies, 4%), quality of life (3 studies, 3%), heart transplantation (3 studies, 3%), and exercise capacity (2 studies, 2%). When stratified by phases, hospital admission and quality of life appeared to be more frequently used in later phase trials whereas arrhythmia, heart transplantation, and exercise capacity tend to be more frequently used in earlier phases.

Across all phases of trials, 22.5% of trials included at least one *Level 1* endpoint as one of their primary endpoints (Table 3). The percentage increased from 12.1% in Phase II, to 22.6% in Phase II/III and 36.8% in Phase III. Surprisingly, only 19.2% of Phase IV trials included at least one *Level 1* endpoint. Across phases, 44% of trials included at least one *Level 2* endpoint as one of their primary endpoints. The percentage increased from 39.3% in Phase II, to 46.8% in Phase II/III and 55.9% in Phase III. 36.5% of Phase IV trials included at least one *Level 2* endpoint. 57.7% of trials across phases included at least one *Level 3* endpoint as one of their primary endpoints. The percentage decreased from 66.4% in Phase II, to

64.5% in Phase II/III and 42.6% in Phase III. 61.5% of Phase IV trials included at least one *Level 3* endpoint.

Figures 1-4 summarized temporal trends of endpoints in heart failure trials registered on ClinicalTrials.gov from 2002 to 2015. Data from before 2002 were not included in the analysis of trend since there were too few trials registered on ClinicalTrials.gov before 2002 for any reliable estimate. Overall, there was a steady increase in the number of Phase II-IV heart failure trials from 2002 to 2014. The biggest increase in the number of trials was seen in Phase II while the number of trials in other phases remained relatively stable over time. *Level 2* and *Level 3* endpoints were used at higher rates than *Level 1* endpoints. From 2002 to 2015, there appeared to be a slightly increasing trend in using *Level 3* endpoints in Phase II-IV heart failure trials and a slightly decreasing trend in using *Level 2* endpoints. The use of *Level 1* endpoints in this time-period remained relatively stable. When broken down by phases, the numbers in each phase were too small for any trend to be visible.

Links to Long-term Survival

Table 3 summarized statistical associations between the *Level 2* and *Level 3* endpoints and survival reported in long-term observational studies in PubMed with a mean follow-up of at least five years. We found 23 estimates of the statistical association between LVEF and long-term survival, six for MI, five for stroke, four for brain natriuretic peptide, three for maximal oxygen consumption, and one each for exercise capacity, and hospital admission. In the studies that we reviewed, we also found links of blood pressure (21 links), NYHA status (15), serum creatinine (11), heart rate (6), smoking (6), body weight or BMI (5), hemoglobin (5), cholesterol (2), angina (1), cardiac index (1), and coenzyme Q10 (1) to long-term survival.

DISCUSSION

Prioritizing limited research funding by investing in studies that have the largest returns is fundamental to increasing efficiency within any research enterprise. Value of information methods is a set of tools that can be used to assess the expected value of research studies. However, valuation within these or any other methods must anchor these comparisons along with some common comprehensive metric of outcomes that are valuable to patients. Long-term survival and quality-adjusted survival are natural candidates for this purpose because of their unbiased and final nature as well as their connection to societal willingness-to-pay. Understanding the links between intermediate endpoints used in clinical trials and long-term comprehensive outcomes begins to refine the value of improving on these endpoints. If a treatment intervention is found to have an impact on an intermediate endpoint in a clinical trial and that intermediate endpoint is linked to long-term survival, that intermediate impact can be extrapolated to impact on survival and thus be valued.

In this review, there were a wide variety of endpoints used in phase II-IV heart failure clinical trials, measuring not only survival but also different aspects of how patients feel and function. The majority of endpoints used in phase II-IV heart failure clinical trials are surrogates with increasing trends in the use of these endpoints mainly in phase II and IV heart failure trials. This increasing ternd in the use of non-mortality endpoints highlights the importance of understanding the links between these surrogate endpoints and level 1 outcomes for the value of information calculations. The trends in different levels of endpoints used in Phase III trials were found to be stable, despite large temporal fluctuations. For Phase IV trials that are expected to demonstrate the effectiveness of interventions instead of efficacy, to our surprise, Level 3 end points were most commonly used and there seem to be an increasing trend in the use of surrogate outcomes after 2011. In our review of recent literature, all of the frequently used surrogate endpoints in heart failure trials had established links to long-term survival in observational studies, though the strength of evidence varied across endpoints. This suggested that it is possible to value an effect or a reduction of uncertainty on a surrogate endpoint through its link to a comprehensive endpoint such as

survival, although the value could be substantially mitigated by the level of uncertainty in the mapping parameters that link the surrogate outcome to the comprehensive endpoints.

Certain limitations of this study should be discussed. We conducted a systematic search of the recent literature on statistical associations between frequently used surrogate endpoints in heart failure trials and long-term mortality, although the search was not exhaustive. We used an arbitrary cut-off of no less than five years of follow-up to define long-term studies to ensure robust event rates. We also only included studies that were published within the last 20 years. While this may have epidemiological advantages, there may have been landmark studies published before that time with still relevant information. Finally, there may be long-term observational studies of interest in other databases, although we only searched PubMed for our literature review.

We view this work to be the first step in establishing a comprehensive framework for valuing information from future clinical trials in heart failure. We have demonstrated the evidence basis of using the most commonly seen surrogate endpoints in heart failure clinical trials by examining their links to Level 1 outcomes, such as long-term survival in observational studies. However, future analyses are needed to examine the links between other less commonly used surrogate endpoints and long-term survival and quality of life, as well as links between one another among surrogate endpoints to complete the picture. Systematic reviews of risk models can help bring to light some of the gaps in these links. Our hope is that future designs of clinical trials on patients with heart failure would explicitly consider and discuss the links of their surrogate endpoints to long-term survival and quality-adjusted survival. Understanding potential impact of different trials on such comprehensive endpoints, such as value of information estimates, can offer funders of research a common quantitative metric to add to their repertoire of criteria used for resource allocation decisions.

Reference:

1. Meltzer D, Basu A, Meltzer HY. Comparative effectiveness research for antipsychotic medications: How much research is enough? *Health Affairs* 2009; 28(5): w794-w808.

2. Meltzer DO, Hoomans T, Chung JW, et al. Minimal Modeling Approaches to Value of Information Analysis for Health Research. *Medical Decision Making* 2011;31(6):E1-E22.

3. Carlson JJ, Thariani R, Roth J, Gralow J, Henry L, Esmail L, et al. Value of Information Analysis within a Stakeholder-Driven Research Prioritization Process in a US Setting: An Application in Cancer Genomics. Med Decis Mak. 2013;33(4):463–71.

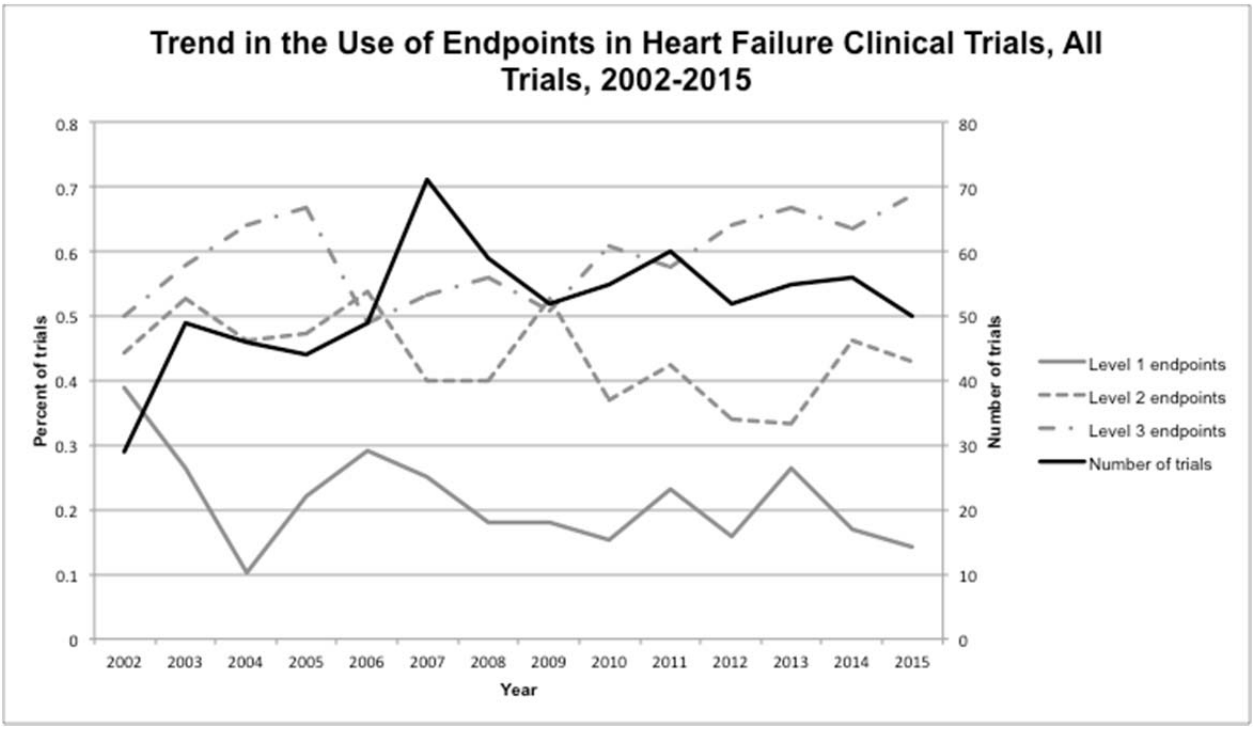
4. Claxton K, Griffin S, Koffijberg H, McKenna C. How to estimate the health benefits of additional research and changing clinical practice. BMJ 2015; 351:h5987.

5. Lauer MS. Investing in clinical sciences: Make way for (not so uncommon) outliers. Annals of Internal Medicine 2014; 60: 651-652.

6. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical Trial Evidence Supporting FDA Approval of Novel Therapeutic Agents, 2005-2012. JAMA. 2014;311(4):368.

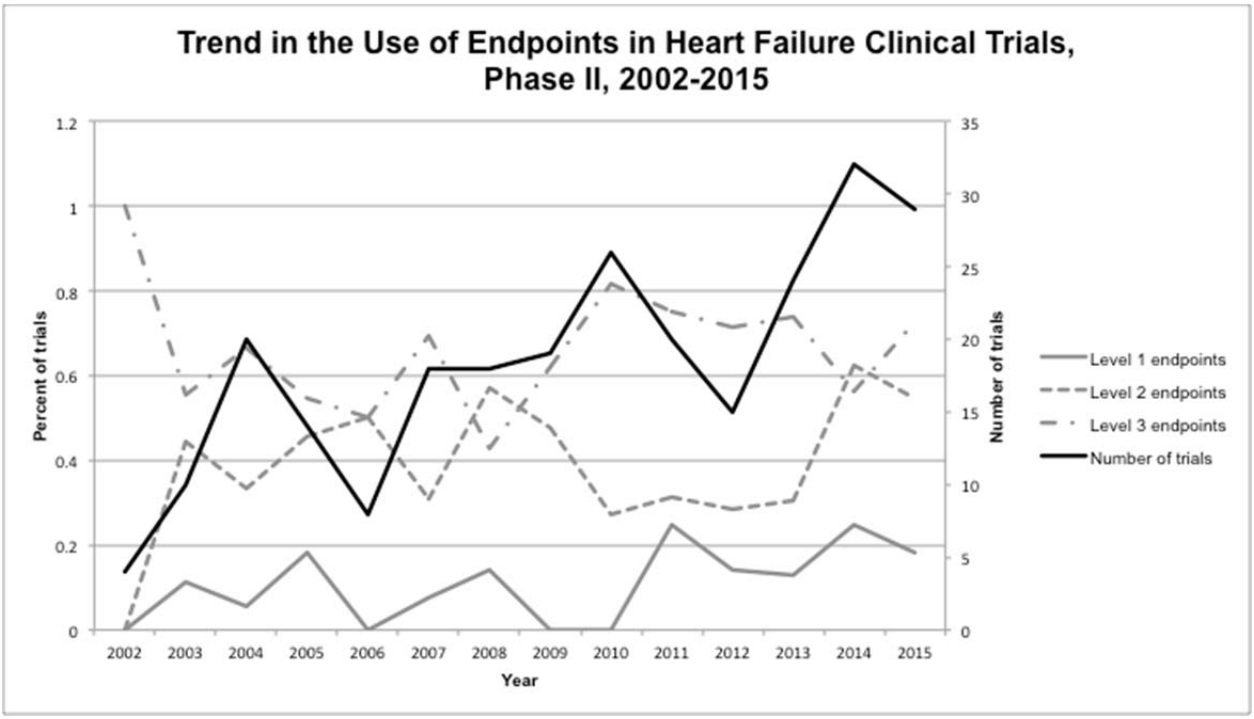
7. Fleming TR, Powers JH. Biomarkers and Surrogate Endpoints In Clinical Trials. Stat Med. 2012;31(25):2973–84.

Figure 1. Trend in the use of endpoints in all phase II-IV heart failure clinical trials



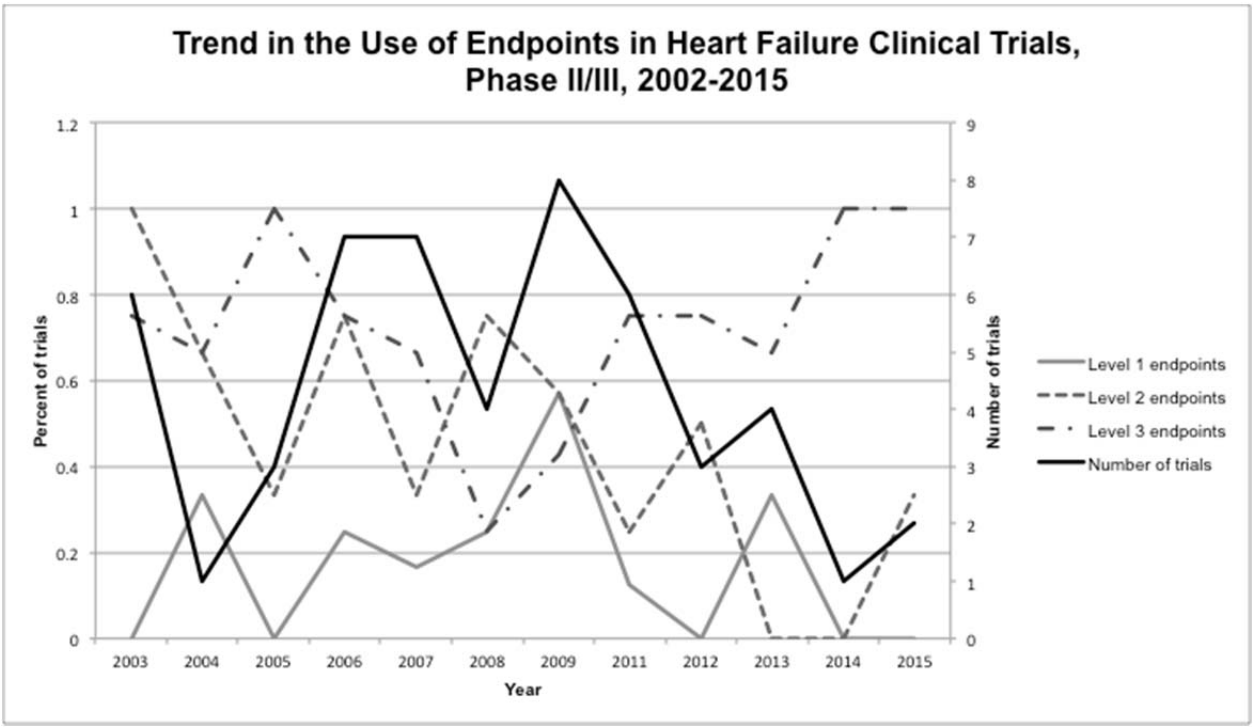
Level 1 endpoints included mortality either as an individual endpoint or as part of a composite endpoint and quality of life. *Level 2* included clinical endpoints such as myocardial infarction, stroke, arrhythmia, dyspnea, as well as emergency department visit, hospital admission, etc. *Level 3* include biological markers such as exercise capacity, maximum oxygen consumption, New York Heart Association status, left ventricular ejection fraction, brain natriuretic peptide, body mass index, pulmonary capillary wedge pressure, mitral regurgitation, blood pressure, left ventricular end systolic volume, urine output, serum creatinine, left ventricular function, glomerular filtration rate, heart rate, etc. Data from before 2002 were not included in the analysis of trend since there were too few trials registered on ClinicalTrials.gov before 2002 for any reliable estimate.

Figure 2. Trend in the use of endpoints in phase II heart failure clinical trials



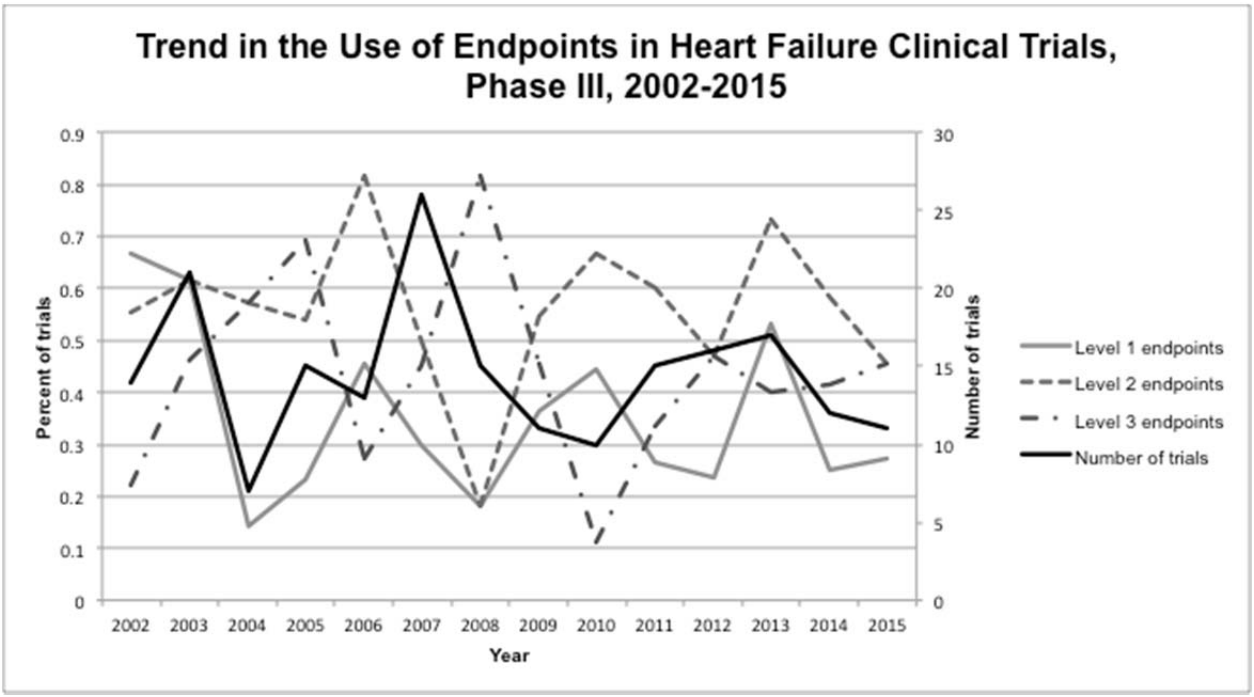
Level 1 endpoints included mortality either as an individual endpoint or as part of a composite endpoint and quality of life. *Level 2* included clinical endpoints such as myocardial infarction, stroke, arrhythmia, dyspnea, as well as emergency department visit, hospital admission, etc. *Level 3* include biological markers such as exercise capacity, maximum oxygen consumption, New York Heart Association status, left ventricular ejection fraction, brain natriuretic peptide, body mass index, pulmonary capillary wedge pressure, mitral regurgitation, blood pressure, left ventricular end systolic volume, urine output, serum creatinine, left ventricular function, glomerular filtration rate, heart rate, etc. Data from before 2002 were not included in the analysis of trend since there were too few trials registered on ClinicalTrials.gov before 2002 for any reliable estimate.

Figure 3. Trend in the use of endpoints in phase II/III heart failure clinical trials



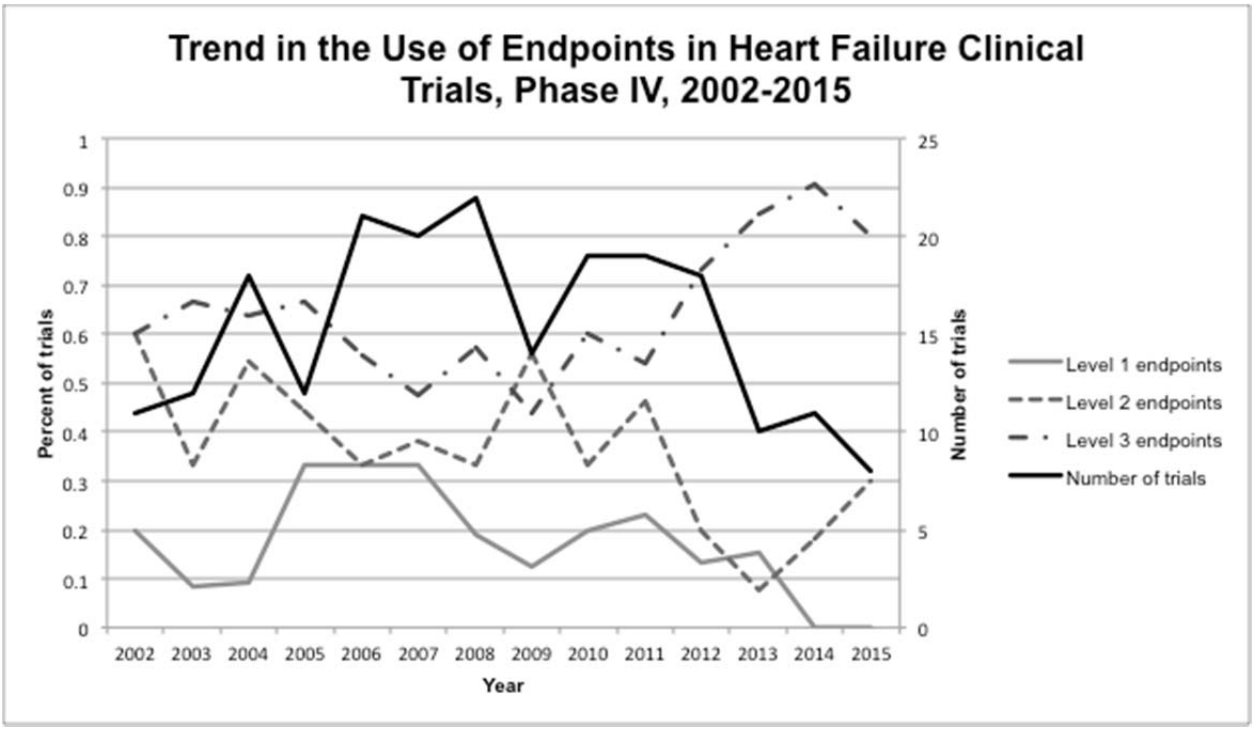
Level 1 endpoints included mortality either as an individual endpoint or as part of a composite endpoint and quality of life. *Level 2* included clinical endpoints such as myocardial infarction, stroke, arrhythmia, dyspnea, as well as emergency department visit, hospital admission, etc. *Level 3* include biological markers such as exercise capacity, maximum oxygen consumption, New York Heart Association status, left ventricular ejection fraction, brain natriuretic peptide, body mass index, pulmonary capillary wedge pressure, mitral regurgitation, blood pressure, left ventricular end systolic volume, urine output, serum creatinine, left ventricular function, glomerular filtration rate, heart rate, etc. Data from before 2002 were not included in the analysis of trend since there were too few trials registered on ClinicalTrials.gov before 2002 for any reliable estimate.

Figure 4. Trend in the use of endpoints in phase III heart failure clinical trials



Level 1 endpoints included mortality either as an individual endpoint or as part of a composite endpoint and quality of life. *Level 2* included clinical endpoints such as myocardial infarction, stroke, arrhythmia, dyspnea, as well as emergency department visit, hospital admission, etc. *Level 3* include biological markers such as exercise capacity, maximum oxygen consumption, New York Heart Association status, left ventricular ejection fraction, brain natriuretic peptide, body mass index, pulmonary capillary wedge pressure, mitral regurgitation, blood pressure, left ventricular end systolic volume, urine output, serum creatinine, left ventricular function, glomerular filtration rate, heart rate, etc. Data from before 2002 were not included in the analysis of trend since there were too few trials registered on ClinicalTrials.gov before 2002 for any reliable estimate.

Figure 5. Trend in the use of endpoints in phase IV heart failure clinical trials



Level 1 endpoints included mortality either as an individual endpoint or as part of a composite endpoint and quality of life. *Level 2* included clinical endpoints such as myocardial infarction, stroke, arrhythmia, dyspnea, as well as emergency department visit, hospital admission, etc. *Level 3* include biological markers such as exercise capacity, maximum oxygen consumption, New York Heart Association status, left ventricular ejection fraction, brain natriuretic peptide, body mass index, pulmonary capillary wedge pressure, mitral regurgitation, blood pressure, left ventricular end systolic volume, urine output, serum creatinine, left ventricular function, glomerular filtration rate, heart rate, etc. Data from before 2002 were not included in the analysis of trend since there were too few trials registered on ClinicalTrials.gov before 2002 for any reliable estimate.

Table 1. Primary endpoints used in at least 1% of 688* heart failure phase II-IV trials registered on ClinicalTrials.gov, 1996-2015

Individual Endpoint, N(%)	Overall (N=688)	Phase II (n=214)	Phase II/III (n=62)	Phase III (n=204)	Phase IV (n=208)
Level 1					
Composite outcome with mortality	101 (14.7)	13 (6.1)	8 (12.9)	55 (27.0)	25 (12.0)
Mortality as an individual outcome	35 (5.1)	6 (2.8)	5 (8.1)	14 (6.9)	10 (4.8)
Quality of life	30 (4.4)	10 (4.7)	2 (3.2)	11 (5.4)	7 (3.4)
Level 2					
Hospital admission	47 (6.8)	12 (5.6)	3 (4.8)	16 (7.8)	16 (7.7)
Myocardial infarction	28 (4.1)	11 (5.1)	2 (3.2)	7 (3.4)	8 (3.8)
Dyspnea	25 (3.6)	9 (4.2)	3 (4.8)	12 (5.9)	1 (0.5)
Length of stay	10 (1.5)	4 (1.9)	0 (0.0)	3 (1.5)	3 (1.4)
Emergency room visit	9 (1.3)	5 (2.3)	0 (0.0)	0 (0.0)	4 (1.9)
Arrhythmia	9 (1.3)	1 (0.5)	0 (0.0)	4 (2.0)	4 (1.9)
Level 3					
Exercise capacity	74 (10.8)	30 (14.0)	4 (6.5)	19 (9.3)	21 (10.1)
Maximal oxygen consumption	39 (5.7)	16 (7.5)	6 (9.7)	6 (2.9)	11 (5.3)
New York Heart Association status	15 (2.2)	2 (0.9)	1 (1.6)	5 (2.5)	7 (3.4)
Left ventricular ejection fraction	47 (6.8)	16 (7.5)	6 (9.7)	8 (3.9)	17 (8.2)
Brain natriuretic peptide	42 (6.1)	11 (5.1)	1 (1.6)	9 (4.4)	21 (10.1)
Body weight or body mass index	28 (4.1)	9 (4.2)	1 (1.6)	7 (3.4)	11 (5.3)
Pulmonary capillary wedge pressure	21 (3.1)	17 (7.9)	0 (0.0)	2 (1.0)	2 (1.0)
Mitral regurgitation	20 (2.9)	16 (7.5)	2 (3.2)	0 (0.0)	2 (1.0)

Blood pressure	18 (2.6)	7 (3.3)	0 (0.0)	4 (2.0)	7 (3.4)
Left ventricular end systolic volume	18 (2.6)	5 (2.3)	1 (1.6)	4 (2.0)	8 (3.8)
Urine output	17 (2.5)	5 (2.3)	1 (1.6)	3 (1.5)	8 (3.8)
Serum creatinine	15 (2.2)	3 (1.4)	2 (3.2)	4 (2.0)	6 (2.9)
Left ventricular function	11 (1.6)	4 (1.9)	2 (3.2)	1 (0.5)	4 (1.9)
Glomerular filtration rate	10 (1.5)	1 (0.5)	1 (1.6)	2 (1.0)	6 (2.9)
Heart rate	8 (1.2)	4 (1.9)	0 (0.0)	3 (1.5)	1 (0.5)

* From 1996-2015, 688 phase II-IV heart failure trials reported a total of 992 primary endpoints.

Table 2. Individual endpoints frequently used as part of a composite endpoint with mortality in phase II-IV heart failure trials

Individual Endpoint, N(%)	**Overall (N=101)**	**Phase II (n=13)**	**Phase II/III (n=8)**	**Phase III (n=55)**	**Phase IV (n=25)**
Hospital admission	68 (67.3)	5 (38.5)	4 (50.0)	35 (63.6)	24 (96.0)
Stroke	9 (8.9)	1 (7.7)	1 (12.5)	7 (12.7)	0 (0.0)
Myocardial infarction	7 (6.9)	1 (7.7)	1 (12.5)	5 (9.1)	0 (0.0)
Arrhythmia	4 (4.0)	1 (7.7)	1 (12.5)	1 (1.8)	1 (4.0)
Quality of life	3 (3.0)	0 (0.0)	0 (0.0)	1 (1.8)	2 (8.0)
Heart transplantation	3 (3.0)	1 (7.7)	1 (12.5)	1 (1.8)	0 (0.0)
Exercise capacity	2 (2.0)	2 (15.4)	0 (0.0)	0 (0.0)	0 (0.0)

Table 3. Number of trials that included at least one *Level 1*, *Level 2*, or *Level 3* endpoint

Endpoint Category, N(%)	Overall (N=688)	Phase II (n=214)	PhaseII/III (n=62)	Phase III (n=204)	Phase IV (n=208)
Level 1	155 (22.5)	26 (12.1)	14 (22.6)	75 (36.8)	40 (19.2)
Level 2	303 (44.0)	84 (39.3)	29 (46.8)	114 (55.9)	76 (36.5)
Level 3	397 (57.7)	142 (66.4)	40 (64.5)	87 (42.6)	128 (61.5)

Level 1 endpoints included mortality either as an individual endpoint or as part of a composite endpoint and quality of life. *Level 2* included clinical endpoints such as myocardial infarction, stroke, arrhythmia, dyspnea, as well as emergency department visit, hospital admission, etc. *Level 3* include biological markers such as exercise capacity, maximum oxygen consumption, New York Heart Association status, left ventricular ejection fraction, brain natriuretic peptide, body mass index, pulmonary capillary wedge pressure, mitral regurgitation, blood pressure, left ventricular end systolic volume, urine output, serum creatinine, left ventricular function, glomerular filtration rate, heart rate, etc.

Table 4. Statistical associations between frequently used surrogate endpoints in heart failure clinical trials and long-term survival

Individual Endpoint	**Number of estimates in PubMed**
Level 2	
Myocardial infarction	6
Stroke	5
Hospital admission	1
Level 3	
Left ventricular ejection fraction	23
Blood pressure	21
New York Heart Association status	15
Serum creatinine	11
Heart rate	6
Smoking	6
Body weight or BMI	5
Hemoglobin	5
Brain natriuretic peptide	4
Maximal oxygen consumption	3
Cholesterol	2
Angina	1
Cardiac index	1
Coenzyme Q10	1
Exercise capacity	1