Elsevier Editorial System(tm) for International Journal of Cardiology
Manuscript Draft

Manuscript Number: IJC-D-14-05386

Title: Composite outcome measures in a pragmatic clinical trial of chronic heart failure management: a comparative assessment

Article Type: Original Article

Keywords: Chronic heart failure; Composite outcome; Outcome assessment

Corresponding Author: Dr. Sungwon Chang, Ph. D., M. Stats., B. Sc.

Corresponding Author's Institution: University of Technology Sydney

First Author: Sungwon Chang, Ph. D., M. Stats., B. Sc.

Order of Authors: Sungwon Chang, Ph. D., M. Stats., B. Sc.; Patricia M Davidson, PhD; Phillip J Newton, PhD; Peter Macdonald, M.D., PhD; Melinda J Carrington , PhD; Thomas H Marwick , M.D., PhD; John D Horowitz , MBBS, PhD, FRACP; Henry Krum, MBBS, PhD, FRACP; Christopher Reid, PhD; Yih Kai Chan, PhD; Paul A Scuffham, PhD; David Sibbritt, PhD; Simon Stewart, PhD

Manuscript Region of Origin: AUSTRALIA

Abstract: Background: A number of composite outcomes have been developed to capture the perspective of the patient, clinician and objective measures of health in assessing heart failure outcomes. To date there has been limited examination in the composition of these outcomes.

Methods and Results: Three commonly used scoring systems in heart failure trials: Packer's composite, Patient Journey and the African American Heart Failure Trial (A-HeFT) scores were compared in assessing outcomes from the Which Heart failure Intervention is most Cost-effective & Consumer Friendly in Reducing Hospital Care ((WHICH(?)) Trial. Comparability and interpretability of these outcomes and the influence of each component to the final outcome were examined. Despite all three composite outcomes incorporating mortality, hospitalisation and quality of life (QoL), the contribution of each individual component to the final outcomes differed. The component with the most influence in deteriorating condition for the Packer's composite was hospitalisation (67.7%), while in Patient Journey it was QoL (61.5%) and for A-HeFT composite score it was mortality (45.4%).

Conclusions: The contribution made by each component varied in subtle, but important ways. This study emphasises the importance of understanding the value system of the composite outcomes to enable meaningful interpretation of results.

Suggested Reviewers:

Cover Letter

December 3, 2014

Professor Andrew Coates
Editor-in chief
International Journal of Cardiology

Dear Professor Coates,

Please find enclosed for your consideration my manuscript, ‘Composite outcome measures in a pragmatic clinical trial of chronic heart failure management: a comparative assessment’ for the upcoming issue of the *International Journal of Cardiology.*

Over the years, a number of composite outcome measures have been developed and used in clinical trials in chronic heart failure. This is in recognition of a need for an outcome that combines the totality of potential benefits and risks such as quality of life, hospitalization and mortality. Yet, little is known about the influences of each of these components on the final composite outcomes. This manuscript seeks to explore these issues by comparing and contrasting three composite outcomes used in heart failure, Packer’s composite, Cleland’s Patient Journey and A-HeFT score. Rather than suggesting which composite outcome is the best, we highlight methodological and interpretive problems in measurement of composite outcomes that may result from the differences due to the most influential component on each respective three composite outcomes. Subsequently, the *International Journal of Cardiology* seems an ideal forum for this manuscript. We hope this manuscript will stimulate interest in methodological innovation in development of the components of the composite outcomes.

This manuscript, I believe, raises the critical importance of methodological complexities in the weighting and calculation of components in measures of composite outcomes. Since this manuscript was written specifically for the *International Journal of Cardiology*, it has neither been previously published nor is currently under consideration for publication by any other journal. Furthermore, all named authors have seen and approved the final version of the manuscript.

I look forward to hearing from you in response.

Kind regards

Sungwon Chang

Author Agreement Form – International Journal of Cardiology

Manuscript Title: Composite outcome measures in a pragmatic clinical trial of chronic heart failure management: a comparative assessment

List of all Authors: Sungwon Chang, PhD, MStats, BSc; Patricia M Davidson, PhD; Phillip J Newton, PhD; Peter Macdonald, M.D., PhD; Melinda J Carrington, PhD; Thomas H Marwick, M.D., PhD; John D Horowitz, MBBS, PhD, FRACP; Henry Krum, MBBS, PhD, FRACP, FCSANZ; Christopher Reid, PhD; Yih Kai Chan, PhD; Paul A Scuffham, PhD; David Sibbritt, PhD; Simon Stewart, PhD

Corresponding Author: Sungwon Chang

This statement is to certify that all authors have seen and approved the manuscript being submitted, have contributed significantly to the work, attest to the validity and legitimacy of the data and its interpretation, and agree to its submission to the *International Journal of Cardiology.*

We attest that the article is the Authors' original work, has not received prior publication and is not under consideration for publication elsewhere. We adhere to the statement of ethical publishing as appears in the International of Cardiology (citable as: Shewan LG, Rosano GMC, Henein MY, Coats AJS. A statement on ethical standards in publishing scientific articles in the International Journal of Cardiology family of journals. Int. J. Cardiol. 170 (2014) 253-254 DOI:10.1016/j.ijcard.2013.11).

On behalf of all Co-Authors, the corresponding Author shall bear full responsibility for the submission. Any changes to the list of authors, including changes in order, additions or removals will require the submission of a new author agreement form approved and signed by all the original and added submitting authors.

All authors are requested to disclose any actual or potential conflict of interest including any financial, personal or other relationships with other people or organizations within three years of beginning the submitted work that could inappropriately influence, or be perceived to influence, their work. If there are no conflicts of interest, the COI should read: “The authors report no relationships that could be construed as a conflict of interest”.

Composite outcome measures in a pragmatic clinical trial of chronic heart failure management: a comparative assessment

Sungwon Chang[1,*]

Patricia M Davidson[1,2,*]

Phillip J Newton[1,*]

Peter Macdonald[3,*]

Melinda J Carrington[4,*]

Thomas H Marwick[5,*]

John D Horowitz[6,*]

Henry Krum[7,*]

Christopher M Reid[7,*]

Yih Kai Chan[4,*]

Paul A Scuffham[8,*]

David Sibbritt[1,*]

Simon Stewart[4,*]

On behalf of the WHICH Investigators

[1] University of Technology Sydney, Australia

[2] Johns Hopkins University, USA

[3] St Vincent's Hospital and Victor Chang Cardiac Research Institute, Sydney, Australia

[4] Baker IDI Heart and Diabetes Institute, Melbourne, Australia

[5] Menzies Research Institute, University of Tasmania, Australia

[6] The Queen Elizabeth Hospital and University of Adelaide, Adelaide, Australia

[7] Monash Centre of Cardiovascular Research and Education in therapeutics, Monash University, Australia

[8] Griffith Health Institute, Griffith University, Logan, Australia

[*] This author takes responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed presentation.

Trial Registration: Australian New Zealand Clinical Trials Registry number 12607000069459 (http://www.anzctr.org.au)

Corresponding Author:

SUNGWON CHANG

Faculty of Health, University of Technology Sydney; PO Box 123, Broadway NSW 2007, Australia.

Tel: +61 2 9514 4655 Fax: +61 2 9514 4474 Email: Sungwon.Chang@uts.edu.au

ACKNOWLEDGEMENT OF GRANT SUPPORT

SC is supported by a Chancellor's Post-Doctoral Research Fellowship from the University of Technology Sydney. The WHICH? Study is funded by the National Health and Medical Research Council of Australia (Number 418967) and individual investigators (SS, CMR and MC) are supported by the National Health and Medical Research Council of Australia.

CONFLICT OF INTEREST

The authors report no relationships that could be construed as a conflict of interest.

KEYWORDS

Chronic heart failure

Composite outcome

Outcome assessment

ABSTRACT

Background: A number of composite outcomes have been developed to capture the perspective of the patient, clinician and objective measures of health in assessing heart failure outcomes. To date there has been limited examination in the composition of these outcomes.

Methods and Results: Three commonly used scoring systems in heart failure trials: Packer's composite, Patient Journey and the African American Heart Failure Trial (A-HeFT) scores were compared in assessing outcomes from the Which Heart failure Intervention is most Cost-effective & Consumer Friendly in Reducing Hospital Care ((WHICH(?)) Trial. Comparability and interpretability of these outcomes and the influence of each component to the final outcome were examined. Despite all three composite outcomes incorporating mortality, hospitalisation and quality of life (QoL), the contribution of each individual component to the final outcomes differed. The component with the most influence in deteriorating condition for the Packer's composite was hospitalisation (67.7%), while in Patient Journey it was QoL (61.5%) and for A-HeFT composite score it was mortality (45.4%).

Conclusions: The contribution made by each component varied in subtle, but important ways. This study emphasises the importance of understanding the value system of the composite outcomes to enable meaningful interpretation of results.

INTRODUCTION

Chronic heart failure (CHF) is a common, complex and multifaceted syndrome.(1) Evaluating interventions in progressive, life limiting conditions challenge traditional approaches of outcome assessment, such as mortality and morbidity. Increasingly, CHF patients and clinicians alike are concerned not only with survival but also the quality of that survival. (2) Currently, there is a lack of consensus on appropriate measures to assess outcomes in clinical trials (3-7), whilst there is an increasing recognition that treatment efficacy needs to be measured by multiple outcomes, especially where management or the outcomes of interventions have various components.(8, 9) As reproducibility is challenged in clinical trials, understanding the reliability, validity and value of outcome assessments is important.

A composite outcome in a clinical trial is where clinically relevant measures are combined into a single outcome that can characterise clinically meaningful benefits of a treatment.(10) The benefits of composite outcomes include a reduced sample size as a consequence of increasing the event rate and hence lower costs of undertaking a trial, and the ability to capture the net benefits of the multiple dimensions into a single summary measure.(9, 11) Using a composite outcome will circumvent the need to make an allocation for multiple hypotheses testing, as one is essentially dealing with a single outcome.(10) In addition, the problem of competing risks can be avoided especially if a clinical outcome such as mortality is combined with morbidity.(10) A rationale for using composite outcome is well described in Neaton.(10) But more importantly, conceptually and theoretically composite outcomes seek to obtain the perspective of the patient, clinician as well as objective biomedical measures.

Composite outcomes are difficult to interpret when the treatment effects vary considerably across the components of the measure.(12) The most extreme case would be when the components are moving in different directions such as an increase in mortality and an improvement in QoL. The problem of interpretation is compounded when components are dissimilar in patient importance.(11) For example, it is useful to consider whether admission to the emergency room is comparable to a catastrophic stroke. Many of these problems may be resolved by choosing clinically relevant components of the composite and applying appropriate weightings of these components.(5, 11)

Using a composite outcome requires considerations, such as the selection of the number and type of clinically relevant components as well as their relative weightings or derivation methods which have important implications in the interpretation of the composite outcome.(5) Although the clinical and statistical challenges to using and interpreting composite outcomes have been discussed (13-16), there is limited discussion on the derivation method of composite outcomes or in establishing the standards for weighting components of a composite outcome.

This paper seeks to provide a better understanding of conceptual and measurement issues in composite outcome assessment by comparing and contrasting Packer's composite, (13) Cleland's Patient Journey (14) and the composite outcome used in the African American Heart Failure Trial (A-HeFT) (15), in a secondary analysis of a prospective, multi-centred randomised controlled trial. These composite outcomes have been chosen because they are commonly known composite outcome models in HF clinical trials and they capture the patient-centred components, namely mortality, hospitalisation and QoL from the perspective of the patient, clinician as well as including objective measures of health. The main objective of this paper was to compare and contrast these three composite outcomes to increase our understanding of the numerous pathways components influence the final outcome in CHF patients. Specifically, this paper does not aim to assess which composite outcome is 'best' or to assess the validity of these composite outcomes but rather to gain insight into the relationship among composite outcomes that measures similar patient-centred components. In addition, we sought to examine the methodological consequence of each component on the final outcomes.

METHOD

Composite outcome measures

Three commonly known composite outcome models were selected for the purpose of this comparative analysis. These were Packer's composite,(17) Cleland's Patient Journey (18) and the composite outcome used in the African American Heart Failure Trial (A-HeFT).(19) Although these composite outcomes incorporate similar components, each uses a different derivation method and/or

different weighting of the components likely reflecting priorities. All components in the composite were examined separately to estimate their relative effect on respective composite outcome.

The Packer's composite outcome measure (17) is perhaps the most widely used in clinical trials.(20) This score combines mortality, heart failure (HF) hospitalisation, change in New York Heart Association (NYHA) classification and a change in patient's global self-assessment of well-being measured in five discrete classes to classify patients as improved, unchanged, or worsened (Table 1)**.**

The Patient Journey (18) is another composite outcome in CHF incorporating information on mortality, hospitalisation and QoL/well-being. Furthermore, this measure incorporates the change in therapy in the scoring scheme.(18) Essentially this measure is a refinement of days alive and out of hospital (DAOH). It incorporates longevity and out of hospital into a single measure in days, and assign for each DAOH score of 100% if the patient reported feeling very good ('well-being' score 1). This score is subsequently reduced by 20% for each decrement in the patient-reported score down to a lowest potential score of 20% ('well-being' score 5). The intensification of diuretic therapy to control symptoms is also integrated by considering patients to be worse in the patient QoL/well-being than actually expressed.(18) However, a reduction in diuretic therapy does not necessarily lead to increase in the QoL/well-being.

The A-HeFT composite outcome (19) includes all-cause mortality, a first HF hospitalisation, and a change in QoL using Minnesota Living with Heart Failure questionnaire (MLWHFQ). A weight given to each component to generate the composite is shown in Table 1. Initial score of 0 is assigned to all patients, which will change depending on patient's experience.(19)

Weighting algorithms for composite scores

The weighting algorithms for each of the components for the composite outcomes are summarised in Table 1. Despite measuring similar concepts, each composite outcome captures and weighs each component differently. In addition, the measurements of the final outcomes were different. The Patient's Journey,(18) is well-being weighted DAOH, where the final outcome is expressed in days whereas the final outcome for the Packer's is a qualitative measure and the A-HeFT composite is a numeric score between -6 and 2.

The extent of differences in measuring and weighting for each components is apparent even in the hospitalisation. In the Packer's (17) and the A-HeFT scores (19), the incident of first HF hospitalisation is used whereas the Patient Journey (18) uses total hospitalisation days for all-causes. For the QoL component, not all composite outcomes use the same instruments and in some cases more than one instrument/measure are used. In the Packer's composite, changes in the NYHA functional class is combined with the changes in patient assessed global well-being, while in the Patient's Journey, increased use in diuretic adjusts QoL/well-being weights which is then applied to DAOH. For the A-HeFT score, greater weight is assigned to a change in QoL than for an incidence of first HF hospitalisation.

Discrete and comparative analysis of the three aforementioned composite outcomes were carried out using the data from Which Heart failure intervention is most Cost-effective & consumer friendly in reducing Hospital care (WHICH(?)), a multicentre randomised controlled study.(21) The main focus of the study was to compare the multidisciplinary CHF management delivered via an outreach, home-based intervention (HBI) with an outpatient or a specialised CHF clinic-based intervention (CBI), in patients with moderate to severe symptoms of HF with at least one admission for acute HF. A detailed description of the rationale and design, baseline findings and primary results for this trial has been published elsewhere.(21, 22)

The WHICH Trial

A total of 280 patients were recruited from three tertiary referral hospitals in three different states in Australia. Of these, 143 patients were randomised to the home-based and 137 to clinic-based post-discharge management. As previously described,(21) baseline characteristics were similar in the two arms. All hospitalisations were adjudicated on the type (elective/unplanned) and the cause, and all deaths were reviewed by a blinded outcome committee. The intervention was found to be not significant on the primary outcome of all-cause mortality or all-cause unplanned hospitalisation (22) during 12-18 months follow-up.

For the purpose of the current study, patients with follow-up greater than 12 months were censored at the date of contact at 12 months to ensure all patients had an equal follow-up duration. This was

necessary for Patient Journey where equivalent follow-up duration was required for all patients. The characteristics of the patients and the components for the three composite outcomes, namely mortality, hospitalisation and QoL at 12 months follow-up were calculated. Below the application of weighting algorithms is provided.

Estimation of Patient Global QoL Measure

For the Packer's composite, the HF specific QoL instrument, MLWHFQ was used in place of a patient global QoL assessment. The MLWHFQ is most widely used heart failure specific instrument with excellent psychometric properties.(23) This self-administered instrument focuses on the physical, socioeconomic and psychological effects of heart failure and its treatment on patients.(24) A minimal clinically important difference score in MLWHFQ is five.(25-27) Consequently, a change of 5 points was considered to be equivalent to a change in one class of patient global QoL assessment. By using the MLWHFQ instead of global patient well-being score, more specific and sensitive QoL measure (26-28) would have been included in the Packer's composite.

Similarly, in the Patient's Journey, instead of using the weights derived from five discrete points of the patient's well-being scales (Table 1), the Australian derived EQ-5D-3L index was used.(29) The EQ-5D-3L instrument is a widely used generic measure of QoL consisting of five components, mobility, self-care, usual activities, pain/discomfort and anxiety/depression with each having three levels.(30) The EQ-5D-3L has been shown to have satisfactory validity and reliability as an outcome measure in the cardiovascular area.(31, 32) The EQ-5D-3L would provide better utility indices than five discrete weights applied to patient global well-being.(33)

Statistical Analysis

Descriptive analysis in the form of counts (and percentage) for each components of the composite for nominal data and the mean, median, standard deviation and inter-quartile range for scale measures were found. Using the data from the WHICH(?) study(21) the final weights (or percentage) assigned to each component (mortality, hospitalisation and QoL) in relation to the total score were calculated for all three composite outcomes. These provide the magnitude of the influence each component has on the final composite outcome. To assess the difference between two arms of the WHICH (?)

study,(21) a Mann-Whitney nonparametric test was used for the A-HeFT scores and the Patient Journey and for the Packer's composite, the chi-square test was used.

The association between the composite outcomes were assessed by Spearman's rho (ρ) and Goodman Kruskal's Gamma (γ) (34) was used. The Patient's Journey as expressed as days lost were calculated for each categories of the Packer's composite and for the A-HeFT scores. In addition, to analyse the relationship between the Packer's and the A-HeFT score a Kruskall-Wallis test was used. The Mann-Whitney U test was performed with an adjusted alpha of 0.017 for multiple tests. All data analyses were performed with Statistical Package for Social Sciences for Windows version 19.0 (SPSS Inc, Chicago, Illinois).

RESULTS

In 12 months follow-up, a total of 57 (57/280; 20.4%) deaths were recorded in the WHICH Trial. Of these, 46 patients (46/57; 80.7%) had at least one unplanned hospitalisation where 39 (39/57; 68.4%) were for HF (Table 2). A total of 200 (71.4%) patients had all cause hospitalisation with 120 (60.0%) having multiple hospitalisation resulting in 3,715 hospital days. 111 (39.6%) patients were hospitalised specifically for worsening HF, resulting in 1,568 hospital days. Of these, 40 (40/111; 36.0%) had multiple HF hospitalisation. The mean duration of hospital stay for HF was 14.1 (±15.1, median=9.0, IQR=15.0) days. NYHA functional class indicated only 8 (2.9%) patients have deteriorated over 12 months follow-up, while 98 (35.0%) patients improved. Changes in MLWHFQ from baseline to 12 months follow-up indicated 51 (18.2%) patients have deteriorated, while 124 (44.3%) have improved in their condition (Table 2).

Modified Packer's composite

Packer's composite demonstrated 30.7% (n=86) of patients have improved, 10.7% (n=30) remained the same and 58.6% (n=164) of patients have worsened. Amongst those classified as improved, 51.2% (44/86) indicated improvement in both NYHA functional class and in MLWHF, suggesting some agreement between patients' and clinicians' assessment of patients' QoL.

The most common reason for being classified in worse category was HF hospitalisation (111/164; 67.7%), followed by deteriorating QoL (35/164; 21.3%) and mortality (18/164; 11.0%) (Table 3). Of note, amongst those who have been hospitalised (i.e. in worse category), 41.4% (46/111) of patients have indicated improvement in their QoL/NYHA class.

Modified Patient Journey

Overall, patients lost 40.9% of total days (41,676 days) from mortality, hospitalisation, QoL measure and a change in diuretic therapy. The largest proportion of days lost was from limited QoL (24,867 days; 59.7%) followed by mortality (12,354 days; 29.6%). Other reasons for days lost include all-cause hospitalisation (3,715; 8.9%) and adjustment for increased use in diuretics (740 days; 1.8%). Notably, the Patient Journey assessed only deteriorating conditions and not of improvement. It assumes all CHF patients have symptoms that can only have negative or no impacts on their QoL. (18) In this study, 14.6% (n=41) patients retained their maximum assigned day even after adjusting with QoL weights.

A-HeFT score

The A-HeFT composite consists of weighted scores for all-cause mortality, a first adjudicated HF hospitalisation, and a change in the QoL. The A-HeFT score indicated 39.5% (n=110) of patients had improved overall, 17.5% (n=49) remained the same and 42.8% (n=119) deteriorated. The mean A-HeFT score was -0.5 (SD=2.1; median =0.0; IQR =3.0). 23.7% (n=66) of patients achieved a score of 2, a highest possible score for A-HeFT composite. This score can only be achieved in the absence of all-cause mortality and HF hospitalisation, and a marked improvement in QoL scores. Only 1.1% (n=3) of patients scored -6, a lowest possible score, which is a result of marked worsening of QoL, first HF hospitalisation and subsequent death. With A-HeFT composite scores, improvement in patient's condition can only be expressed with an increase in QoL. However indication of deterioration (negative score) can be a result of mortality, first HF hospitalisation and worsening in QoL or any combination of these.

Despite all three composite outcomes incorporating mortality, hospitalisation and QoL, the contribution of each individual component to the final outcomes were different. Using the data from the WHICH(?) trial (21), the component with the most influence on the Packer's composite (17) was hospitalisation (67.7%) while for the Patient Journey (18) it was QoL (61.5%) and for the A-HeFT composite score (19) it was mortality (45.4%) (Table 3).

Application of Composite outcomes to compare CBI and HBI

No significant differences were found between two arms for each of three composite outcomes and their components (Packer's composite, p-value=0.50; A-HeFT score, p-value=0.30 and Patient Journey, p-value=0.21) (Supplementary Material).

Relationship between Packer's composite, Patient Journey and A-HeFT

The correlation coefficients demonstrated moderate associations amongst all three composite outcomes. The correlation between the Packer's composite and the Patient Journey was moderate (γ =0.49). For the worse category of Packer's composite, 50.9% of all potential days were lost due to mortality, hospitalisation and impaired QoL, while in the same category, 23.4% of the days were lost and in the better category, 28.2% (Table 4). This substantial difference in days lost in the worse category in contrast to the same and the better categories is driven by mortality (20.7% in worse while 0.0% in the same and better category) and hospitalisation (5.1% in worse, 1.5% in same and 1.6% in better category). Similar pattern and magnitude in proportion of days lost were observed for the same and better categories of Packer's composite (Table 4).

The correlation between the Patient Journey and the A-HeFT score was moderate (ρ = 0.54). For lower A-HeFT scores (from -6 to -3) more than 50% of days were lost to mortality, hospitalisation and impaired QoL driven mainly by days lost to mortality (Table 5). In fact, for A-HeFT scores between -6 and -3, the largest proportion of days lost was due to mortality followed by impaired QoL. However for A-HeFT scores between -2 and 2, the greatest days lost was from impaired QoL followed by mortality and hospitalisation.

The overall correlation between the Packer's and the A-HeFT score was γ =0.86. The Krusal-Wallis test was highly significant (χ^2 (2, N=278) = 156.967, p<001), indicating differences in A-HeFT scores among the three categories of Packer's composite. Three post hoc comparisons using the Mann-Whitney tests were all statistically significant (p<0.01), where a lowest median score was achieved in the worse category, followed by the same and then in the improved category (Table 6).

DISCUSSION

This study has compared and contrasted three composite outcomes, the Packer's composite,(17) the Patient Journey (18) and the A-HeFT score (19) which incorporates mortality, hospitalisation and QoL. The analysis has also examined the methodological issues in deriving each of composite outcomes to gain insights into the relationship amongst them. As there is no established gold standard for assessing the absolute effect on any outcome measure, it would be premature to assess which composite outcome is the 'best'. Most likely the 'best' measure would be the one that addresses the research question most appropriately.

Interestingly, all three composite outcomes provided similar result for the comparison between HBI and CBI arms. This may be due to synergies in the outcomes mortality, hospitalisation and QoL, despite differences in derivation methods used.

Packer's composite

The Packer's composite is perhaps the most well-known and widely used composite outcome in CHF.(20) Deriving this composite requires two stages. First, it involves monitoring patients until the event of mortality or first HF hospitalisation in the follow up period to classify them into worse category. Second, those who have not been hospitalised for HF or died (i.e. only those who have not been classified into worse category) will be assessed for a change in their QoL score and/or NYHA functional class. Depending on the magnitude and the direction of the change in QoL/NYHA, patients will be classified into worse, same or improved category. As such, this composite outcome provides only a qualitative assessment.

In the first stage of the derivation method, mortality and HF hospitalisation are considered to have the same weight despite the fact that patients may view these components of the composite outcome very differently. As HF hospitalisation occurs more frequently than mortality, the patient's final outcome may be determined more often by HF hospitalisation than less frequent but more serious component, mortality. A patient who has a single, short, early admission is placed into worse category similar to mortality, when in fact a short HF hospitalisation may reflect early detection of problems and hence a good care rather than an adverse outcome. This implies that hospitalisation does not necessarily indicate worse outcomes.

The information captured on HF hospitalisation is as an indicator variable only. Hence, crucial information such as duration and severity of the HF hospitalisation are not captured nor used in the derivation of the Packer's composite. Furthermore, the Packer's composite only considers the first HF hospitalisation, disregarding subsequent HF hospitalisations. In this study, 36.0% of patients had multiple HF hospitalisations.

In the second stage of categorisation, assessing changes in NYHA functional class and patient assessed QoL components would only apply to patients who have not been censored by mortality or HF hospitalisation. As such, mortality and HF hospitalisation is prioritised above QoL and the changes in QoL component would only determine the final outcome for those who have survived and have not been hospitalised for HF. In this study, 41.4% of patients who were hospitalised for HF and hence placed into worse category, reported improvement in their QoL. Not surprisingly, Packer's composite is most influenced by first HF hospitalisation rather than mortality or QoL components (Table 3).

One of the strengths of the Packer's composite is that it considers the change in QoL component from both the patient's and clinician's perspective. This however can potentially create a problem when they differ significantly or contradict each other. In this study, just over 50% agreement was observed between NYHA functional class and the patient's QoL assessment.

Patient Journey

The Patient Journey is different to other composite outcomes. The final outcome measure is the total days per comparison groups rather than the mean days. As such, in addition to ensure the duration of

the total follow-up days is comparable between groups, there is a need to ensure the total number of patients in each comparison arms to be similar.

In the metric of the Patient's Journey, the DAOH are adjusted by QoL weights and often these weights may have the greatest influence on the final outcome. Yet, these weights have not been validated (35) and in general, there would be disagreement among clinicians and patients about the value and the appropriateness of these weights. In present study, EQ-5D-3L index was used as QoL weights which may have been more sensitive and appropriate.(33)

Depending on the clinical condition of the patients, the Patient Journey may lead to a highly skewed outcome with many patients attaining near maximum potential days. In this study, 14.6% (n=41) patients kept their maximum potential days. Such skewed data are usually difficult to analyse and often less powerful nonparametric methods would need to be utilised.(5)

In the metric of the Patient Journey, QoL component is assessed on one occasion only and as such a change from baseline to the follow-up is not used in its calculation. This has the advantage of avoiding the problem of recall bias and of the possible variability that may result due to temporary changes in patient's QoL. However, if a substantial change in QoL component is to occur during the study period, this would not be captured in the final outcome of the Patient Journey. Furthermore, the Patient Journey can only examine the deteriorating condition and not of improvement.

A-HeFT score

A major strength of an A-HeFT score is that patients can contribute to all components of the outcome albeit only first HF hospitalisation is captured in its calculation. This would avoid multiple HF hospitalisations to add up to a score equivalent to death.(19) Indeed, in this composite outcome, death is considered as the worst outcome with the score of -3.

One of the interesting features of the A-HeFT score is a wide range of weights assigned to the changes in QoL; a substantial change in QoL has twice the weight assigned to a first HF hospitalisation. A major disadvantage of A-HeFT score is that the weights assigned to each

component have not been validated. Hence the clinically meaningful effect would be difficult to define, which would hinder clear interpretations of the result.

Although there was a moderate correlation between the Packer's composite and the Patient Journey, and also between the A-HeFT score and the Patient Journey in this study, there was no clear pattern when patients have improved or remained the same. In all three composite outcomes the focus was on deteriorating clinical status. Hence their use is limited. This is especially the case in the Patient Journey which only considers deteriorating condition. Even in the Packer's composite and the A-HeFT score, only component that would determine patients as improved or same is QoL component.

In this study, the only pair of scores with high correlations with some emerging pattern was between the A-HeFT score and the Packer's composite. This may be due to using same definition of components, namely all-cause mortality and first HF hospitalisation and QoL assessed in similar way, albeit with different scoring system.

In planning a study, one of the most important decisions that investigators make is the choice of the outcome measure. Besides aiming to include outcomes that are important to patients, providers and health care system, they need to consider the feasibility of measuring and interpreting the efficacy of the intervention (6). Hence in choosing the composite outcomes, understanding the value system of the composite will enable potential users to choose appropriately. In this study, the hospitalisation was the most influential component in determining deteriorating condition in Packer's ordinal composite score (17) while QoL component was in Patient Journey (18) and mortality in A-HeFT.(19) Furthermore, this study demonstrates how results can potentially be influenced by component measures selected. This information will aid in the interpretation of these composite outcomes as well as provide a rationale for the choice of the components to be included in composite outcomes.

Limitation

The present study is a secondary data analysis which is an important limitation. The patient global assessment was not available to be used in calculation of the Packer's composite (17) or the Patient Journey.(18) Consequently the results of Packer's composite (17) and Patient Journey (18) in this study are estimates of these composite outcomes. However, using the MLWHFQ instead of the

patient global assessment for Packer's composite (17) may have provided a more detailed description of emotional and physical aspects of QoL than the one-item patient's global assessment.(36) Similarly in calculation of the Patient Journey,(18) EQ-5D-3L (30) was used. Given EQ-5D-3L (30) provide better utility measure than arbitrary weights assigned to a five point scale,(36) the result would have provided better reflection of the patient experience.

The derivation of composite outcomes and the examination was limited to a single study (21). This may limit the generalisability of the findings. However, the aim of the study was to obtain a better understanding of the composition and performance of composite outcome assessment and not to examine validity of these composite outcomes.

Although each of the three composite outcomes Packer's,(11) Patient Journey (12) and A-HeFT (13) score use three similar components (mortality, hospitalisation and QoL), there is no validation study to ensure they measure same concepts, nor to compare against a gold standard for assessing the totality of the interventions. This study highlights the subtle differences in outcomes as a result of difference in weighting of components of composite outcomes. This analysis will likely inform the development of outcome measures in the future.

CONCLUSION

There is a widespread interest in using a composite outcome as a primary outcome in clinical trials to avoid multiplicity issues, obtain a comprehensive perspective of the patient experience, and pragmatically for reducing sample size. Moreover, in the context of complex chronic conditions, such as CHF assessing morbidity, mortality and QoL are very important. However, trials with a composite primary outcome can be complex and raise challenging issues in group comparisons and making recommendations for clinical practice. This study has examined the structural elements of composite outcomes consisting of patient centred components mortality, hospitalisation and QoL in a well-controlled clinical trial. Although, each of the composite outcomes has a varying degree of assigning weights to each component, there was a considerable agreement amongst these composite outcomes when estimating deteriorating condition but not when estimating improvements. Appreciating methodological issues in the derivation and interpretation of composite outcomes is

important in advancing the science of outcome measurement. This analysis emphasises the importance of achieving consensus in the weighting and calculation of components in measures of composite outcomes to allow comparison of results across clinical trials.

ACKNOWLEDGEMENTS

We thank all the cardiac nurses, health care professionals and patients who participated, and staff at Baker IDI who contributed to data management.

REFERENCE

1. Clark RA, McLennan S, Dawson A, Wilkinson D, Stewart S. Uncovering a hidden epidemic: a study of the current burden of heart failure in Australia. Heart Lung Circ. 2004;13:266-73.
2. Lewis EF, Johnson PA, Johnson W, Collins C, Griffin L, Stevenson LW. Preferences for quality of life or survival expressed by patients with heart failure. The Journal of heart and lung transplantation. 2001;20(9):1016-24.
3. Allen LA, Hernandez AF, O'Connor CM, Felker GM. End points for clinical trials in acute heart failure syndromes. J Am Coll Cardiol. 2009;53(24):2248-58.
4. Anand IS, Florea VG. End points in chronic heart failure clinical trials. Dialogues in Cardiovascular Medicine. 2010;15(2):81-102.
5. Cohn J, Cleland JGF, Lubsen J, Borer JS, Steg PG, Perelman M, et al. Unconventional end points in cardiovascular clinical trials: should we be moving away from morbidity and mortality? J Card Fail. 2009;15(3):199-205.
6. Allen LA, Spertus JA. End Points for comparative effectiveness research in heart failure. Heart Fail Clin. 2013;9(1):15-28.
7. Chang S, Newton P, Inglis S, Luckett T, Krum H, Macdonald P, et al. Are all outcomes in chronic heart failure rated equally? An argument for a patient-centred approach to outcome assessment. Heart Failure Reviews. 2014;19(2):153-62.
8. Allen LA, Gheorghiade M, Reid KJ, Dunlay SM, Chan PS, Hauptman PJ, et al. Identifying patients hospitalized with heart failure at risk for unfavorable future quality of life. Circulation: Cardiovascular Quality & Outcomes. 2011;4(4):389-98.
9. Buzney EA, Kimball AB. A critical assessment of composite and coprimary endpoints: a complex problem. J Am Acad Dermatol. 2008;59(5):890-6.
10. Neaton JD, Gray G, Zuckerman BD, Konstam MA. Key issues in end point selection for heart failure trials: composite end points. J Card Fail. 2005;11(8):567-75.
11. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007;334(7597):786-8A.

12. Zannad F, Garcia AA, Anker SD, Armstrong PW, Calvo G, Cleland JGF, et al. Clinical outcome endpoints in heart failure trials: a European Society of Cardiology Heart Failure Association consensus document. Eur J Heart Fail. 2013;15(10):1082-94.
13. Chi GYH. Some issues with composite endpoints in clinical trials. Fundam Clin Pharmacol. 2005;19(6):609-19.
14. Ferreira-González I, Permanyer-Miralda G, Busse JW, Bryant DM, Montori VM, Alonso-Coello P, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. J Clin Epidemiol. 2007;60(7):651-7.
15. Ferreira-Gonzalez I, Permanyer-Miralda G, Busse JW, Devereaux PJ, Guyatt GH, Alonso-Coello P, et al. Composite outcomes can distort the nature and magnitude of treatment benefits in clinical trials. Ann Intern Med. 2009;150(8):566-7.
16. Freemantle N, Calvert MJ. Interpreting composite outcomes in trials. BMJ. 2010;341.
17. Packer M. Proposal for a new clinical end point to evaluate the efficacy of drugs and devices in the treatment of chronic heart failure. J Card Fail. 2001;7(2):176-82.
18. Cleland JGF. How to assess new treatments for the management of heart failure: composite scoring systems to assess the patients' clinical journey. Eur J Heart Fail. 2002;4(3):243-7.
19. Taylor AL. The African-American heart failure trial (A-HeFT): rationale and methodology. J Card Fail. 2003;9(5, Supplement 3):S216-S9.
20. Arya A, Block M, Kautzner J, Lewalter T, Mortel H, Sack S, et al. Influence of home monitoring on the clinical status of heart failure patients: design and rationale of the IN-TIME study. Eur J Heart Fail. 2008;10(11):1143-8.
21. Stewart S, Carrington MJ, Marwick T, Davidson PM, Macdonald P, Horowitz J, et al. The WHICH? trial: rationale and design of a pragmatic randomized, multicentre comparison of home- vs. clinic-based management of chronic heart failure patients. Eur J Heart Fail. 2011;13(8):909-16.
22. Stewart S, Carrington MJ, Marwick TH, Davidson PM, MacDonald P, Horowitz JD, et al. Impact of home versus clinic-based management of chronic heart failure: The WHICH? (Which heart failure intervention is most cost-effective & consumer friendly in reducing hospital care) multicenter, randomized trial. J Am Coll Cardiol. 2012;60(14):1239-48.

23. Garin O, Ferrer M, Pont À, Rué M, Kotzeva A, Wiklund I, et al. Disease-specific health-related quality of life questionnaires for heart failure: a systematic review with meta-analyses. Qual Life Res. 2009;18(1):71-85.
24. Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota living with heart failure questionnaire: Reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. Pimobendan Multicenter Research Group. Am Heart J. 1992;124(4):1017-25.
25. Rector TS, Tschumperlin LK, Kubo SH, Bank AJ, Francis GS, McDonald KM, et al. Use of the living with heart failure questionnaire to ascertain patients' perspectives on improvement in quality of life versus risk of drug-induced death. J Card Fail. 1995;1(3):201-6.
26. Bennett SJ, Oldridge NB, Eckert GJ, Embree JL, Browning S, Hou N, et al. Comparison of quality of life measures in heart failure. Nurs Res. 2003;52(4):207-16.
27. Riegel B, Moser DK, Glaser D, Carlson B, Deaton C, Armola R, et al. The Minnesota Living With Heart Failure Questionnaire: sensitivity to differences and responsiveness to intervention intensity in a clinical population. Nurs Res. 2002;51(12131233):209-18.
28. Asadi-Lari M, Rao A, Gray D. Health-related quality-of-life tools in heart failure. Expert Review of Pharmacoeconomics & Outcomes Research. 2005;5(3):267-70.
29. Viney R, Norman R, King MT, Cronin P, Street DJ, Knox S, et al. Time trade-off derived EQ-5D weights for Australia. Value in Health. 2011;14(6):928-36.
30. EuroQuol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.
31. Ellis JJ, Eagle KA, Kline-Rogers EM, Erickson SR. Validation of the EQ-5D in patients with a history of acute coronary syndrome*. Curr Med Res Opin. 2005;21(8):1209-16.
32. Schweikert B, Hahmann H, Leidl R. Validation of the EuroQol questionnaire in cardiac rehabilitation. Heart. 2006;92(1):62-7.
33. Stevenson LW, Lewis E. Mapping the journey. J Am Coll Cardiol. 2006;47(8):1612-4.
34. Goodman LA, Kruskal WH. Measures of association for cross classifications. Journal of the American Statistical Association. 1954;49(268):732-64.
35. Cleland JGF, Charlesworth A, Lubsen J, Swedberg K, Remme WJ, Erhardt L, et al. A comparison of the effects of carvedilol and metoprolol on well-being, morbidity, and mortality

(the "Patient Journey") in patients with heart failure: a report from the Carvedilol Or Metoprolol European Trial (COMET). J Am Coll Cardiol. 2006;47(8):1603-11.

36. Jaarsma T, Lesman-Leegte GAT, Cleuren GVJ, Lucas CMHB. Measuring quality of life in heart failure: one versus multiple items. Netherlands Heart Journal. 2005;13(10):338-42.

Table(s)

Table 1. Derivation method of Packer's composite, Patent Journey, and A-HeFT composite

Dimension	Packer's composite	Patient Journey	A-HeFT composite
Mortality	All-cause mortality expressed as an indicator variable (N=0, Y=1)	Days dead: The number of days from all-cause mortality to the end of study.	All-cause mortality expressed as an indicator variable (N=0, Y=1)
Hospitalisation	First HF Hospitalisation expressed as an indicator variable (N=0, Y=1)	Days in hospital: Total time in hospital for all causes Add the durations of all individual hospital stay	First HF Hospitalisation expressed as an indicator variable (N=0, Y=1)
Quality of Life	Change in patient global assessment and change in NYHA functional class	Average NYHA functional class over the duration of the study moderated by the increased use of diuretics	Change in MLWHFQ from baseline to follow-up.
Derivation method	Patients are classified as worse, same or better as: Worse Experienced death or HF hospitalisation during the planned duration of treatment or reported worsening of their NYHA class* or global assessment by at least one class at the final visit compared to the baseline. Same	Initially, DAOH will be calculated. The patient journey incorporated a patient's functional status by allocating each day of the DAOH to the last known NYHA status of the patient for that day. Calculation of DAOH: Total days in the study: number of days from randomization until the	This composite outcome consists of composite score of weighted values of all-cause mortality, first HF hospitalisation and change in QoL score using MLWHFQ. Scoring scheme • All-cause mortality (at any time during the trial) (-3 points)

Dimension	Packer's composite	Patient Journey	A-HeFT composite
	Neither improved nor worse (ie. Did not experience death or HF hospitalisation and no change in patient global assessment of QoL or NYHA class) Better Experienced a *favorable change* in NYHA class or in the patient global assessment by at least one class from the baseline but did not experience death or HF hospitalisation during the course of the trial.	date of the final patient examination (if alive) or end of study. DAOH = Total days in the study – (days dead + days in hospital) A score of 100% was assigned for each DAOH if the patient reported feeling very good The score was reduced by 20% for each decrement in the patient-reported score down to a lowest potential score of 20%[29]	• First HF hospitalisation (adjudicated) (-1 point) • Change in QoL Improved by ≥10 units (2 points) Improved by 5-9 units (1 point) Change by <5 units (0 point) Worsened by 5-9 units (-1 point) Worsening by ≥10 units (-2 points)
Final Outcome	An ordinal outcome of • Worse • Same • Better	QoL/well-being (Symptom) adjusted DAOH	-6 to 2

A-HeFT, African American Heart Failure Trial; DAOH, Days Alive and Out of Hospital; QoL, Quality of Life; NYHA, New York Heart Association; MLWHFQ, Minnesota Living with Heart Failure Questionnaire

Table 2. Characteristics of individual component used in composite outcomes (Packer's score, Patient Journey and A-HeFT score) in 12 months follow-up (n=280)

Outcomes	n (%)
All cause death	57 (20.4)
Hospitalisation	
All cause	200 (71.4)
1 hospitalisation	80 (28.6)
> 1 hospitalisation	120 (42.8)
Length of stay – Mean (Median; SD)	18.6 (9.5; 21.4)
Unplanned	175 (62.5)
1 hospitalisation	83 (29.3)
> 1 hospitalisation	92 (33.2)
Length of stay – Mean (Median; SD)	17.8 (10.0; 20.5)
Heart Failure Hospitalisation	111 (39.6)
1 hospitalisation	71 (25.4)
> 1 hospitalisation	40 (14.2)
Length of stay – Mean (Median; SD)	14.1 (9.0; 15.4)
Change in MLWHF (Baseline – Followup) - Mean (SD)*	9.2 (1.0;22.4)
Improvement by 10 units or more	104 (37.1)
Improvement by 5-9 units	20 (7.1)
Change by <5 units	103 (37.1)
Worsening by 5-9 units	7 (2.5)
Worsening by 10 units or more	44 (15.7)
Change in the New York Heart Association functional class	
Improved by two class	19 (6.8)
Improved by one class	79 (28.2)
Same	174 (62.1)
Worsened by one class	8 (2.9)
Worsened by one class	0 (0.0)

Change in diuretic use	
Increase	29 (10.4)
Same	221 (79.0)
Decrease	30 (10.7)

*+ve value indicates improvement; MLWHF, Minnesota Living with Heart Failure; SD, Standard deviation

Table 3. Percentage contribution of each dimensions to Packer's score, Patient Journey and A-HeFT score for deteriorating conditions (n=280)

Dimensions	Percentage (%) contribution to deteriorating condition		
	Packer's score	Patient Journey	A-HeFT
Death	11.0	29.6	45.4
Hospitalisation	67.7	8.9	29.4
QoL	21.3	61.5	25.2
Total	100.0	100.0	100.0

A-HeFT, American Heart Failure Trial; QoL, Quality of Life

Table 4. Cross-tabulation of Patient Journey by Packer's score (n=280)

Patient Journey	Packer's score					
	Worse (n=164)		Same (n=30)		Better (n=86)	
	Total	%	Total	%	Total	%
Potential days	59,599		10,950		31,238	
Days lost to						
Death	12,354	20.7%	-	0.00%	0	0.0%
Hospitalisation	3,056	5.1%	166	1.5%	493	1.6%
Impaired QoL	14,497	24.3%	2,347	21.4%	8,023	25.7%
Diuretic adjustment	403	0.7%	53	0.5%	284	0.9%
Total days lost	30,310	50.9%	2,566	23.4%	8,800	28.2%
Patient Journey	29,289	49.1%	8,384	76.6%	22,438	71.8%

QoL, Quality of Life

Table 5. Cross-tabulation of Patient Journey by A-HeFT composite score (n=280)

Patient Journey	A-HeFT Score Days (%)								
	-6	-5	-4	-3	-2	-1	0	1	2
Potential days	1,095	-	11,315	11,278	12,308	7,263	17,871	15,967	23,960
Days lost to									
Death	269 (24.6)	-	7,003 (61.9)	4,642 (41.2)	392 (3.2)	48 (0.7)	-	-	-
Hospitalisation	81 (7.4)	-	676 (6.0)	732 (6.5)	324 (2.6)	446 (6.1)	338 (1.9)	782 (4.9)	281 (1.2)
Impaired QoL	246 (22.4)	-	1,181 (10.4)	2,387 (21.2)	3,885 (31.6)	2,057 (28.3)	4,424 (24.8)	50,807 (31.9)	5,438 (22.7)
Diuretic adjustment	-	-	-	70 (0.6)	131 (1.1)	75 (1.0)	106 (0.6)	97 (0.6)	236 (1.0)
Total days lost	596 (54.4)	-	8,860 (78.3)	7,831 (69.4)	4,732 (38.4)	2,627 (36.2)	4,869 (27.2)	5,966 (37.4)	5,956 (24.9)
Patient Journey	499 (45.6)	-	2,455 (21.7)	3,447 (30.6)	7,576 (61.6)	4,636 (63.8)	13,002 (72.8)	10,001 (62.6)	18,004 (75.1)

A-HeFT, African American Heart Failure Trial; QoL, Quality of Life

Table 6. Descriptive statistics of A-HeFT score by Packer's composite (n=280)

Packer score	A-HeFT scores				
	n	Mean	Median	SD	Range
Worse	164	-1.77	-2.00	1.89	8
Same	30	0.00	0.00	0.00	0
Better	86	1.62	2.00	0.71	2
Total	278	-0.53	0.00	2.14	8

A-HeFT, African American Heart Failure Trial; SD, Standard Deviation

e-component

Click here to download e-component: Supplementary material final.rtf