

## **COMPOSITE OUTCOMES: PROBLEMS AND SOLUTIONS THAT INCORPORATE QUALITY OF LIFE ADJUSTED SURVIVAL ANALYSES**

Dean T. Eurich, Sumit R. Majumdar, Finlay A. McAlister; Ross T. Tsuyuki, Yutaka Yasui, Jeffrey

A. Johnson

Dean T. Eurich, BSP, PhD

Assistant professor

Department of Public Health Sciences, School of Public Health, 13-103 Clinical Sciences

Building, University of Alberta, Edmonton, Alberta, Canada, T6G 2G3; [deurich@ualberta.ca](mailto:deurich@ualberta.ca)

Sumit R. Majumdar, MD, MPH

Associate Professor

Division of Internal Medicine, Department of Medicine, Faculty of Medicine and Dentistry, 2F1

WMC, University of Alberta Hospital, 8440 – 112 Street, Edmonton, Alberta, Canada, T6G 2B7;

[me2.majumdar@ualberta.ca](mailto:me2.majumdar@ualberta.ca)

Finlay A. McAlister, MD, MSc

Associate Professor

Division of Internal Medicine, Department of Medicine, Faculty of Medicine and Dentistry, 2F1

WMC, University of Alberta Hospital, 8440 – 112 Street, Edmonton, Alberta, Canada, T6G 2B7;

[finlay.mcalister@ualberta.ca](mailto:finlay.mcalister@ualberta.ca)

Ross T. Tsuyuki, BSc(Pharm), PharmD, MSc

Professor

Division of Cardiology, Faculty of Medicine and Dentistry, 2F1 WMC, University of Alberta

Hospital, 8440 – 112 Street, Edmonton, Alberta, Canada, T6G 2B7; [ross.tsuyuki@ualberta.ca](mailto:ross.tsuyuki@ualberta.ca)

Yutaka Yasui, PhD

Professor

Department of Public Health Sciences, School of Public Health, 13-103 Clinical Sciences

Building, University of Alberta, Edmonton, Alberta, Canada, T6G 2G3;

[yutaka.yasui@ualberta.ca](mailto:yutaka.yasui@ualberta.ca)

Jeffrey A Johnson, PhD

Professor

Department of Public Health Sciences, School of Public Health, 13-103 Clinical Sciences

Building, University of Alberta, Edmonton, Alberta, Canada, T6G 2G3; [jeff.johnson@ualberta.ca](mailto:jeff.johnson@ualberta.ca)

Work was performed at the University of Alberta, Edmonton, AB, Canada

**Mail Correspondence and Reprint Requests to:**

Dr. Dean Eurich  
Department of Public Health Sciences  
School of Public Health  
13-103 Clinical Sciences Building  
University of Alberta  
Edmonton, Alberta, Canada, T6G 2G3  
Tel: (780) 492-6333 Fax: (780) 492-7455  
Email: [deurich@ualberta.ca](mailto:deurich@ualberta.ca)

**Word Count (total): 4842; Word Count (text): 2901; Word Count (Abstract): 250**

**Tables: 2; Figures: 2; Appendix 2;**

**Running Title:** QAS Analysis of Composite Outcomes

## ABSTRACT

**Background:** Composite outcomes that weigh each component equally are commonly used to study treatment effects. We hypothesized that each component of the composite outcome would impact differentially on patients' overall health-related quality of life (HRQL) and tested this in two published heart failure (HF) studies.

**Methods:** We adapted the Q-twist methodology to incorporate HRQL data and accommodate the differential weights for composite outcome components (e.g., death or hospitalization). For each of two HF studies, the composite outcome was partitioned into its components and assigned literature-derived utility weights. Total quality-adjusted survival (QAS) time was determined for each treatment and compared with results from traditional survival analyses.

**Results:** In an observational study of metformin in HF, there was a reduced hazard of the composite outcome of 'death or hospitalization' [160 (77%) events compared to 658 (85%) events in sulfonylurea users; HR 0.83 (95%CI 0.70-0.99)]. Net gain of 0.82 years (95%CI 0.26-1.37) with traditional survival analyses; the difference in QAS time was a smaller 0.54 years (95%CI 0.20-0.89). In a randomized study of digoxin, there was a reduced hazard of the composite outcome [1291 (38%) events compared to 1041 (31%) events in the placebo group; HR 0.75 (95%CI 0.69-0.82)]. A net gain of 0.10 years (95%CI 0.02-0.16) but this time, the difference in QAS time was greater at 0.14 years (95%CI 0.07-0.19).

**Conclusion:** Studies assuming equal weighting of composite outcome components may potentially misrepresent treatment effects. By incorporating HRQL into survival analyses, the impact of the composite outcome components can be assessed more directly.

Composite outcomes are common in clinical research, especially in cardiovascular disease. Potential advantages of using composite outcomes include increased event rates with improved power and statistical efficiency and integration of clinically important events into a single quantifiable outcome. However, several concerns also exist, particularly in interpretation of results when the components of the composite outcome impact patients' health differentially or are associated with competing risks.<sup>(1)</sup> For example, the composite outcome of 'death or hospitalization' is often used in studies of heart failure. Difficulties in comparing the effects of two treatments could exist if Treatment A reduced mortality but increased hospitalizations while Treatment B had no effect on mortality but reduced hospitalizations.

The concern is with most composite outcomes is the assumption that each component of the outcome (i.e., each health state) is equally important. Although valid in certain cases, more often the different components of a composite outcome impact patients' overall quality of life differently. One approach to address this situation is to account for potential differences in health states by assigning differential weights to the components. This has often been attempted where the weights assigned are based on expert opinion.<sup>(2-5)</sup> However, controversy exists as to the appropriateness of these expert-derived weights, which has likely contributed to the limited adoption of weighted methods in the literature. Since patients ultimately experience the events in question, it seems reasonable to incorporate the patients' perspective when weighting composite outcomes in clinical research.

A method for assigning weights that incorporates the patients' perspective is the use of health related quality of life (HRQL). By incorporating HRQL into survival analyses, index measures (i.e., utilities) could adjust for the unequal impact of health states used in composite outcomes providing a 'weighted' survival outcome that incorporates different degrees of quality and

quantity of life.(6) Furthermore, incorporation of HRQL into survival analysis would align clinical research with the methods recommended for economic evaluations(7), thereby improving assessment of health care interventions. This concept of using HRQL to adjust survival analyses for different health outcomes has been used in oncology(6), where it is referred to as Quality Adjusted Survival (QAS) Analysis.(8) It has not, to our knowledge, been employed in cardiovascular disease, specifically among patients with heart failure. Therefore, we introduce this method using heart failure as our base example since many heart failure studies evaluate the composite outcome of ‘hospitalization or death’ as the primary endpoint. Most patients and clinicians, however, would consider these disparate endpoints and therefore analyses of such composite outcomes may be improved using the QAS approach.

## **Methods**

### Overview

We adapted and applied the QAS technique (8) to the commonly reported composite outcome of ‘death or hospitalization’ using two previously published studies - an observational study of antidiabetic therapies in diabetes and heart failure (Metformin Study)(9) and a randomized controlled trial comparing digoxin with placebo in patients with heart failure (DIG Study).(10) A summary of each study and a detailed overview of the QAS method are found in Appendix A. Ethical approval was granted for both studies by the University of Alberta; data for the Digitalis Investigation Group (DIG) study were obtained from the National Heart, Lung and Blood Institute.

### Weighted Composite Outcome

We applied the QAS analysis to the composite outcome of “all-cause mortality or hospitalization” in the metformin study and to the composite outcome of “heart failure-specific mortality or hospitalization” in the DIG study. In both of our case studies, we considered three potential successive health states patients may transition through: 1) state  $H_1$  represents the patients’ initial health state prior to a hospitalization, death or censoring at study end (i.e., the mean event-free survival time corresponding to the area under the composite outcome curve); 2) state  $H_2$  represents the patients’ health state after a hospitalization until either death or censoring at study end (i.e., area between the composite outcome survival curve and the mortality survival curve); 3) state  $H_3$  was the final health state dead (no time is associated with this health state).

For the QAS analyses, we applied validated utility coefficients for heart failure patients taken from the literature as HRQL measurements had not been collected in either of our case studies. For state  $H_1$ , the utility coefficient was set at 0.81, based on Health Utilities Index Mark 3 (HUI3) scores observed for subjects with heart disease and diabetes in the Canadian population(11), and is similar to utility weights observed in patients with heart failure alone.(12) In patients with heart failure, hospitalization reduces HRQL by 30% (12) resulting in a utility coefficient of 0.57 for state  $H_2$ . By convention, the health state dead ( $H_3$ ) was assigned a utility score of 0.(11) For the traditional survival analysis that assumes each health state (i.e.,  $H_1$  and  $H_2$ ) is of equal importance, a utility equal to 1 (i.e., perfect health) was used.

### Analysis

For the metformin cohort study, the transitional survival functions for each health state (i.e., all-cause death or all-cause hospitalization) for the metformin and sulfonylurea groups were

estimated using Cox proportional hazards models (i.e., PHREG) with adjustments for the confounding covariates (Appendix A) and the product-limit estimate of the baseline survival function.(8) The survival functions for the treatment groups were estimated using overall mean values for each covariate. For the DIG Study, there were no significant differences between the baseline characteristics of the digoxin and placebo patients due to the randomized design.(10) As a result, the survival functions were estimated using Cox proportional hazards models without covariate adjustments.

For each health state, the mean time spent in the health state was calculated by integrating the estimated survival function from zero to the maximum follow-up of 8.0 years for the metformin study and 4.8 years for the DIG Study (i.e., last composite event occurrence). Mean survival time in state  $H_1$  was estimated by the integrated survival time for the composite outcome. Mean survival time in state  $H_2$  was calculated by integrating the overall survival time of death and subtracting the mean time spent in state  $H_1$ . QAS time in each health state was then calculated by multiplying the mean time spent in each health state by the respective utility coefficient.(8) Summation of these quality adjusted times provided an estimate of the overall QAS time during the study period for each treatment group.

To generate estimates of the variability of the mean QAS time for the treatment groups, we used 500 bootstrap samples for both studies with the 95 percent confidence interval (CI) and corresponding p-values calculated according to the percentile method.(13)

### *Sensitivity Analysis*

Sensitivity analyses were performed to evaluate the effect of different utility coefficients for the health states on the results for both the metformin and DIG Study. The utility coefficients for

health states  $H_1$  and  $H_2$  were varied by increasing or decreasing the utility coefficients by 0.03, considered a clinically important difference on the HUI3.(14) All analyses were conducted using SAS for Windows Version 9.1, Cary, NC, USA.

## Results

### *i. Metformin use in Heart Failure*

Using traditional survival analysis, which assumes that those who are alive are in perfect health (i.e., utility of 1 for all health states) until the time of death, sulfonylurea users would have an expected mean adjusted survival of 4.10 years and the metformin users would have an expected mean adjusted survival of 4.92 years in this dataset. This translates into an average gain of 0.82 life-years (95% bootstrapped CI 0.26-1.37) associated with metformin use compared to sulfonylurea use. However, sulfonylurea users spent a mean of 1.21 years in health state  $H_1$  at an expected utility of 0.81 and 2.89 years in health state  $H_2$  at an expected utility of 0.57, for a total QAS time of 2.63 years (Figure 1a, Table I). Conversely, metformin users spent a mean 1.52 years in health state  $H_1$  and 3.40 years in health state  $H_2$ , for a total QAS time of 3.17 years (Figure 1b, Table I). As a result, metformin users exhibited a net increase of 0.54 quality adjusted life years (QALYs) as compared to the sulfonylurea therapy users, which is statistically significant (95% bootstrapped CI 0.20 – 0.89,  $p < 0.002$ ). In this case, traditional survival analysis method led to a 34% over-estimate of potential benefits related to treatment of diabetic heart failure patients with metformin.



## ***ii. DIG Study***

Mean survival times were 4.34 years for the placebo group and 4.44 years for the digoxin group; for the composite outcome, estimated total hospitalization-free survival time was 3.32 years for the placebo group and 3.66 years for the digoxin group (Figure 2a and 2b, Table II). Assuming perfect health prior to death as in traditional survival analysis, digoxin users had an apparent net survival benefit of 0.10 years (95% bootstrapped CI 0.02-0.16) compared to placebo. However, the placebo users spent a total time of 3.32 years in health state  $H_1$  at an expected utility of 0.81 and 1.02 years at an expected utility of 0.57 for health state  $H_2$ , resulting in a total QAS time of 3.27 years. Conversely, the digoxin group spent a total time of 3.66 years in health state  $H_1$  and 0.78 years in health state  $H_2$ , resulting in a total QAS time of 3.41 years. Therefore, after taking into account the greater time spent before first hospitalization, the digoxin group had a net gain of 0.14 QALYs compared to the placebo group, which is statistically significant (95% bootstrapped CI 0.07-0.19,  $p < 0.002$ ). Furthermore, this estimate is 40% higher than the standard survival estimate based on equally weighted outcomes. In this case, traditional analytic methods led to an under-estimate of potential benefits related to treatment of heart failure patients with digoxin.

## ***Sensitivity analyses***

In the base-case analyses described above, we applied deterministic utility coefficients to our health states taken from the literature. Therefore, we conducted a sensitivity analysis to assess the effect of varying the utility coefficients by a clinically important difference (0.03) on the QALY estimates for health state  $H_1$  and  $H_2$  for the metformin and DIG studies. Variations in the utility coefficients resulted in changes to both the individual QALYs calculated for each health

state and consequently to the differences between study groups. All sensitivity analyses confirmed that the standard survival estimates over-estimated benefits by 30% to 37% for the observational metformin study and under-estimated benefits of digoxin by 30% to 60% in the DIG study (See Appendix B for full sensitivity results).

## **Discussion**

Using heart failure as an example, we have adapted and demonstrated a method that quality-adjusts survival to deal with the potentially unequal impact of the individual components of composite outcomes. By incorporating into survival analyses an estimate of the impact on patients HRQL, a weighted composite outcome may provide a more patient-important representation of the benefits (or harms) associated with a therapy. In the two case studies we examined, incorporation of HRQL into the survival analyses resulted in quality adjusted survival estimates which ranged from 34% lower to 40% higher than those derived from traditional survival analyses using equally weighted components for a composite outcome.

Composite outcomes are commonly reported, but the tacit assumption that each component is of equal importance to patients, providers, and payers seems untenable. Moreover, increasing evidence suggests that the use of equal weights may lead to biased conclusions and consistently lead to overestimates of treatment effects in cardiovascular-related trials.<sup>(5)</sup> Our results also illustrate that bias may occur in either direction.

It is important to realize, however, that our primary focus in undertaking this study was to examine the differences in the magnitude of the QAS estimates between the treatment groups (e.g., digoxin or placebo) and determine how differential weighting of composite outcomes changes the interpretation of the data. The QAS approach provides survival time associated with

'perfect health' as opposed to overall survival in the traditional unweighted approach. For example, the DIG trial reported that the risk of heart failure hospitalization or death was significantly lower in the digoxin group [HR 0.75 (95% CI 0.69-0.82)].<sup>(10)</sup> However, the analysis we outlined above quantifies this benefit as a gain of only 0.10 life years using the traditional unweighted approach and only 0.14 quality adjusted life years (QALYs) using the weighted QAS approach – a difference of perhaps questionable clinical significance in this case, however this may not always be the case. Conversely, the gain of 0.54 QALYs observed with metformin therapy is similar to that observed with the use of other proven efficacious therapies versus placebo in heart failure.<sup>(15)</sup> Further, the difference in the QAS estimates for the unweighted and weighted approach for metformin therapy (0.82 vs 0.54) would be similar to the HRQL effects of choosing not to prescribe ACE inhibitors in patients with heart failure.<sup>(15)</sup> Although in our examples, both therapies remained beneficial after adjustment for HRQL, one could easily foresee clinical trials where marginal benefits using unweighted composite outcomes could be non-significant using the weighted approach or vice versa. In addition, the QAS approach is directly amendable to economic evaluations (e.g., cost utility analysis which incorporate QALYs) unlike most traditional survival analyses evaluating composite outcomes. Thus, by incorporating measures of clinical impact into summary effect estimates, we believe that better informed treatment decisions may result.

Importantly, however, appropriate selection of utility weights is required, as shown by the modest effect on the QAS estimates in our sensitivity analyses. Prospective collection of utilities in the setting of randomized controlled trials or cohort studies would permit direct incorporation of study-specific utilities into the QAS method. For example, by collecting utility estimates at baseline, at the time of clinically important events and at regularly scheduled time periods (e.g.,

every 6 months) a more complete evaluation of the effect of therapies on HRQL in clinical trials could be completed. Further, since the periods of evaluation are successive, these time specific health states could be easily incorporated into the QAS approach. In studies where prospective collection of utility estimates is not possible, carefully selected literature-based estimates or threshold utility analyses for unknown utility estimates(6;8) may be used but they should be justified and subject to sensitivity analyses such as we illustrated.

Although there are several advantages to using QAS analyses, there are also some limitations. First, the quality adjusted survival estimate we described must be restricted to a set time limit. Thus, in our analyses, the integration range was limited to the maximum follow-up of 8.0 years for the metformin study and 4.8 years for the DIG study - identical to that used in the traditional survival analysis. As such, this method cannot provide 'lifetime' estimates associated with the therapy in itself, but could be incorporated into projected life expectancy models to provide these estimates. Second, we used the simplest presentation of this method, which assumes a progressive health state model where hospitalization preceded death. Although this is appropriate for many disease conditions and most composite outcomes, it may not be suitable for all (for example, consider a composite outcome of hospitalization or coronary artery bypass grafting). Parametric methods have been developed, however, that overcome these limitations. (16) In addition, these models may be extended to account for repeated failure times and the use of time-varying covariates.(8) Third, all limitations and assumptions associated with Cox proportional hazards regression also apply to the QAS method.(17) Fourth, we used literature derived utility estimates and therefore these weights may not be representative for our study populations. Importantly, other preference based utility measures have been shown to be responsive to HRQL changes in patients with heart failure.(18) Therefore, although the actual

utility weights assigned to the respective health states may vary, the relative change in HRQL and associated differences in QAS between groups would be expected to be similar irrespective of utility measure; this is supported by the sensitivity analysis utilizing different utility coefficients. In addition, we have assumed that utilities remain constant within each health state over time. Limited evidence exists but it is likely improvements and decrements in HRQL are observed in patients with heart failure over time; however, given the high mortality in patients with heart failure, decrements in HRQL over the longer term would still be expected. Thus, this may have potentially over- or underestimated the true impact of the health state on patients. As a result, prospective collection of health utilities should be considered in heart failure trials, as previously suggested. Further, the QAS method incorporates average group utility estimates and not individual patient utilities. As a result, although this will provide appropriate estimates for overall treatment effects, it may not be appropriate for assessment of treatment impact at the individual level. Finally, our examples were restricted to one type of composite outcome evaluated for only one common condition. However, this method can be extended to incorporate multiple endpoints which comprise the composite outcome. Although computationally more complex, conceptually it is easiest to consider that in the area of cardiovascular research, most composite outcomes are comprised of non-fatal and fatal endpoints. Thus, complex composite outcomes could be distilled into simple 2 health state models with the utility assigned to the non-fatal endpoint weighted according to the distribution of the non-fatal events.

In conclusion, thoughtful and well-constructed composite outcomes are important in observational studies and clinical trials, but traditional methods of analysis may not show the whole picture. By incorporating patient-reported HRQL into survival analyses, the potential impact of the individual components of the composite outcome on the patient's health can be

assessed more directly; therefore, the potential benefits, harms, or costs associated with therapy may be more precisely quantified and transparent to patients, providers, and policy-makers.

## **Funding**

Dr. Eurich held a Doctoral Research award with Canadian Institutes for Health Research (CIHR). Dr. Majumdar and Dr. McAlister are New Investigators with the Canadian Institutes for Health Research (CIHR) and are Health Scholars with the Alberta Heritage Foundation for Medical Research (AHFMR). Dr. McAlister and Dr. Tsuyuki hold the Merck Frosst/Aventis Chair in Patient Health Management. Dr. Yasui holds a Canada Research Chair in Biostatistics and Epidemiologic Methods and is a Health Scholar support by AHFMR. Dr. Johnson holds a Canada Research Chair in Diabetes Health Outcomes, is a Health Scholar supported by AHFMR, and is Chair of a New Emerging Team (NET) grant to the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD). The ACHORD NET grant is sponsored by the Canadian Diabetes Association, the Heart and Stroke Foundation of Canada, The Kidney Foundation of Canada, the CIHR - Institute of Nutrition, Metabolism and Diabetes and the CIHR - Institute of Circulatory and Respiratory Health.

These study sponsors did not play any role in study design or conduct; collection, analysis, interpretation of data; writing of the report; or in the decision to submit the paper for publication.

We would also like to extend our gratitude to Qi Lui for her assistant with the data analysis.

This study is based on non-identifiable data provided by the Saskatchewan Department of Health. The interpretation and conclusions contained herein do not necessarily represent those of the Government of Saskatchewan or the Saskatchewan Department of Health.

The Digitalis Investigation Group (DIG) is conducted and supported by the NHLBI in collaboration with the DIG Study Investigators. This Manuscript was prepared using a limited access dataset obtained from the NHLBI and does not necessarily reflect the opinions or views of the DIG study or the NHLBI.

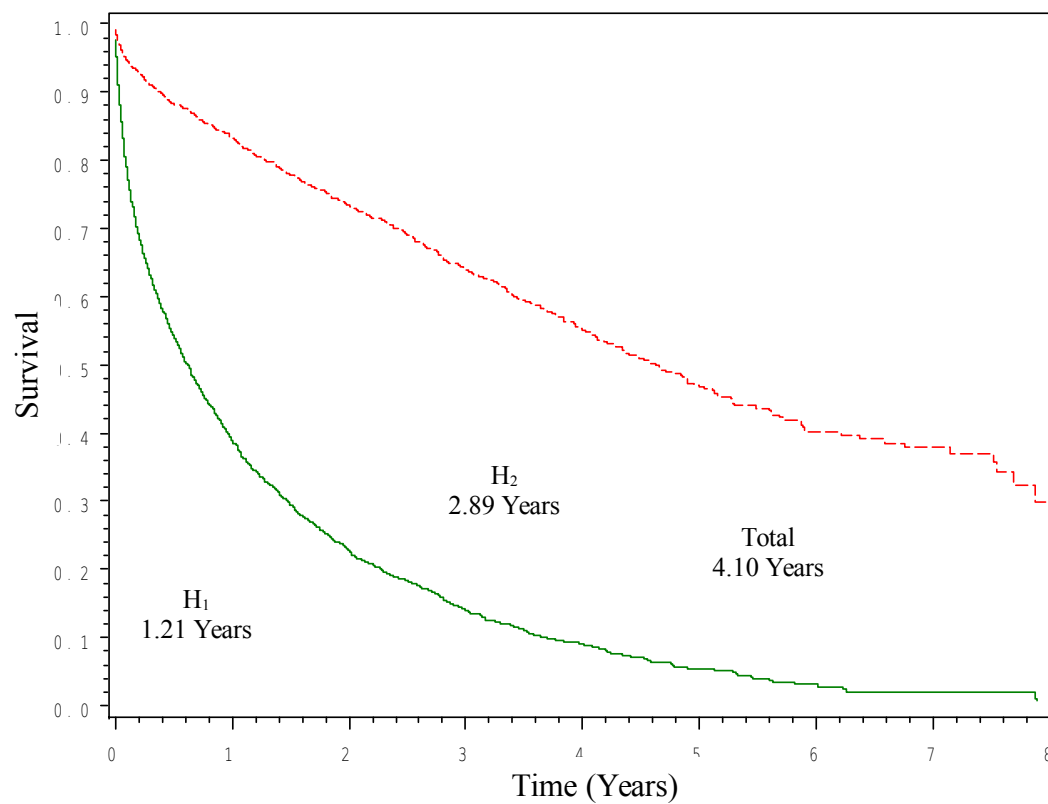
**Competing Interest Statement**

All authors declare that they have no competing interests and therefore have none to declare.

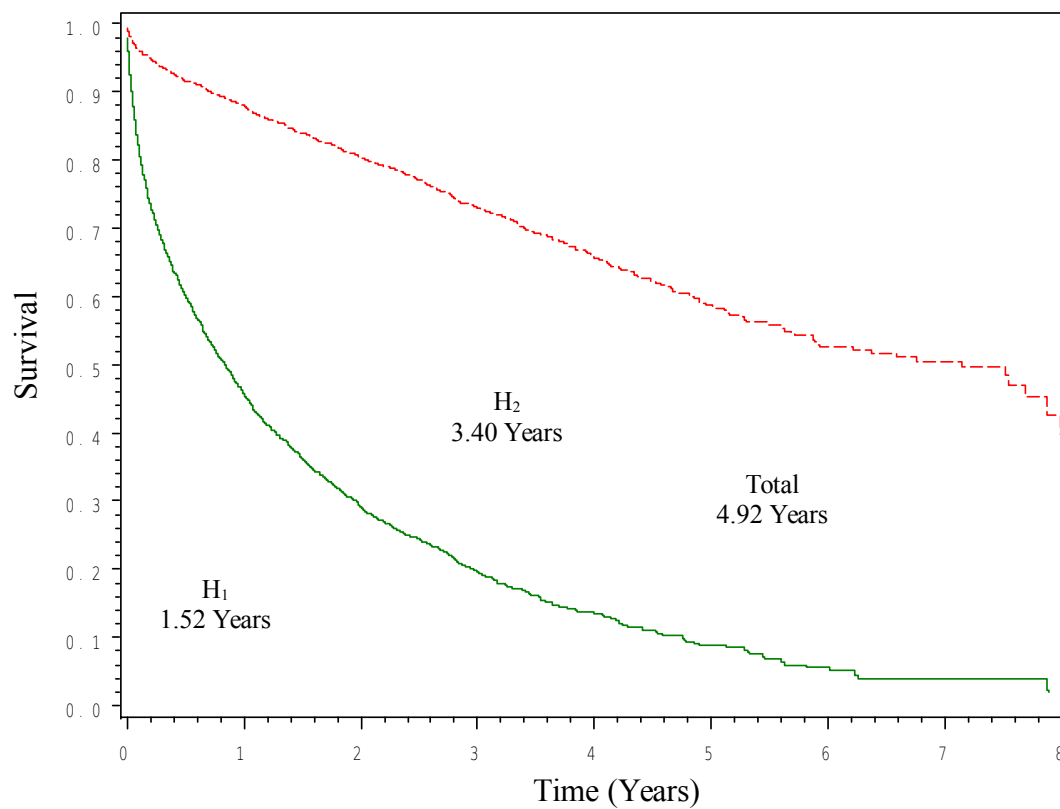


**Figure 1. Partitioned Survival Functions for All-Cause Mortality and All-Cause Hospitalization in Observational Study of Antidiabetic Agents in Patients with Heart Failure.**

**A. Sulfonylurea Therapy**

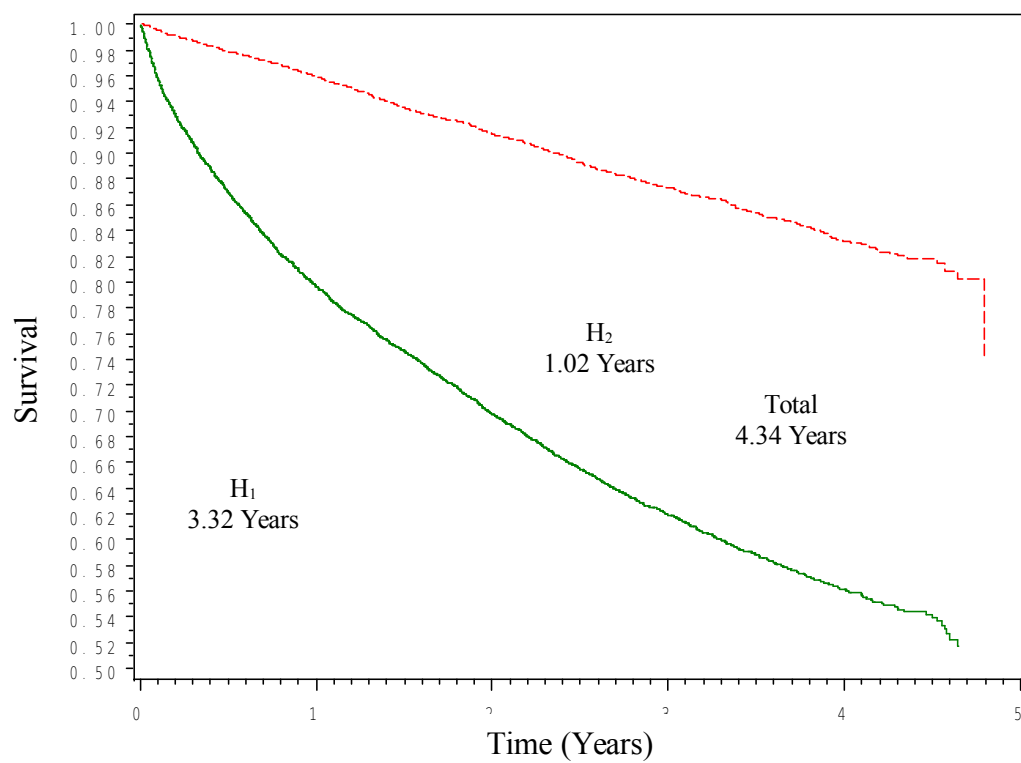


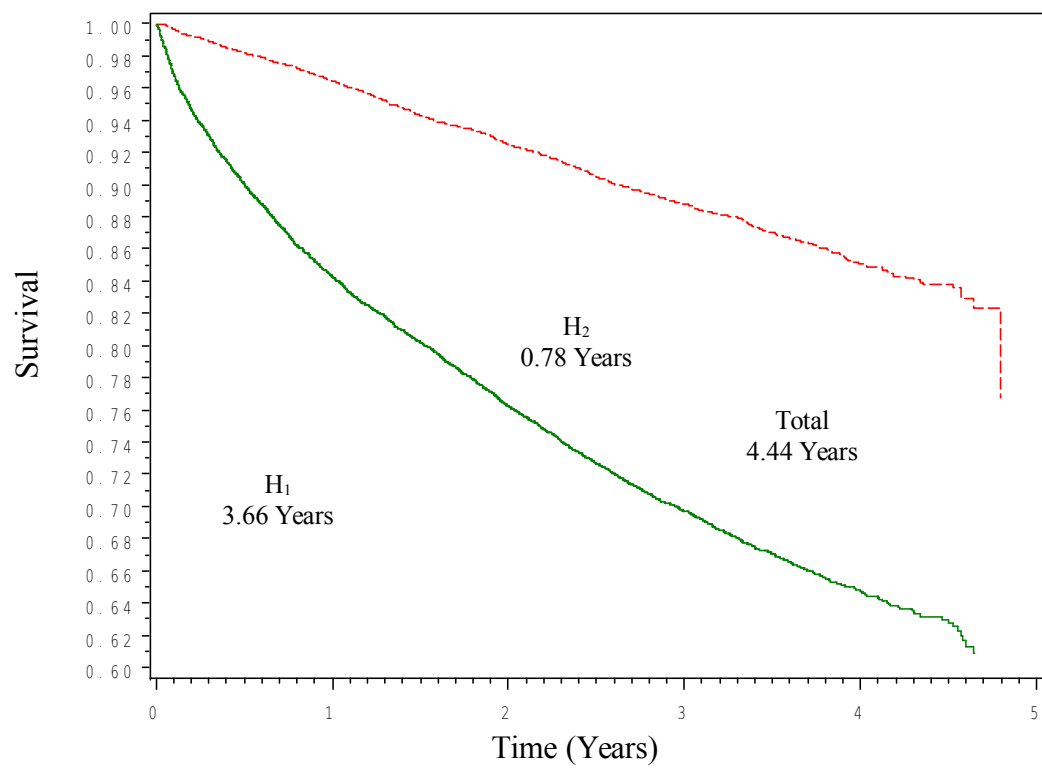
**B. Metformin Therapy**



**Figure 2. Partitioned Survival Functions for Death due to Worsening Heart Failure or Hospitalization Related to that Diagnosis in a Randomized Trial of Digoxin.**

**A. Placebo Group**



**B. Digoxin Group**

**Legend:**

Figure 1 A &amp; B

All-Cause Mortality – — — — —

All-Cause Hospitalization or Mortality- —————

Figure 2 A &amp; B

Heart Failure Related Mortality – — — — —

Heart Failure Related Hospitalization or Mortality - —————

**Table I. Integrated Survival Time for Sulfonylurea and Metformin Groups per Health State**

<b>Group</b>	<b>Survival in H<sub>1</sub></b> <b>(95% CI)</b>	<b>Survival in H<sub>2</sub></b> <b>(95% CI)</b>	<b>Total Survival</b> <b>(95% CI)</b>	<b>Quality Adjusted</b> <b>Survival (95%</b> <b>CI)</b>
<b>Sulfonylurea</b>	1.21 (1.06-1.32)	2.89 (2.65-3.17)	4.10 (3.82-4.41)	2.63 (2.45-2.82)
<b>Metformin</b>	1.52 (1.27-1.82)	3.40 (2.81-3.92)	4.92 (4.35-5.41)	3.17 (2.81-3.48)

Note- All Values in Years; 95% Confidence Intervals (CI) generated from the bootstrap procedure (500 samples)

**Table II. Integrated Survival Time for Placebo and Digoxin Groups per Health State**

<b>Group</b>	<b>Survival in H<sub>1</sub> (95% CI)</b>	<b>Survival in H<sub>2</sub> (95% CI)</b>	<b>Total Survival (95% CI)</b>	<b>Quality Adjusted Survival (95% CI)</b>
<b>Placebo</b>	3.32	1.02	4.34	3.27
<b>Digoxin</b>	(3.13-3.27) 3.66	(0.98-1.09) 0.78	(4.30-4.39) 4.44	(3.23-3.31) 3.41
	(3.44-3.56)	(0.72-0.83)	(4.38-4.48)	(3.36-3.44)




Note- All Values in Years; 95% Confidence Intervals (CI) generated from the bootstrap

procedure (500 samples)






## Figure Legends

### Figure 1 A & B

-   All-Cause Mortality
-  All Cause Hospitalization or Mortality

### Figure 2 A & B

-   HF-Related Mortality
-  HF-Related Hospitalization or Mortality

## Appendix A - QAS Method

This method is an adaptation of the Q-TWIST (quality adjusted time without symptoms and toxicity) methodology.(6) The Q-Twist method was originally designed to evaluate the impact of treatments on Health Related Quality of Life (HRQL) in breast cancer trials. The Quality Adjusted Survival (QAS) Analysis by Cole et al. was an adaptation of this methodology which extended the standard Q-Twist methodology to Cox's proportional hazards regression to allow for adjustment of covariates.(8)

To compute the QAS endpoint, health states which differ in HRQL are first identified. Once identified, each health state is assigned a corresponding utility coefficient or 'weight' representing the degree of health relative to the state of perfect health. By convention, a utility coefficient of 1 indicates perfect health and zero represents death. Negative utility coefficients are possible which represents health states worse than death. The quality adjusted survival time is computed by multiplying the time spent in each health state by its corresponding utility coefficient or 'weight'. These quality adjusted survival times for each health state are then summated to provide a measure of the total QAS.(8)

To calculate the time spend in each corresponding health state, Cox proportional hazards regression (with or without adjustment for covariates) is used to generate survival functions (i.e., survival curves) for the successive health states (transition times) originating from the same time point.(8) Since the area under the survival curve for an event of interest represents an estimate of the mean survival time for a subject being free from the event, the mean duration of a health state can be estimated.(19) The area between the survival curves for the successive health states represents the mean time spent in each health state, also known as a partitioned survival analysis. (6) Due to censoring, the entire survival curve cannot often be calculated. Thus, the area under

the survival curve is restricted to an upper limit.(8) Since the total QAS time is influenced by the range of integration, it is critical that the upper limit of integration is identical for both treatment groups to avoid introducing bias. Furthermore, as the survival curves continue to change over time, the QAS difference observed between treatment groups may potentially vary depending on the time range. The restricted limit may represent a specific time period (e.g. 5 years) or may represent the upper limit of follow-up. Therefore, the mean survival time is computed by integrating the survival curve from zero to the upper limit.(8)

To compare QAS between treatment groups, the QAS estimates are derived by including the treatment group as a covariate in the model or by stratifying on the treatment variable if the proportional hazards assumption is violated. By convention, if one of the health states is censored for the patient, all subsequent health states are also censored.(8) Importantly, the proportional hazards assumption should be evaluated for each health state.

Previous research has shown that the mean QAS time estimates are asymptotically normally distributed, under widely-applicable regularity conditions.(8;20) Resampling methods, such as the bootstrap method(13;21) are used to then generate variance estimates around the QAS estimate. It is important to realize, however, that it is neither tenable nor appropriate to compare the confidence intervals of QAS from the unweighted and weighted methods and infer their underlying QAS differences. This is because the two confidence intervals are derived from the same sample and, therefore, are not statistically independent of one another. This is analogous to comparing two confidence intervals in a paired t-test scenario, where an overlap of the confidence intervals of the two groups is irrelevant to the intended inference.

See Hinkley DV. Bootstrap Methods. *Journal of the Royal Statistical Society Series B-*

*Methodological* 1988; 50(3):321-337 or Efron B. Censored-Data and the Bootstrap. *Journal of*

*the American Statistical Association* 1981;76(374):312-319 for a detailed description of the bootstrap method.

### Manuscript Examples

#### Data Sources

##### *i. Metformin Use in Heart Failure*

The data sources and population studied were previously described in detail.<sup>(9)</sup> Briefly, between January 1, 1991 and December 31, 1999, 1,833 eligible subjects newly treated with oral antidiabetic agents and had incident heart failure were identified using the administrative databases of Saskatchewan Health. Subjects were categorized into three mutually exclusive groups according to oral antidiabetic prescription claims: 773 (42%) were treated with sulfonylurea therapy alone, 208 (11%) with metformin alone, and 852 (47%) were treated with combinations of sulfonylurea and metformin. For the illustrative purposes of this paper, only individuals who received either metformin or sulfonylurea monotherapy were included. All subjects were prospectively followed until death, termination of Saskatchewan Health coverage, or December 31, 1999, providing a maximum follow-up of 9 years. However, for illustrative purposes, the upper limit was truncated to 8 years to ensure comparability in follow-up time between treatment groups.

The mean age of the study cohort at baseline was 74 (SD 10) years, 59% were male, and mean follow-up was 2.3 (SD 1.9) years following the diagnosis of heart failure. The sulfonylurea group was slightly older, had less comorbidity, and had fewer prescription claims for heart failure-related medications compared to the metformin monotherapy group (Table 1).

Using standard Cox proportional hazards regression techniques, after adjustment for potentially confounding variables (i.e., age, sex, a modified Chronic Disease Score (CDS) (22;23), therapies known to affect heart failure outcomes: ACE inhibitors; angiotensin II blockers; beta-blockers; antiplatelet agents; nitrates; lipid-lowering therapies; antiarrhythmic agents; and spironolactone, and total physician visits prior to heart failure diagnosis), a reduction in the hazard of events in favor of the metformin group compared to sulfonylurea therapy was observed for both all-cause mortality [69 (33%) vs. 404 (52%); hazards ratio (HR) 0.70 (95% CI 0.54-0.91)] and the composite outcome of “all-cause mortality or all-cause hospitalization” [160 (77%) vs. 658 (85%); HR 0.83 (95% CI 0.70-0.99)].(9)

## *ii. The DIG Study*

The rationale, design, and results of the DIG study has been previously described in detail. (10;24) A total of 6800 patients with heart failure and a left ventricular ejection fraction  $\leq 0.45$  percent were randomly assigned to receive either digoxin or placebo. The mean age of the DIG study participants was 63 (SD 11) years, 78% were male, and mean follow-up was 37 (range 28 to 58) months following randomization. There were no significant differences between the baseline characteristics between the digoxin or placebo patients (Table 2). (10) After an average follow-up of 37 months, there was no difference between the study groups with respect to the primary outcome of all-cause mortality [1194 (35%) in placebo group versus 1181 (35%) in digoxin group; HR 0.99 (95% CI 0.91-1.07)]. There was a trend toward a lower risk of heart failure related mortality in the digoxin group compared to the placebo group [394 (12%) vs. 449 (13%); HR 0.88 (95% CI 0.77-1.01)]. In addition, the risk for the composite outcome of death due to worsening heart failure or hospitalization related to that diagnosis was lower in the

digoxin group [1041 (31%) in digoxin group vs. 1291 (38%) in placebo group; HR 0.75 (95% CI 0.69-0.82)].(10)

### Weighted Composite Outcome

We applied the QAS analysis to the all-cause mortality or hospitalization composite outcome in the metformin study and to the heart failure-specific mortality or hospitalization composite outcome in the DIG study. In these two examples, we considered three potential successive health states patients may transition through during the periods of the study: 1) state  $H_1$  was the initial health state of the patient and represents the state of health prior to a hospitalization, death or censoring at the end of follow-up; 2) state  $H_2$  was the health state of the patient after a hospitalization until either death or censoring at the end of follow-up; 3) state  $H_3$  was the final health state dead.

Each health state is associated with a different HRQL, represented by a utility coefficient.(8) Traditional survival analysis assumes that each health state before and after a hospitalization (i.e.,  $H_1$  and  $H_2$ ) is of equal importance with a corresponding utility coefficient equal to 1 (i.e., perfect health). The QAS approach, however, accounts for potential differences in health states by assigning unequal utility coefficients to the health states to reflect the differential impact on the patients' quality of life. For state  $H_1$ , the utility coefficient was set at 0.81, based on Health Utilities Index Mark 3 (HUI3) scores observed for subjects with heart disease and diabetes in the Canadian population,(11) and is similar to utility weights observed in patients with heart failure alone.(12) In patients with heart failure, a hospitalization is associated with a 30% reduction in the patients HRQL(12) resulting in a utility coefficient of 0.57 for state  $H_2$ . By convention, the health state dead ( $H_3$ ) was assigned a utility score of 0.(11)

For the metformin study, the transitional survival functions for each health state (i.e., all-cause death or all-cause hospitalization) for the metformin and sulfonylurea groups were estimated using Cox proportional hazards models (i.e., PHREG) with adjustments for the confounding covariates and the product-limit estimate of the baseline survival function.<sup>(8)</sup> The survival functions for the treatment groups were estimated at the overall mean values of the covariates. For the DIG study, due to the randomized design, the survival functions were estimated with no adjustment for confounding covariates. Proportional hazards assumption were checked for both the metformin and DIG study using log-log plots and time interactions with no violations for either study.

For each study, the Cox proportional hazards regression survival function (i.e., survival curves) for the health state dead ( $H_3$ ) as the event of interest was first estimated in each group. The survival function for the composite outcome (hospitalization or death) as the event of interest was then estimated for each treatment group. The mean time spent in each health state was calculated by integrating the estimated survival function from zero to the maximum follow-up of 8.0 years and 4.8 years for the metformin study and DIG study, respectively. Time in state  $H_1$  (initial health state of the patient and represents the state of health prior to a hospitalization, death or censoring at the end of follow-up) was estimated by the integrated survival time for the composite outcome (hospitalization or death). Time in state  $H_2$  (health state of the patient after a hospitalization until either death or censoring at the end of follow-up) was calculated by integrating the overall survival time of death and subtracting the mean time spent in state  $H_1$ . QAS time in each health state was then calculated by multiplying the mean time spent in each

health state by the respective utility coefficient.(8) Summation of these quality adjusted times provided an estimate of the overall QAS time during the study period for each treatment group.

To estimate the variability of the mean QAS time for the treatment groups, we used the bootstrap procedure. The bootstrap procedure gives an approximate sampling distribution (and, therefore, standard errors) of the statistic of interest (in our case the mean QAS) by a resampling procedure. Specifically, a sample of  $n$  observations, called the bootstrap sample, is drawn from the data of  $n$  observations (i.e., the observed metformin or DIG Study data) by resampling of the original  $n$  observations *with replacement*. Thus, the bootstrapped sample is an approximate of the original source population. Each observation in the original data has an equal probability of being selected for the bootstrap sample. The bootstrap sample contains the same total number of observations as the original sample. However, due to the resampling with replacement, some observations will appear once, more than once, or not at all in the bootstrap sample. Once the bootstrap sample has been obtained, the entire QAS method is applied to the bootstrap sample, generating an estimate of the mean QAS for the treatment group of interest in the bootstrapped data. Of note, the time spent in each health state was calculated to the end of the follow-up time observed or to the maximum of 8.0 years and 4.8 years for the metformin study and DIG study, respectively, for subjects included in each boot sample. This bootstrap sampling and QAS calculation was repeated 500 times for each of the metformin and DIG Study datasets. The variability in the 500 QAS estimates generated from the 500 bootstrap samples is an estimate the variability of the mean QAS time obtained in the original data. The 95 percent confidence interval and corresponding p-values were then calculated according to the percentile method.(13) The QAS estimates which span the 2.5 and 97.5 percentiles of the bootstrapped distribution are used as the 95 percent confidence interval.





**Table 1. Study Cohort Characteristics for Metformin Use in Heart Failure**

Characteristic	Sulfonylurea	Metformin	P Value*
	Monotherapy	Monotherapy	
	(n = 773)	(n = 208)	
	No. (%) or Mean $\pm$ SD		
Age – yrs	74.8 $\pm$ 10.1	72.5 $\pm$ 10.6	<0.001
Sex – male	451 (58)	123 (59)	0.40
Duration of Follow Up after diagnosis of heart failure (yrs)	2.3 $\pm$ 2.0	2.3 $\pm$ 1.8	<0.001
Chronic Disease Score	10.7 $\pm$ 3.7	11.6 $\pm$ 3.6	<0.001
Median	10.0	11.0	
Total Physician Visits†	41.6 $\pm$ 44.5	48.0 $\pm$ 40.0	<0.001
Myocardial Infarction	72 (9)	20 (10)	0.645
Ischemic Heart Disease	152 (16)	32 (15)	0.874
Cerebrovascular Disease	88 (11)	19 (9)	0.490
Other Diseases of Arteries, Arterioles and Capillaries	27 (4)	6 (3)	0.91
Medications‡			
Thiazide Diuretics	214 (28)	59 (11)	0.36
Loop Diuretics	595 (77)	157 (76)	0.061
ACE Inhibitors	476 (62)	148 (71)	<0.001
ARBs	38 (5)	17 (8)	0.008
Antiplatelet Therapy	300 (39)	92 (44)	0.24
Antiarrhythmic Agent	369 (48)	109 (52)	0.45
Beta Blockers	251 (33)	90 (43)	<0.001
Spironolactone	113 (15)	29 (14)	0.77
Lipid Therapy	123 (16)	49 (24)	<0.001
Nitroglycerin	357 (46)	106 (51)	0.04

\* omnibus p-values from  $\chi^2$  test or ANOVA

† total physician visits prior to HF diagnosis

**Table 2. Selected Study Characteristics for DIG Trial**

Characteristic	Placebo Group	Digoxin Group	P Value*
	(n = 3403)	(n = 3397)	
	No. (%) or Mean $\pm$ SD		
Age – yrs	63.5 $\pm$ 10.8	63.4 $\pm$ 11.0	0.63
Sex (male)	2639 (78)	2642 (78)	0.82
Nonwhite Race	504 (14.8)	487 (14.3)	0.58
Ejection Fraction (%)	28.4 $\pm$ 8.9	28.6 $\pm$ 8.9	0.38
Duration of heart failure (mo)	29.8 $\pm$ 36.5	30.5 $\pm$ 37.2	0.42
NYHA Class			
I	442 (13)	465 (14)	0.61
II	1854 (55)	1810 (53)	
III	1039 (31)	1042 (31)	
IV	66 (2)	76 (2)	
Medical history			
Previous myocardial infarction	2221 (65)	2198 (65)	0.64
Current angina	899 (26)	922 (27)	0.50
Diabetes	972 (29)	961 (28)	0.80
Hypertension	1557 (46)	1527 (45)	0.51
Primary cause of heart failure			
Ischemic	2398 (71)	2405 (71)	0.77
Nonischemic	1005 (29)	992 (29)	
Medications			
Diuretics	2797 (82)	2759 (81)	0.30
ACE Inhibitors	3225 (95)	3197 (94)	0.24
Nitrates	1466 (43)	1432 (42)	0.44
Hydralazine	64 (2)	22 (2)	0.26

\* omnibus p-values from  $\chi^2$  test or ANOVA

## Appendix B - Sensitivity Analyses According to Changes in Utility Coefficients

### A. Metformin Study

Parameter varied	Group	Utility H <sub>1</sub>	QALY H <sub>1</sub>	Utility H <sub>2</sub>	QALY H <sub>2</sub>	Total QALYs	Difference in Total QALYs	Difference from Standard Analysis (%)
<b>Base Case</b>	Sulfonylurea	0.81	0.98	0.57	1.65	2.63	0.54	34
	Metformin	0.81	1.23	0.57	1.94	3.17		
<b>H<sub>1</sub> increased 0.03; H<sub>2</sub> base</b>	Sulfonylurea	0.84	1.02	0.57	1.65	2.66	0.55	33
	Metformin	0.84	1.28	0.57	1.94	3.21		
<b>H<sub>1</sub> decreased 0.03; H<sub>2</sub> Base</b>	Sulfonylurea	0.78	0.94	0.57	1.65	2.59	0.53	33
	Metformin	0.78	1.19	0.57	1.94	3.12		
<b>H<sub>1</sub> base; H<sub>2</sub> increased 0.03</b>	Sulfonylurea	0.81	0.98	0.60	1.73	2.71	0.56	32
	Metformin	0.81	1.23	0.60	2.04	3.27		
<b>H<sub>1</sub> base; H<sub>2</sub> decreased 0.03</b>	Sulfonylurea	0.81	0.98	0.54	1.56	2.54	0.53	33
	Metformin	0.81	1.23	0.54	1.84	3.07		
<b>H<sub>1</sub> increased 0.03; H<sub>2</sub> decreased 0.03</b>	Sulfonylurea	0.84	1.02	0.54	1.56	2.58	0.54	34
	Metformin	0.84	1.28	0.54	1.84	3.11		
<b>H<sub>1</sub> decreased 0.03; H<sub>2</sub> increased 0.03</b>	Sulfonylurea	0.78	0.94	0.60	1.73	2.68	0.55	33
	Metformin	0.78	1.19	0.60	2.04	3.23		
<b>H<sub>1</sub> increased 0.03; H<sub>2</sub> increased 0.03</b>	Sulfonylurea	0.84	1.02	0.60	1.73	2.75	0.57	30
	Metformin	0.84	1.28	0.60	2.04	3.32		
<b>H<sub>1</sub> decreased 0.03; H<sub>2</sub> decreased 0.03</b>	Sulfonylurea	0.78	0.94	0.54	1.56	2.50	0.52	37
	Metformin	0.78	1.19	0.54	1.84	3.02		

**B. DIG Study**

Parameter varied	Group	Utility H <sub>1</sub>	QALY H <sub>1</sub>	Utility H <sub>2</sub>	QALY H <sub>2</sub>	Total QALYs	Difference in Total QALYs	Difference from Standard Analysis (%)
<b>Base Case</b>	Placebo	0.81	2.69	0.57	0.58	3.27	0.14	40
	Digoxin	0.81	2.96	0.57	0.44	3.41		
<b>H<sub>1</sub> increased 0.03; H<sub>2</sub> base</b>	Placebo	0.84	2.79	0.57	0.58	3.37	0.15	50
	Digoxin	0.84	3.07	0.57	0.44	3.52		
<b>H<sub>1</sub> decreased 0.03; H<sub>2</sub> Base</b>	Placebo	0.78	2.59	0.57	0.58	3.17	0.13	30
	Digoxin	0.78	2.85	0.57	0.44	3.30		
<b>H<sub>1</sub> base; H<sub>2</sub> increased 0.03</b>	Placebo	0.81	2.69	0.60	0.61	3.30	0.13	30
	Digoxin	0.81	2.96	0.60	0.47	3.43		
<b>H<sub>1</sub> base; H<sub>2</sub> decreased 0.03</b>	Placebo	0.81	2.69	0.54	0.55	3.24	0.15	50
	Digoxin	0.81	2.96	0.54	0.42	3.39		
<b>H<sub>1</sub> increased 0.03; H<sub>2</sub> decreased 0.03</b>	Placebo	0.84	2.79	0.54	0.55	3.34	0.16	60
	Digoxin	0.84	3.07	0.54	0.42	3.50		
<b>H<sub>1</sub> decreased 0.03; H<sub>2</sub> increased 0.03</b>	Placebo	0.78	2.59	0.60	0.61	3.20	0.12	20
	Digoxin	0.78	2.85	0.60	0.47	3.32		
<b>H<sub>1</sub> increased 0.03; H<sub>2</sub> increased 0.03</b>	Placebo	0.84	2.79	0.60	0.61	3.40	0.14	40
	Digoxin	0.84	3.07	0.60	0.47	3.54		
<b>H<sub>1</sub> decreased 0.03; H<sub>2</sub> decreased 0.03</b>	Placebo	0.78	2.59	0.54	0.55	3.14	0.14	40
	Digoxin	0.78	2.85	0.54	0.42	3.28		

## Reference List

- (1) O'Connor CM, Gattis WA, Ryan TJ. The role of clinical nonfatal end points in cardiovascular phase II/III clinical trials. *Am Heart J* 2000 Apr;139(4):S143-S154.
- (2) Taylor AL, Ziesche S, Yancy C, et al. Combination of isosorbide dinitrate and hydralazine in blacks with heart failure. *N Engl J Med* 2004 Nov 11;351(20):2049-57.
- (3) DeMets DL, Califf RM. Lessons Learned From Recent Cardiovascular Clinical Trials: Part I. *Circulation* 2002 Aug 6;106(6):746-51.
- (4) Neaton JD, Gray G, Zuckerman BD, et al. Key issues in end point selection for heart failure trials: composite end points. *J Card Fail* 2005 Oct;11(8):567-75.
- (5) Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007 Apr 14;334(7597):786.
- (6) Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Stat Med* 1990 Nov;9(11):1259-76.
- (7) Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals in Canada. 2nd ed. Ottawa: Canadian Coordinating Office for Health Technology Assessment (CCOHTA); 1997.
- (8) Cole BF, Gelber RD, Goldhirsch A. Cox regression models for quality adjusted survival analysis. *Stat Med* 1993 May 30;12(10):975-87.
- (9) Eurich DT, Majumdar SR, McAlister FA, et al. Improved Clinical Outcomes Associated With Metformin in Patients With Diabetes and Heart Failure. *Diabetes Care* 2005 Oct 1;28(10):2345-51.
- (10) The Digitalis Investigation Group. The effect of digoxin on mortality and morbidity in patients with heart failure. *N Engl J Med* 1997 Feb 20;336(8):525-33.
- (11) Maddigan SL, Feeny DH, Johnson JA. Health-related quality of life deficits associated with diabetes and comorbidities in a Canadian National Population Health Survey. *Qual Life Res* 2005 Jun;14(5):1311-20.
- (12) Nichol G, Kaul P, Huszti E, et al. Cost-effectiveness of cardiac resynchronization therapy in patients with symptomatic heart failure. *Ann Intern Med* 2004 Sep 7;141(5):343-51.
- (13) Efron B. Censored-Data and the Bootstrap. *Journal of the American Statistical Association* 1981;76(374):312-9.
- (14) Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI(R)): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003 Oct 16;1(1):54.

- (15) Boyko WL, Jr., Glick HA, Schulman KA. Economics and cost-effectiveness in evaluating the value of cardiovascular therapies. ACE inhibitors in the management of congestive heart failure: comparative economic data. *Am Heart J* 1999 May;137(5):S115-S119.
- (16) Cole BF, Gelber RD, Anderson KM. Parametric approaches to quality-adjusted survival analysis. International Breast Cancer Study Group. *Biometrics* 1994 Sep;50(3):621-31.
- (17) Southern DA, Faris PD, Brant R, et al. Kaplan-Meier methods yielded misleading results in competing risk scenarios. *J Clin Epidemiol* 2006 Oct;59(10):1110-4.
- (18) Eurich DT, Johnson JA, Reid KJ, et al. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. *Health Qual Life Outcomes* 2006;4:89.
- (19) Kaplan EL, Meier P. Nonparametric estimation from icomplet observations. *Journal of the American Statistical Association* 1958;54:457-81.
- (20) Tsiatis AA. Estimating the distribution of quality-adjusted life with censored data. *Am Heart J* 2000 Apr;139(4):S177-S181.
- (21) Hinkley DV. Bootstrap Methods. *Journal of the Royal Statistical Society Series B-Methodological* 1988;50(3):321-37.
- (22) Eurich DT, Majumdar SR, Tsuyuki RT, et al. Reduced mortality associated with the use of ACE inhibitors in patients with type 2 diabetes. *Diabetes Care* 2004 Jun;27(6):1330-4.
- (23) Johnson JA, Majumdar SR, Simpson SH, et al. Decreased mortality associated with metformin use compared to sulfonylurea monotherapy in type 2 diabetes mellitus. *Diabetes Care* 2002;25:2244-8.
- (24) The Digitalis Investigation Group. Rationale, design, implementation, and baseline characteristics of patients in the DIG trial: a large, simple, long-term trial to evaluate the effect of digitalis on mortality in heart failure. *Control Clin Trials* 1996 Feb;17(1):77-97.