UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DIPARTIMENTO DI SCIENZE CLINICHE E SPERIMENTALI

DOTTORATO DI RICERCA IN INTELLIGENZA ARTIFICIALE IN MEDICINA E INNOVAZIONE NELLA RICERCA CLINICA E METODOLOGICA

settore scientifico disciplinare MEDS-24/A Statistica medica

XXXVII CICLO

Patient-Reported Outcomes in cardiovascular trials for the treatment of Heart Failure: alternative approaches for quantitative synthesis and joint evaluation with survival.

CANDIDATE
Chiara Oriecuia
Matr. N. 738877

SUPERVISOR
Prof. Marco Metra
Dipartimento di Specialità Medico-Chirurgiche,
Scienze Radiologiche e Sanità Pubblica,
Università degli Studi di Brescia

CO-SUPERVISOR
Prof.ssa Claudia Specchia
Dipartimento di Medicina Molecolare e Traslazionale,
Università degli Studi di Brescia

Abstract

Patient-Reported Outcomes (PROs) have become essential in clinical research, directly capturing patients' perspectives on symptoms, quality of life (QoL), and health-related aspects to enhance understanding of treatment impacts and support informed decision-making. In heart failure (HF) research, the Kansas City Cardiomyopathy Questionnaire (KCCQ) has emerged as a preferred, FDA-approved instrument for assessing QoL due to its sensitivity and specificity. However, PROs introduce considerable methodological challenges: differences in questionnaire formats, scoring scales, and follow-up intervals make it difficult to align data across studies without potential biases or information loss, complicating direct comparisons and data synthesis. Additionally, in clinical trials, the interplay between QoL and survival presents unique complexities. Since PROs are only collected from surviving patients, early mortality among those with lower QoL can lead to an overestimation of average QoL, distorting perceived treatment effects. Furthermore, survival models that rely solely on baseline PRO values may fail to capture the dynamic nature of patient health, potentially missing crucial changes in QoL that might correlate with prognosis.

The overall aim of this doctoral thesis was to address the complexities introduced by PROs in both meta-analyses and survival analysis of clinical studies.

The first objective was to propose a meta-analytic methodology capable of comparing continuous PROs between treatment groups at predefined timepoints while minimizing information loss. As a motivating example, it was applied to PRO data from published randomized controlled trials (RCTs) comparing SGLT2 inhibitors (SGLT2i) with standard of care (SOC) in patient with HF. The effect size of interest was the difference in mean change of the KCCQ overall score (KCCQ-OSS) between the SGLT2i and SOC groups at 3 and 6 months from baseline. Two complementary methods were applied: a traditional aggregated data (AD) approach with specific enhancements for

handling missing data and a meta-analytical approach based on the reconstructed pseudo individual patient data (IPD). The synthesis was then conducted using two distinct meta-analytical approaches: a two stage approach, where study-specific estimates were first obtained from the pseudo-IPD separately for each study and then combined by a traditional meta-analytical model; and a one stage approach based on a linear mixed-effects model (LMM), which provided pooled estimates of treatment differences at predefined timepoints and explored interactions between time and treatment effects. Results demonstrated that pseudo-IPD reconstruction and the LMM improve the accuracy and precision of pooled estimates in longitudinal QoL data, enabling a clearer understanding of treatment effects over time that would be missed with AD. Using methods to minimize information loss, including graphical extraction or linear interpolation at specified timepoints, was essential for achieving standardized results across studies with varying follow-up intervals. These refinements could assist healthcare providers in gaining clearer insights into both the magnitude and timing of QoL improvements, thereby supporting more informed clinical decisions.

The second objective was to evaluate statistical models for analyzing survival and QoL outcomes, addressing specific challenges posed by the MITRADVANCE-HF trial, an RCT assessing the effectiveness of percutaneous mitral valve repair in patients with advanced HF. Within this framework, joint models were recently proposed to capture the dependency structure between longitudinal predictors and survival outcome, linking a mixed-effect model for repeated measures and a survival model for event-time data. As data collection for MITRADVANCE-HF trial is still ongoing, simulated survival and QoL data were used to explore several possible scenarios that may emerge from the RCT.

To assess the robustness and statistical proprieties of joint models for analyzing these data, comparisons were made with traditional methods, such as the Cox model with time-dependent covariates, typically used to estimate

the effect of variables collected during follow-up. The joint model demonstrated superior accuracy and robustness in estimating treatment effects and in capturing the protective influence of QoL on survival, whereas simpler models exhibited substantial bias or failed to accurately represent these relationships.

Together, the findings from the two applications highlight the value of integrating advanced analytical techniques for PRO data across different study designs. These methods not only strengthen evidence-based conclusions but also promote a more comprehensive understanding of treatment impacts on both patient QoL and survival, ultimately supporting a more holistic approach to patient care.

Abstract (italiano)

I Patient-Reported Outcomes (PRO), ovvero gli esiti riferiti direttamente dai pazienti, sono diventati indispensabili nella ricerca clinica, poiché permettono di cogliere la prospettiva dei pazienti su sintomi, qualità della vita (QoL) e altri aspetti legati alla salute. Questo approccio consente di comprendere meglio gli effetti dei trattamenti e supporta decisioni più informate. Nell'ambito dello scompenso cardiaco (HF), il *Kansas City Cardiomyopathy Questionnaire* (KCCQ) si è affermato come strumento di riferimento grazie alla sua sensibilità e specificità, ricevendo anche l'approvazione della FDA. Tuttavia, i PRO comportano notevoli sfide metodologiche: le differenze nei formati dei questionari, nelle scale di punteggio e negli intervalli di follow-up rendono difficile confrontare i dati tra studi diversi senza introdurre bias o perdita di informazioni, complicando il confronto diretto e la sintesi dei dati. Inoltre, nei trial clinici, l'interazione tra QoL e sopravvivenza presenta complessità particolari. Poiché i PRO vengono raccolti solo dai pazienti sopravvissuti, la mortalità precoce tra coloro con QoL più bassa può portare a una sovrastima della QoL media, distorcendo così la percezione degli effetti del trattamento. Inoltre, i modelli di sopravvivenza basati unicamente sui valori iniziali dei PRO possono non cogliere la natura dinamica dello stato di salute dei pazienti, rischiando di perdere cambiamenti importanti nella QoL che potrebbero essere correlati alla prognosi.
L'obiettivo generale di questa tesi di dottorato era quello di affrontare le complessità introdotte dai PRO sia nelle metanalisi che nell'analisi di sopravvivenza degli studi clinici.

Il primo obiettivo era proporre una metodologia meta-analitica in grado di confrontare i PRO (continui) tra gruppi di trattamento in momenti predefiniti, riducendo al minimo la perdita di informazioni. A titolo esemplificativo è stata applicata ai dati provenienti da studi clinici randomizzati (RCT) pubblicati che confrontavano gli inibitori SGLT2 (SGLT2i) con lo standard di cura (SOC) nei pazienti con HF. La misura d'interesse era la differenza nel cambiamento

medio del punteggio KCCQ complessivo (KCCQ-OSS) tra i gruppi SGLT2i e SOC a 3 e 6 mesi dal basale. Sono stati applicati due metodi complementari: un approccio tradizionale su dati aggregati con miglioramenti specifici per gestire i dati mancanti e un approccio basato sui dati pseudo-individuali ricostruiti (pseudo-IPD). Per quest'ultimo la sintesi è stata condotta utilizzando due approcci meta-analitici distinti: uno a due stadi, dove le stime specifiche per ciascun studio sono state prima ottenute dagli pseudo-IPD separatamente e poi combinate in un modello meta-analitico tradizionale; e un approccio a uno stadio basato su un modello lineare a effetti misti (LMM), che ha fornito le stime pooled delle differenze fra trattamenti in momenti predefiniti esplorando anche le interazioni tra gli effetti temporali e quelli del trattamento. I risultati hanno dimostrato che la ricostruzione degli pseudo-IPD e il modello LMM migliorano la precisione delle stime pooled nei dati longitudinali di QoL, consentendo una visione più chiara e completa degli effetti del trattamento nel tempo, che non sarebbe stata possibile usando solo i dati aggregati. L'uso di metodi per minimizzare la perdita di informazioni, tra cui l'estrazione grafica o l'interpolazione lineare in momenti specifici, è stato essenziale per ottenere risultati standardizzati tra studi con diversi intervalli di follow-up. Questi affinamenti possono aiutare i medici a comprendere meglio sia l'entità dei miglioramenti nella QoL sia i tempi in cui si verificano, supportando così decisioni cliniche più informate.

Il secondo obiettivo era valutare modelli statistici per analizzare congiuntamente la sopravvivenza e la QoL, affrontando le specifiche sfide poste dal trial MITRADVANCE-HF, un RCT che valuta l'efficacia della riparazione percutanea della valvola mitrale in pazienti con HF avanzato. In questo contesto, sono stati recentemente proposti modelli congiunti per catturare la relazione tra predittori longitudinali e l'esito di sopravvivenza, collegando un LMM per misure ripetute e un modello di sopravvivenza per i dati relativi ai tempi degli eventi. Poiché la raccolta dei dati per il trial MITRADVANCE-HF è ancora in corso, sono stati utilizzati dati simulati di sopravvivenza e QoL per esplorare vari possibili scenari. Per valutare la

robustezza e le proprietà statistiche dei modelli congiunti nell'analisi di questi dati, sono stati confrontati con metodi tradizionali, come il modello di Cox con covariate tempo-dipendenti, utilizzato per stimare l'effetto delle variabili raccolte durante il follow-up. Il modello congiunto ha dimostrato maggiore precisione e robustezza nella stima degli effetti del trattamento e nel rilevare l'effetto protettivo della QoL sulla sopravvivenza, mentre i modelli più semplici hanno mostrato rilevanti bias o non sono riusciti a rappresentare accuratamente queste relazioni.

Nel complesso, i risultati delle due applicazioni evidenziano l'importanza di integrare tecniche analitiche avanzate per l'analisi dei PRO in vari disegni di studio. Questi metodi non solo rafforzano le conclusioni basate sull'evidenza, ma favoriscono anche una comprensione più completa degli effetti dei trattamenti sulla QoL dei pazienti e sulla sopravvivenza, promuovendo così un approccio più olistico alla cura del paziente.

Index

1 Introduction

1.1 Patient's Reported Outcomes

Patient-Reported Outcomes (PROs) represent a significant advancement in healthcare research, offering direct insights into patients' perspectives on their symptoms, quality of life (QoL), and various health-related aspects. Unlike traditional clinical assessments, PROs are data reported by patients themselves, thereby eliminating interpretation biases that may arise from healthcare provider assessments.

These outcomes are typically collected through structured questionnaires or direct interviews, providing a comprehensive understanding of the patient experience during treatment. They capture diverse dimensions such as symptom severity, functional status, autonomy in daily activities, mental health, and overall perception of life quality. By utilizing metrics that assign values to different patient responses, PROs enable the generation of composite scores, allowing for quantitative analysis.[1]

PROs have gained prominence, particularly in contemporary randomized clinical trials (RCTs), where they are increasingly used to complement traditional efficacy indicators. Within RCTs, PROs serve to enhance the understanding of treatment effectiveness by incorporating patients' subjective experiences alongside objective clinical measures, contributing to the balance between positive therapeutic effects and adverse effects. The integration of PROs in RCTs represents a paradigm shift, empowering patients by providing them with a central role in decision-making processes regarding their treatment.[1]

There are numerous questionnaires available for collecting PROs, ranging from generic ones like the EuroQoL-5 Dimension (EQ-5D)[2] to those specific to particular pathologies.

1.2 PROs in Heart Failure

Heart failure (HF) stands as a significant health concern, notable for its prevalence, impact on morbidity and mortality, and extensive utilization of healthcare services. Currently affecting approximately 2 to 3% of the population, HF's prevalence escalates with advancing age, impacting as many as 10 to 20% of individuals aged over 65. Notably, in developed nations, HF's prevalence is on the rise, fueled by population aging, prolonged patient survival, and the effectiveness of secondary prevention efforts. Projections suggest a staggering 46% increase in HF prevalence from 2012 to 2030. In essence, HF is a widespread condition with profound implications for patient prognosis and lifestyle, presenting a mounting challenge for healthcare policymakers.[3,4] Therefore, there is a pressing need to comprehensively study its effects, including through the utilization of PROs, which are increasingly acknowledged as clinically meaningful endpoints.[5]

In HF, the most commonly used questionnaires are the Kansas City Cardiomyopathy Questionnaire (KCCQ)[6] and the Minnesota Living with Heart Failure Questionnaire (MLHFQ)[4], as they have been shown to have greater sensitivity in capturing QoL compared to generic tools such as the EQ-5D or the 36-item Short Form Health Survey (SF-36)[7].

Given the complexity of HF as a clinical syndrome and its diverse physiological effects, it's understandable that generic instruments may struggle to adequately capture QoL issues in these patients. The EQ-5D and SF-36 primarily focus on general health[2], while the MLHFQ and KCCQ specifically address how HF impacts patients' daily activities, providing a more nuanced understanding of their experience. For instance, the MLHFQ explores symptoms such as shortness of breath (a common and severe symptom of HF) while the KCCQ goes even further by posing three questions related to shortness of breath to capture different aspects, including severity, frequency, and impact on sleep. Moreover, while the EQ-5D assesses patients' current health status, the MLHFQ and KCCQ prompt reflection on the past four or two

weeks, respectively, incorporating events that, for example, lead to their hospitalization.[4,6,8] This approach offers a more comprehensive assessment of the impact of HF.

MLHFQ

The MLHFQ, designed in 1984 by Thomas S. Rector and Jay N. Cohn, consists of 21 items, each scored from 0 to 5, covering various physical, emotional, and socioeconomic aspects of how HF can adversely affect a patient's life during the past month. Specifically, the 21 questions assess symptoms like shortness of breath, fatigue, and peripheral edema. It also explores HF's impact on everyday physical and social activities, including mobility, household chores, sleep, work, social interactions, recreational pursuits, sexual activities, and diet. Additionally, it includes questions about cognitive function, emotional well-being (with a particular focus on depression), and also the financial burden of HF treatment.[9] The questionnaire is scored simply by summing up all 21 responses. As a result, MLHFQ scores range from 0 to 105, with higher scores indicating a greater adverse impact of HF on the respondent's life.[4] Studies have shown a direct relationship between escalating MLHFQ scores and the deterioration of HF symptoms, as corroborated by worsening New York Heart Association (NYHA) classifications.[9]

KCCQ

The KCCQ, developed in 1996 and published in 2000[10], employs a 2-week recall period to account for the day-to-day variability in HF symptoms. Comprising 23 items across seven domains, including symptom frequency, burden, stability, physical and social limitations, quality of life (KCCQ-QoL), and self-efficacy, the KCCQ provides a comprehensive assessment of the patient's health status. Symptoms frequency and burden can be combined to create the total symptom score (KCCQ-TSS), which, together with physical impairment, forms the clinical summary score (KCCQ-CSS). The symptoms, physical limitations, social limitations, and QoL domains can also be combined

to generate an overall summary score (KCCQ-OSS).[11] Additionally, to enhance the feasibility of implementing the KCCQ in routine clinical care, a shorter 12-item version, known as the KCCQ-12, was developed. This abbreviated version also covers symptom frequency, physical and social limitations, and QoL domains, offering excellent concordance with the scores of the full 23-item instrument.[11] All KCCQ scores are depicted on a 0-to-100-point scale, where lower scores denote more severe symptoms and/or limitations, while scores of 100 signify no symptoms, no limitations, and excellent QoL. The scoring process involves assigning an ordinal value to each response, and summing items within each domain. Missing values are replaced with the average of answered items within the same domain. Scale scores are then transformed to a 0 to 100 range through standardization.[10] The psychometric properties of the KCCQ, including its validity, reliability, and responsiveness, have been demonstrated across various HF etiologies, including HF with preserved ejection fraction (HFpEF) and HF with reduced ejection fraction (HFrEF).[11] Studies[10,12] have demonstrated that the KCCQ, including both the KCCQ-23 and KCCQ-12, is a reliable tool for assessing clinical changes in HF patients, surpassing the accuracy of the NYHA classification and the 6-minute walking test (6-MWT).

In recent years, the KCCQ has emerged as the preferred choice in clinical trials for patients with HF, following the U.S. Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) approval in April 2020.[5] The psychometric properties of the KCCQ are well-established in HF, leading the FDA to qualify it as a clinical outcome assessment in this population.[13] Compared to the MLHFQ, the KCCQ demonstrates substantially greater sensitivity[8] and prognostic value for outcomes such as death, transplant, left ventricular assist device utilization, and hospitalization, particularly in patients with HFrEF and, to a lesser extent, in those with HFpEF[14].

The case report form of the questionnaire, including the various questions administered to patients, is provided in the appendix for reference.

1.3 Methodological challenges introduced by PROs

The increasing use of PROs has introduced new methodological challenges in both quantitative synthesis and data analysis.

In terms of quantitative synthesis, several complexities arise from the following:

1. The questionnaires used to collect PROs are numerous and employ different scales: sometimes the score can range from 0 to 100, other times from 0 to 105. Moreover, interpretations may vary, with higher scores indicating improved health status in some instances and the opposite in others.[4,6]

2. PROs are longitudinal measures, collected repeatedly throughout a study to monitor patients' QoL over time. Due to the absence of precise guidelines, researchers autonomously determine measurement intervals. Therefore, for example, study A may collect PROs every 3 months, study B every 4 months, and study C at 1, 3, and 8 months. Meta-analyses in the literature[15–18] often overlook this variability by considering only the final available PRO measurement, thereby treating studies with disparate follow-up durations equally. Alternatively, they may perform a quantitative synthesis of the available PROs at a specific time, excluding studies that collected them at different times. Both approaches lead to significant information loss and potentially biased or incomplete synthesis.

3. Sometimes PROs are only reported graphically. When this happens, they are generally not included in the analysis, encountering the same limitations described in the previous point.

Another distinct challenge arises in clinical studies, both in RCTs and non-randomized studies, where PROs are often analyzed separately from survival outcomes. Since PROs can only be collected for patients who are still alive at the time the questionnaire is administered, this approach introduces bias in estimating the treatment effect.[19]

2 Aims and motivation examples

2.1 First aim: example of a meta-analysis on SGLT2i

The first objective was to implement various meta-analytical approaches to address some of the challenges associated with PROs, aiming to achieve a quantitative synthesis that minimizes information loss and provides additional insights beyond traditional meta-analyses (i.e. expected improvement in QoL at a given time, interaction between time and treatment effects).

To illustrate the application of these methods, a systematic meta-analysis of RCTs was conducted to evaluate the impact of initiating sodium-glucose co-transporter-2 inhibitors (SGLT2i) in addition to the standard of care (SOC) on QoL (as assessed using the KCCQ) among patients with HF, regardless of left ventricular ejection fraction (LVEF) and clinical setting.

SGLT2i, also known as gliflozins, inhibit glucose and sodium reabsorption by binding to the relative carrier protein in the proximal renal tubule. Initially developed as anti-diabetic drugs, SGLT2i have consistently shown benefits on cardiovascular outcomes in patients with HF regardless of comorbid type 2 diabetes.[20] SGLT2i are currently recommended for HF treatment, with rising evidence supporting their use across the entire LVEF spectrum.[21–23] While benefits in reducing HF hospitalization, as well as cardiovascular and all-cause death, are well established[24], their impact on short-term and long-term QoL has been less systematically evaluated through meta-analysis, leaving a gap in the current understanding. Despite consistent positive findings in RCTs, which vary in design, primary endpoints, and patient populations, a quantitative pooled estimate of this effect is lacking.

2.2 Second aim: example of joint modeling of survival and QoL data

The second objective was to explore statistical models for the joint analysis of survival outcomes and QoL. This line of research emerged from a practical challenge: how to best analyze data from the MITRADVANCE-HF trial, a multicenter clinical study designed to evaluate the percutaneous treatment of mitral regurgitation in patients with advanced HF. Given the trial's dual focus on survival and QoL, it became essential to investigate appropriate statistical methods to handle the inherent complexities of these combined endpoints.

MITRADVANCE-HF (NCT05292716) is a multicenter, prospective, randomized, controlled clinical trial that aims to assess the effectiveness of percutaneous mitral valve repair with MitraClip in patients with advanced HF and secondary mitral regurgitation (SMR)[25]. All participants receive guideline-directed optimal medical therapy (OMT) for HF.[21] Patients are randomly assigned in a 1:1 ratio to either the intervention group, which receives MitraClip in addition to OMT, or the control group, which receives OMT alone. The primary objective of the trial is to determine whether the addition of MitraClip to OMT leads to an early improvement in health status, measured by the absolute change in KCCQ-OSS at 3 months from baseline.

Secondary objectives include evaluating the persistence of these benefits over a one-year period, as well as changes in survival, HF hospitalization, functional status, symptom burden, echocardiographic parameters, and optimization of medical therapies.

The trial plans to enroll a total of 172 patients, with 86 in each arm, and approximately 20 centers across Italy are involved. Follow-up visits occur at baseline and at 3, 6, and 12 months post-randomization in both arms, with a total follow-up period of 2 years.

The study is currently in the recruitment phase and has not yet gathered sufficient data for the analysis. Therefore, simulated joint survival and QoL

data were created, exploring various scenarios, to test the statistical methods that will be applied once real data from the MITRADVANCE-HF trial become available. The study is expected to complete data collection, including the final follow-up for all patients, by April 2026.

3 Meta-analysis on SGLT2i

The methodologies and results presented in this section are adapted and expanded from the published work of Oriecuia et al. (2023)[26]. Additional content has been included here to further elaborate on the methodology and findings.

3.1 Meta-analysis: concepts and methods

Meta-analysis refers to the quantitative process of synthesizing data from multiple individual studies to obtain an overall estimate of the effect size.[27,28] It combines the results from these studies by calculating a weighted average of their effect estimates, with the weight of each study generally being the inverse of its variance. This method increases the effective sample size, leading to improved statistical power and more precise estimates than those from individual studies alone.

When there is no significant statistical heterogeneity between studies, a fixed-effects model can be fitted. This model assumes that the true effect is consistent across all studies (they all come from the same population), and any differences observed are due solely to random sampling error. However, since studies often involve different methodologies, and clinical contexts, assuming they all estimate the same effect is rarely appropriate. In such cases, a random-effects model is preferred. This model accounts for both within-study variability (due to random error) and between-study variability (caused by differences in populations, interventions, and study designs). This approach assumes that each study may be estimating a different, yet related, true effect. By incorporating this variability, the random-effects model provides a more flexible and realistic estimate of the overall effect, though it also results in greater standard errors and wider confidence intervals, reflecting the increased uncertainty introduced by differences between studies.

In a fixed-effects model, the weights for each study i are determined by the inverse of the squared standard error of the effect size $SE(\theta_i)$, calculated as:

$$\omega_i = \frac{1}{SE(\theta_i)^2} \qquad i = 1, \ldots . k$$

The pooled estimate $\widehat{\theta_{FE}}$ is then calculated as:

$$\widehat{\theta_{FE}} = \frac{\sum_{i=1}^{k} \omega_i \, \theta_i}{\sum_{i=1}^{k} \omega_i}$$

with θ_i being the parameter of interest in each study (i = 1, … , k).

In cases where between-study variability exists, a random-effects model adjusts the weights to account for this heterogeneity by introducing an additional variance component, denoted as τ^2, which represents the variance between studies. The adjusted weights are given by:

$$\omega_i' = \frac{1}{SE(\theta_i)^2 + \tau^2} \qquad i = 1, \ldots . k$$

τ^2 is calculated using the Q-statistic, which tests whether the variability in effect sizes across studies is due to chance alone and it is calculated as:

$$Q = \sum_{i=1}^{k} \omega_i \, (\theta_i - \widehat{\theta_{FE}})^2$$

where $\widehat{\theta_{FE}}$ is the pooled effect estimate from a fixed-effects model, and ω_i are the fixed-effects weights. The Q-statistic follows a chi-squared distribution with k−1 degrees of freedom. If Q>k−1, this indicates the presence of heterogeneity, and τ^2 can be computed as follows (using the DerSimonian and Laird method)[29]:

$$\tau^2 = \frac{Q - (k-1)}{\sum_{i=1}^{k} \omega_i - \frac{\sum_{i=1}^{k} \omega_i{}^2}{\sum_{i=1}^{k} \omega_i}}$$

Once τ^2 and ω_i' are calculated, the overall pooled estimate $\widehat{\theta_{RE}}$ for the random effect model is given by:

$$\widehat{\theta_{RE}} = \frac{\sum_{i=1}^{k} \omega_i' \theta_i}{\sum_{i=1}^{k} \omega_i'}$$

Note that when τ^2 = 0, weights reduce to those given in a fixed-effects model (ω_i' = ω_i).

The 100(1 − α)% confidence interval (CI) for the overall estimate $\widehat{\theta_{RE}}$ is given by the equation:

$$\left[\widehat{\theta_{RE}} - z_{1-\frac{\alpha}{2}} * SE\left(\widehat{\theta_{RE}}\right);\ \widehat{\theta_{RE}} + z_{1-\frac{\alpha}{2}} * SE\left(\widehat{\theta_{RE}}\right)\right]$$

where $SE\left(\widehat{\theta_{RE}}\right)$ is the standard error of the pooled estimate under the random-effects model, calculated as:

$$SE\left(\widehat{\theta_{RE}}\right) = \frac{1}{\sqrt{\sum_{i=1}^{k} \omega_i'}}$$

Similarly, in the fixed-effects model, the CI for the overall estimate is constructed using the same formula, but with the standard error $SE\left(\widehat{\theta_{FE}}\right)$ calculated with ω_i.

Heterogeneity between studies can be quantified using the Q statistic, as described above, but also the I^2 statistic, which quantify the proportion of total variation across studies that is due to heterogeneity rather than chance and is calculated as:

$$I^2 = 100 * \frac{Q - (k-1)}{Q}$$

I^2 values closer to 0% suggest minimal heterogeneity, while higher values indicate substantial variability between studies. When I^2 is high, a random-effects model is more appropriate, as it accounts for both within-study and between-study variability, providing a more reliable pooled estimate.

3.2 Search strategy and inclusion criteria

This meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines[30] and the recommendations of the Setting International Standards in Analyzing Patients-Reported Outcomes and Quality of Life (SISAQOL) Endpoints Data Consortium[31].

During the literature search, two distinct strategies were followed:

1) one on ClinicalTrials.gov followed by semi-automatic data screening using SAS software to identify relevant trials completed or ongoing;
2) one on PubMed followed by manual screening to identify published articles.

1) On ClinicalTrials.gov, keywords related to the clinical condition of interest, such as 'heart failure', 'cardiac insufficiency', and 'cardiomyopathy' were entered to generate an initial list of trials. Additional filters were applied to include only interventional studies conducted on adult or elderly populations. For each trial, information including the ClinicalTrials.gov identifier (NCT number), trial title, status, clinical condition investigated, interventions, outcome measures, study design, phase, and number of enrolled patients were extracted.

   Subsequently, a semi-automatic screening process was conducted using SAS software (version 9.3). Keywords were searched within the exported dataset to identify relevant trials. First, those still recruiting patients or canceled, as well as phase 1 trials, were excluded. During the subsequent phase of eliminating non-randomized trials, manual intervention was required due to inconsistencies in the trial descriptions (i.e. Allocation=Randomized, Model=Single group assignment). Regarding interventions, trials comparing drugs with

procedures, genetic treatments, devices, or behavioral interventions were eliminated, retaining only those comparing drugs with other drugs or placebo. Among the drugs, only trials related to SGLT2i were retained. Finally, a filter was applied based on outcome measures, selecting those reporting key terms related to QoL and the designated questionnaire. Manual screening was again conducted during this phase to ensure the inclusion of all the relevant trials.

Once the list of trials meeting the inclusion criteria was obtained, corresponding publications, for each trial, were retrieved.

2) The search strategy for PubMed was developed collaboratively with cardiologists, incorporating relevant keywords to define the clinical condition of HF (e.g., 'heart failure,' 'cardiac insufficiency,' and 'cardiomyopathy'), along with terms identifying SGLT2i (e.g., 'Sodium glucose co-transporter 2', 'SGLT2 inhibitor', 'SGLT2i', 'gliflozins', 'dapagliflozin', 'empagliflozin', 'canagliflozin', and 'sotagliflozin'), PROs and QoL endpoints (e.g., 'quality of life,' 'QoL', 'patient reported outcome,' 'PRO', and 'Kansas City Cardiomyopathy Questionnaire', 'KCCQ'). Additional filters were applied to select RCTs or clinical trials.

PubMed and ClinicalTrials.gov were systematically searched from their inception up to March 15, 2023.

Comprehensiveness was ensured by employing a multidisciplinary approach, thoroughly reviewing references from included papers, and examining other systematic reviews in the literature.

Studies were considered eligible for inclusion if they were RCTs meeting the following criteria:

(i) inclusion of patients affected by HF (both stable chronic and acute or worsening HF);

(ii) comparison of SGLT2i with a matching placebo added to SOC treatment of HF;
(iii) use of the KCCQ-OSS to assess QoL;
(iv) a minimum follow-up of 3 months.

KCCQ-OSS was selected as primary endpoint because, unlike other domains (i.e. KCCQ-TSS, KCCQ-CSS), which only assess symptom aspects, it incorporates the KCCQ-QoL subdomain, which independently evaluates QoL, a crucial aspect of this investigation. KCCQ-QoL could not serve as the primary endpoint due to its limited availability in the included studies. Therefore, KCCQ-OSS was chosen as a comprehensive measure of QoL consistently reported across a wide range of studies.

3.2.1 Testing ASReview as a screening tool

Recently quite a few tools driven by artificial intelligence (AI) have been developed to streamline the study screening process for systematic reviews and meta-analyses[32].

ASReview (Active Learning for Systematic Reviews)[33], for example, is an open-source tool that employs machine learning, specifically an active learning algorithm, to assist researchers in efficiently selecting relevant studies. A key feature of ASReview is its ability to dynamically learn and enhance its performance as more data is processed, thereby prioritizing the most pertinent studies and reducing the number of items researchers must manually screen.

The process begins with the researcher manually reviewing the title and abstract of a small subset of studies labelling them as either relevant or irrelevant. ASReview utilizes this initial input to train its algorithm, which then analyzes the remaining studies and ranks them according to their predicted relevance. As more studies are labeled, the algorithm improves its predictive accuracy, allowing reviewers to concentrate on the most relevant studies early in the screening process. This significantly cuts down on the time spent on

manual screening, as the tool highlights studies likely to be important and minimizes the need to examine large quantities of irrelevant papers, making it especially valuable when dealing with large datasets in systematic reviews.
By combining machine learning with human input, ASReview is presented as a highly efficient alternative to traditional manual screening methods, accelerating the review process while maintaining high levels of precision in study selection.

An exploratory evaluation was conducted to assess the effectiveness of ASReview in identifying studies found through the conventional approach. A broader search string was used to capture a wider range of studies, including those less focused on the primary research question. This approach aimed both to evaluate ASReview's ability to retrieve all relevant papers from the initial search and to identify any additional studies that might have been overlooked.

3.3 Data retrieval and quality assessment

Full-text articles were independently reviewed by two investigators (C.O. and D.T.) with any discrepancies resolved through discussion. When duplicate publications were identified, only the most recent and complete report was included. To ascertain risk of bias, the methodologic quality of each trial was assessed using the Cochrane Risk of Bias 2 tool.[34]
Data on the following variables were extracted: study name, first author and year of publication, study design, number of patients, follow-up time, treatment groups, type of HF (including LVEF), KCCQ domains used, and KCCQ-OSS results.

3.4 Statistical analysis methods

A meta-analysis was conducted considering, as effect size of interest, the difference in mean change of the KCCQ-OSS between the SGLT2i group and the SOC group at 3 and 6 months from baseline.

To ensure the robustness and comprehensiveness of the findings, both a traditional aggregated data (AD) approach with specific enhancements for handling missing data and a more refined meta-analytical approach based on the reconstructed pseudo individual patient data (IPD) were employed.

3.4.1 AD approach

For each study, the difference in mean change at 3 and 6 months was extracted, if available.

In cases where data were missing at the 3 or 6 month timepoints, the mean change at month m in the intervention group $d_{I(m)}$ and in the control group $d_{C(m)}$ were retrieved, along with their respective standard deviations $SD(d_{I(m)})$ and $SD(d_{C(m)})$. The standard error of the mean change in the intervention group was calculated as $SE(d_{I(m)})=SD(d_{I(m)})/\sqrt{n_{I(m)}}$, where $n_{I(m)}$ represents the number of respondents in the intervention arm at the specific month m ; $SE(d_{C(m)})$ was similarly derived. The effect size for each study was estimated as $\delta_m=(d_{I(m)}-d_{C(m)})$, and its standard error as $SE(\delta_m) = \sqrt{SE(d_{I(m)})^2 + SE(d_{C(m)})^2}$.

For studies reporting only the mean μ_m and the standard deviation SD_m of the KCCQ-OSS separately for each arm, the mean change at month m was calculated as $d_m = \mu_m - \mu_0$. The standard deviation of the mean changes was estimated taking into account the correlation between the paired measures, as $SD(d_m) = \sqrt{SD_0^2 + SD_m^2 - 2\rho SD_0 SD_m}$. The correlation coefficient (ρ) was imputed as the average coefficient derived from the studies reporting the SD of the scores both at baseline and at month m, along with the SD of the mean difference between month m and baseline.

When the arm-specific mean changes or absolute values were only graphically reported, the web-based validated semi-automated tool WebPlotDigitizer[35], available for free, was used to extract the relevant information.
If KCCQ-OSS was only available at timepoints other than those selected for pooled synthesis (i.e., 3- and 6-month follow-up), the required scores were estimated using linear interpolation between the previous and subsequent timepoints, if available.
Random-effect models were used to calculate the pooled estimates, weighting each study estimate by the inverse of its variance. A pooled difference in mean change greater than 0 indicated a greater benefit in KCCQ-OSS for the SGLT2i-containing arm. Heterogeneity among studies was assessed using the I^2 index. An I^2 value of 50% or more was considered indicative of a considerable level of heterogeneity.

To assess the robustness of the findings and evaluate the influence of individual studies on the overall pooled estimates, a leave-one-out sensitivity analysis was conducted. This analysis systematically excluded one study at a time from the dataset, re-estimating the pooled effect size to determine the impact of each study on the results.
Additionally, an exploratory analysis was performed to investigate the effects of SGLT2i on QoL across the LVEF spectrum (HFrEF or at most mildly reduced vs. HFpEF or at most mildly reduced) and according to clinical status (stable chronic HF vs. acute or worsening HF). The impact of different types of SGLT2i on QoL was also evaluated.

3.4.2 Pseudo-IPD reconstruction

Meta-analyses based on IPD are considered the gold standard, offering significant advantages over those based on AD. The use of IPD allows for (a) conducting covariance analysis (ANCOVA), which can adjust for any imbalances in PROs between treatment and control groups at baseline, accounting for the correlation between baseline and change scores, and (b)

exploring potential covariate interactions, for example, between time and treatment effects. However, obtaining IPD directly from trial investigators is often impractical due to the resource-intensive nature of this process.

The algorithm developed by Papadimitropoulou et al.[36], published in 2020, offers a practical solution for reconstructing pseudo-IPD from reported AD for continuous outcomes. This method utilizes the sufficient statistics from an ANCOVA model, including the means and SDs of scores at baseline ($\bar{y}_{B,}sd_B$) and follow-up ($\bar{y}_{F,}sd_F$) for each treatment arm, along with the group correlation (r) between baseline and subsequent values.

The algorithm works as follows:

1. Two samples, Y_{i1}^* (i=1,…,n) and Y_{i2}^* (i=1,…,n), are simulated from a certain distribution, typically a standard normal distribution.
2. The simulated samples are standardized, resulting in means $\bar{y}_1^* = \bar{y}_2^* = 0$ and standard deviations $sd_1^* = sd_2^* = 1$. The correlation r^*between the two samples (Y_{i1}^* and Y_{i2}^*) is calculated.
3. Y_{i2}^* is regressed on Y_{i1}^* to determine the regression coefficient $\hat{\beta}$ and the residuals $\widehat{\varepsilon_i}$. Given that $sd_1^* = sd_2^* = 1$, $\widehat{\beta} = r^*$ and $\varepsilon_i = Y_{i2}^* - r^*Y_{i1}^*$. The residuals are also uncorrelated to Y_{i1}^* and have a variance of $1-r^{*2}$.
4. An additional sample Y_{i3}^* is generated using the equation $Y_{i3}^*=Y_{i1}^* * r + \widehat{\varepsilon_i}\sqrt{1-r^2}\,[\sqrt{1-r^{*2}}]^{-1}$.

   In practice, group correlations are rarely reported, however, when sd_B, sd_F and the standard deviation of the change from baseline sd_{change} are available, r can be calculated using the formula: $r = \frac{sd_B^2+sd_F^2-sd_{change}^2}{2*sd_B*sd_F}$.
5. Finally, the pseudo baseline values are generated as $Y_{Bi} = Y_{i1}^* * sd_B + \bar{y}_B$, while the pseudo values at the follow-up timepoint as $Y_{Fi} = Y_{i3}^* * sd_F + \bar{y}_F$.

This procedure was used to obtain pseudo-IPD that, by design, closely reflected the original data (same means, standard deviations, and correlations), enabling analyses to be conducted as if the actual IPD were available. If mean values at the specified timepoints were missing, the interpolation techniques described in the previous section were utilized to estimate the missing data.

3.4.3. One-stage/two-stage approaches based on pseudo-IPD

Once the pseudo-IPD were reconstructed, the synthesis phase was conducted using two distinct meta-analytical approaches: a) the two-stage and b) the one-stage ANCOVA methods.

a) The two-stage approach involves first obtaining study-specific estimates from the pseudo-IPD of each study, which are then combined using a traditional meta-analytic model to derive the overall effect size.

   Specifically, the regression model for each study j was expressed as:

$$Y_{Fij} = \beta_{0j} + \beta_{1j} Y_{Bij} + \beta_{2j} trt_{ij} + \varepsilon_{ij} \qquad i = 1, \ldots, n_j$$

   where Y_{Fij} represents the KCCQ-OSS score at follow-up (either 3 or 6 months) for the i-th patient in the j-th study, β_{0j} is the study-specific intercept, β_{1j} is the coefficient for the baseline score Y_{Bij}, and β_{2j} captures the study-specific treatment effect. ε_{ij} is the error term. The parameter estimates $\widehat{\beta_{2j}}$ and their standard errors were then extracted and combined using a traditional meta-analysis to estimate the pooled treatment effect.

b) The one-stage approach integrates data from all studies simultaneously using a linear mixed-effects model (LMM), that accounts for the clustering among patients within studies.

The LMM used allowed to estimate the treatment effect while accounting for potential baseline imbalances, as well as to explore the interaction between follow-up timepoints and treatment effects. The model was expressed as:

$$Diff_{ijk} = \gamma_0 + u_j + \gamma_1(Y_{Bij} - \overline{Y_B}) + \gamma_2 trt_{ij} + \gamma_3 time_k + \gamma_4(trt_{ij} * time_k) + \varepsilon_{ijk}$$

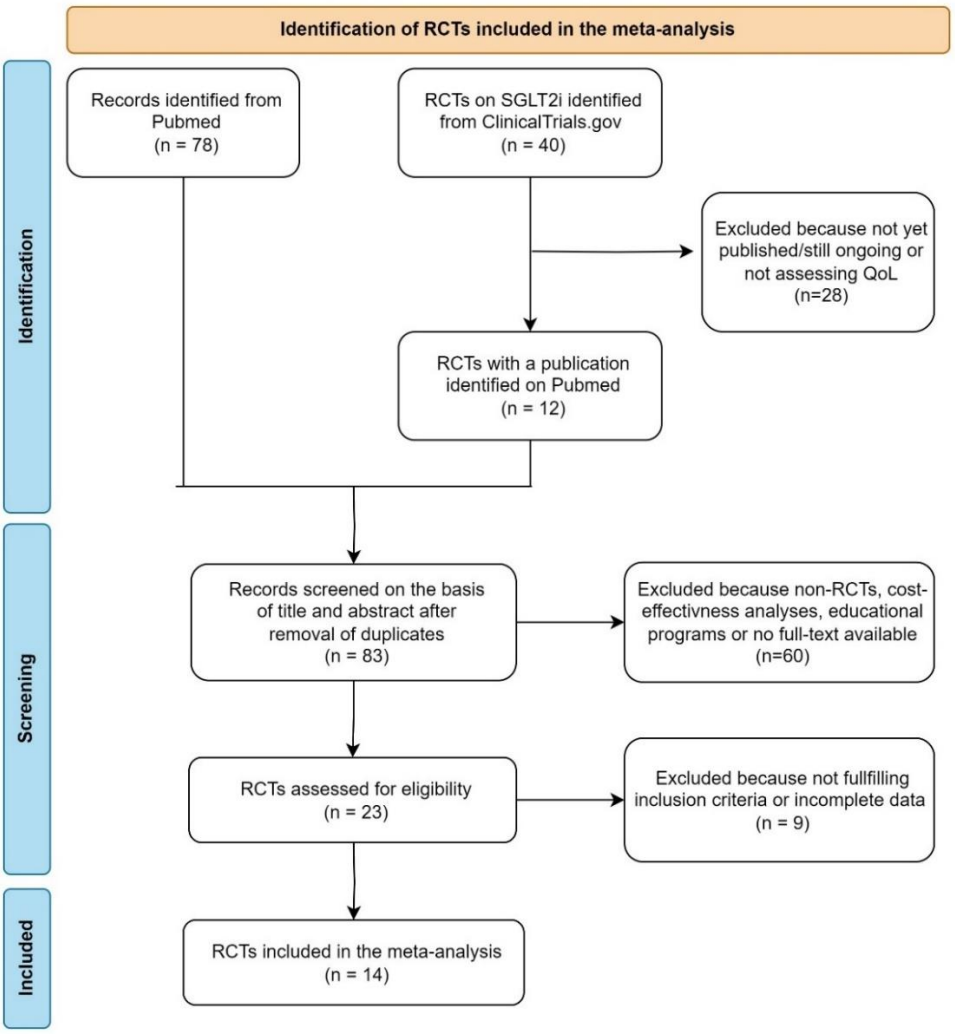
where $Diff_{ijk}$ represents the difference between the follow-up and baseline KCCQ-OSS scores for the i-th patient in the j-th study at timepoint k (either 3 or 6 months). γ_0 is the overall intercept, γ_1 adjusts for the baseline KCCQ-OSS score Y_{Bij}, which has been centered on its mean $\overline{Y_B}$, γ_2 estimates the overall treatment effect, γ_3 captures the effect of the k-th timepoint, while γ_4 represents the interaction effect between treatment and timepoint. u_j is a random effect for the j-th study, accounting for between-study variability, and ε_{ijk} is the residual error term.

A two-tailed p-value of <0.05 was considered statistically significant. All analyses were performed using SAS (version 9.4) and R (version 4.2.1).

3.5 Results

A total of 78 articles were identified through the PubMed search, and 40 RCTs were identified from ClinicalTrials.gov. After the removal of duplicates and irrelevant studies, 83 records were screened on the basis of title and abstract. Out of the 23 identified RCTs assessed for eligibility, 8 did not meet the inclusion criteria (4 RCTs used a different type of questionnaire to assess QoL, 3 reported different domains of KCCQ but not the KCCQ-OSS, and 1 had a maximum follow-up time of only 1 month) and 1 did not report complete data (**Figure 1**).

Figure 1. PRISMA flow chart.

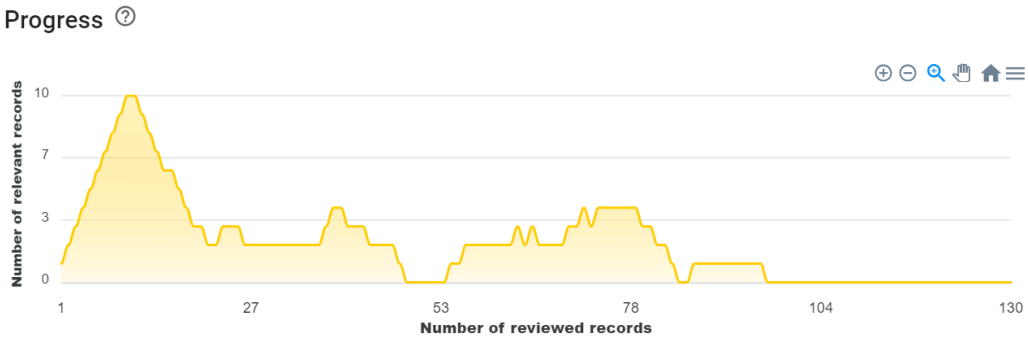


Abbreviations: QoL, quality of life; RCT, randomized controlled trial; SGLT2i, sodium glucose co-transporter 2 inhibitors.

In the analysis, 14 trials[26-38] were included, involving a total of 21737 patients, with 10935 receiving SGLT2i plus SOC and 10802 receiving SOC.

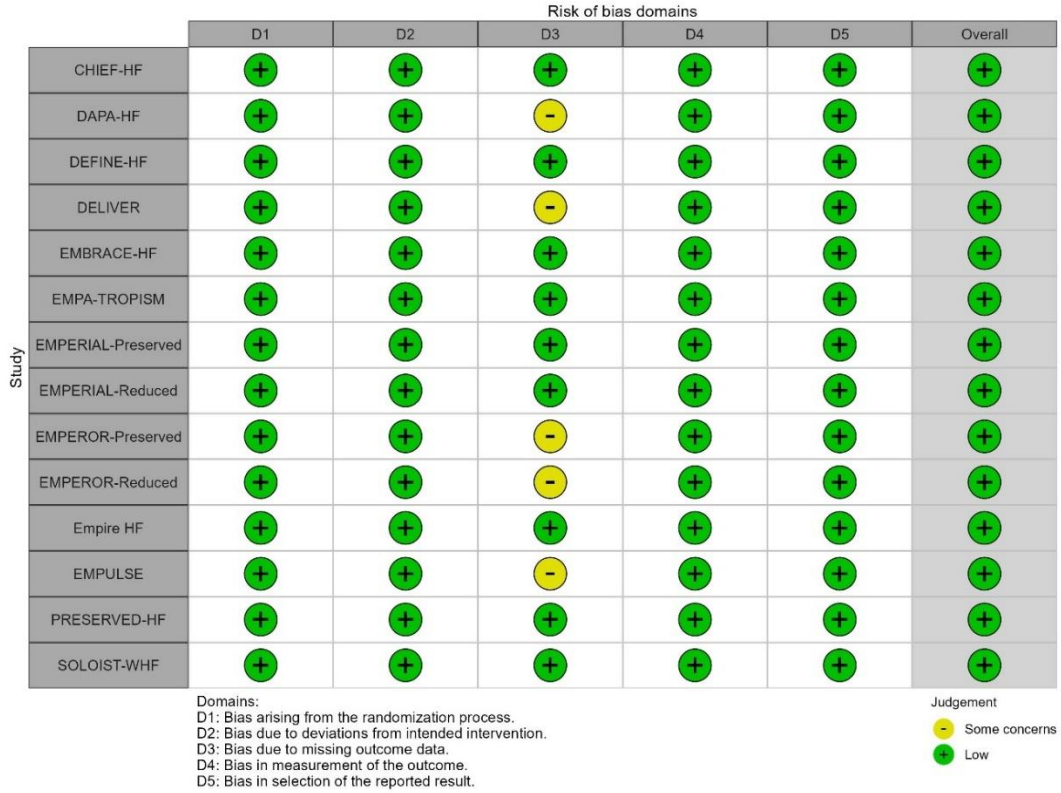
To evaluate ASReview's performance, a supplementary assessment was conducted using an expanded search string to capture a broader array of studies. **Figure 2** illustrates the ASReview process, showing the number of studies screened over time and the articles identified as relevant. Initially, there was a sharp increase in relevant studies that corresponded to trials included in the analysis, with additional relevant items consisting primarily of systematic reviews and meta-analyses on the topic. After screening approximately 130 studies, a plateau was observed in the number of relevant articles, indicating that additional screening beyond this point would likely be unnecessary, so the process was stopped. ASReview successfully identified all relevant studies from the original search except one (EMPIRE-HF), which at that point was ranked 203rd. However, while this trial did not appear among the prioritized items, it would likely have been captured through systematic reviews or meta-analyses that had already been marked as relevant. Albeit the broader search, ASReview did not reveal any additional eligible trials, supporting confidence that all relevant trials for this meta-analysis were already included.

Figure 2. ASReview screening process: cumulative count of reviewed and relevant papers.



Overall, the trials demonstrated a high quality, with low risk of bias in selection, deviations from intended interventions, outcome measurement, and result reporting (**Table 1**). The only potential bias that might have affected some of these RCTs was related to missing outcome data, although it is worth noting that all these trials ensured that the percentage of missing data was balanced between patients receiving SGLT2i plus SOC and patients receiving only SOC.

Table 1. Risk of bias assessment in the RCTs included in the analysis.



Key studies characteristics are summarized in **Table 2**. Among the included trials, 6 (43%) trials enrolled patients with HFrEF (or at most mildly reduced), 4 (29%) trials enrolled patients with HFpEF (or at most mildly reduced), and 4 (29%) trials enrolled patients with HF regardless their LVEF. Twelve studies focused on chronic HF patients, while 2 studies (EMPULSE and SOLOIST-WHF) addressed patients with acute HF or worsening HF (the DELIVER trial

primarily enrolled patients with chronic HFpEF, but it also included a subset of 654 patients with acute HF; however, stratified estimates of KCCQ-OSS for this subgroup were not available). In terms of SGLT2i agents, empagliflozin was administered in 8 studies, dapagliflozin in 4 studies, whereas canagliflozin and sotagliflozin (a dual inhibitor of SGLT2 and SGLT1) in 1 study each.

Table 2. Characteristics of the RCTs included in the meta-analysis.

Trial name	LVEF categories	Experimental arm [a] (N pts[b])	Control arm (N pts[b])	KCCQ-OSS evaluation at 3 months	KCCQ-OSS evaluation at 6 months
CHIEF-HF [37]	All	Canagliflozin (222)	SOC (226)	Yes	No
DAPA-HF [38]	HFrEF	Dapagliflozin (2222)	SOC (2221)	No*	No*
DEFINE-HF [39]	HFrEF	Dapagliflozin (131)	SOC (132)	Yes	No
DELIVER [40]	HFpEF (or HFmrEF)	Dapagliflozin (2903)	SOC (2892)	No*	No*
EMBRACE-HF [41]	All	Empagliflozin (33)	SOC (32)	Yes	No
EMPA-TROPISM [42]	HFrEF (or HFmrEF)	Empagliflozin (40)	SOC (40)	No*	Yes
EMPERIAL-Preserved [43]	HFpEF (or HFmrEF)	Empagliflozin (157)	SOC (158)	Yes	No
EMPERIAL-Reduced [43]	HFrEF	Empagliflozin (156)	SOC (156)	Yes	No
EMPEROR-Preserved [44]	HFpEF (or HFmrEF)	Empagliflozin (2884)	SOC (2867)	Yes	No*
EMPEROR-Reduced [45]	HFrEF	Empagliflozin (1776)	SOC (1753)	Yes	No*
Empire HF [46]	HFrEF	Empagliflozin (94)	SOC (92)	Yes	No
EMPULSE [47]	All	Empagliflozin (245)	SOC (250)	Yes	No
PRESERVED-HF [48]	HFpEF (or HFmrEF)	Dapagliflozin (162)	SOC (162)	Yes	No
SOLOIST-WHF [49]	All	Sotagliflozin (608)	SOC (614)	No*	No

a In addition to SOC.
b Number of patients with KCCQ-OSS evaluation at baseline.
* KCCQ-OSS estimated using available information from the previous and subsequent time points.
Abbreviations: HFmrEF, heart failure with mildly reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; KCCQ-OSS, Kansas City Cardiomiopathy Questionnaire overall summary score; LVEF, left ventricular ejection fraction; SOC, standard of care.

Regarding the patients included in the assessment of QoL, 4 studies involved a substantial number of patients (over 3000), 3 studies included fewer than 200

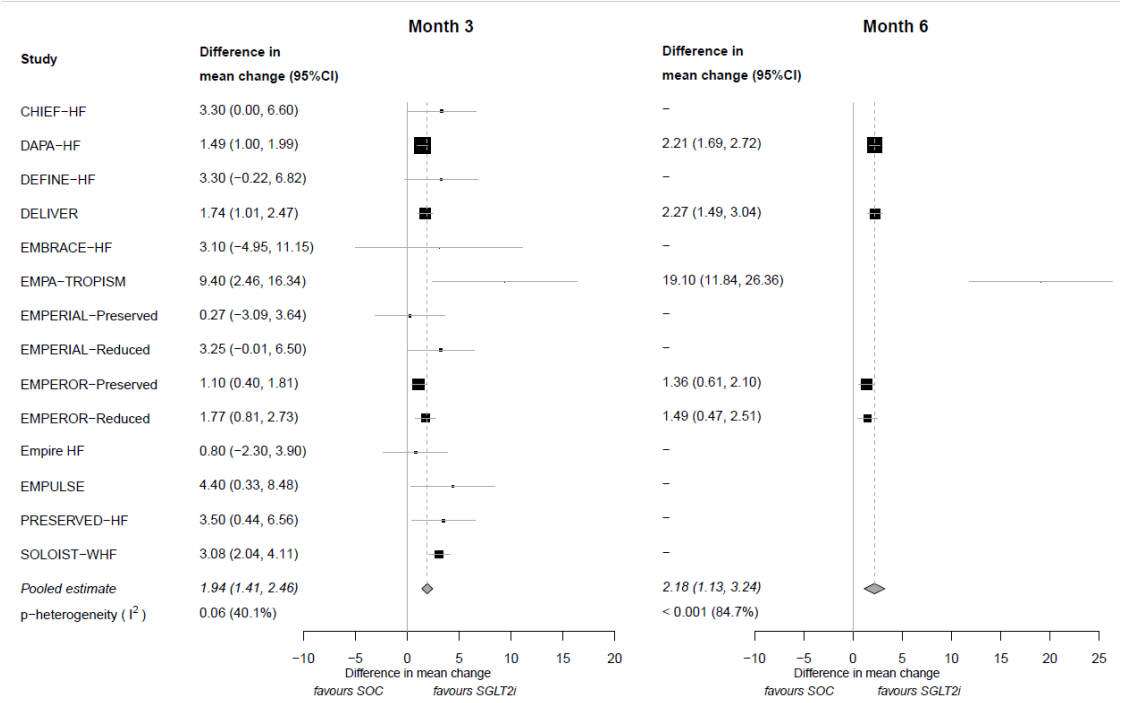
patients, while the remaining studies had varying numbers of patients in between. The shorter version of the questionnaire, KCCQ-12, was employed in 3 trials (i.e. EMPA-TROPISM, PRESERVED-HF, and SOLOIST-WHF), while the standard version with 23 items was used in the remaining studies. The median follow-up time across the included trials was 3 months, with a maximum follow-up of 12 months. In several studies, the KCCQ-OSS scores at 3 or 6 months had to be estimated using linear interpolation, relying on data from previous and subsequent assessments (**Table 2, Figure 4**).

3.5.1 Pooled estimates based on AD

At 3 months, there was a significant between-group difference in mean change of KCCQ-OSS from baseline, favoring the SGLT2i-treated group, with a pooled estimate of 1.94 (SE=0.27) (**Figures 3 and 4**). Moderate heterogeneity was observed among single study estimates (I^2 = 40.1%, p-heterogeneity =0.06).

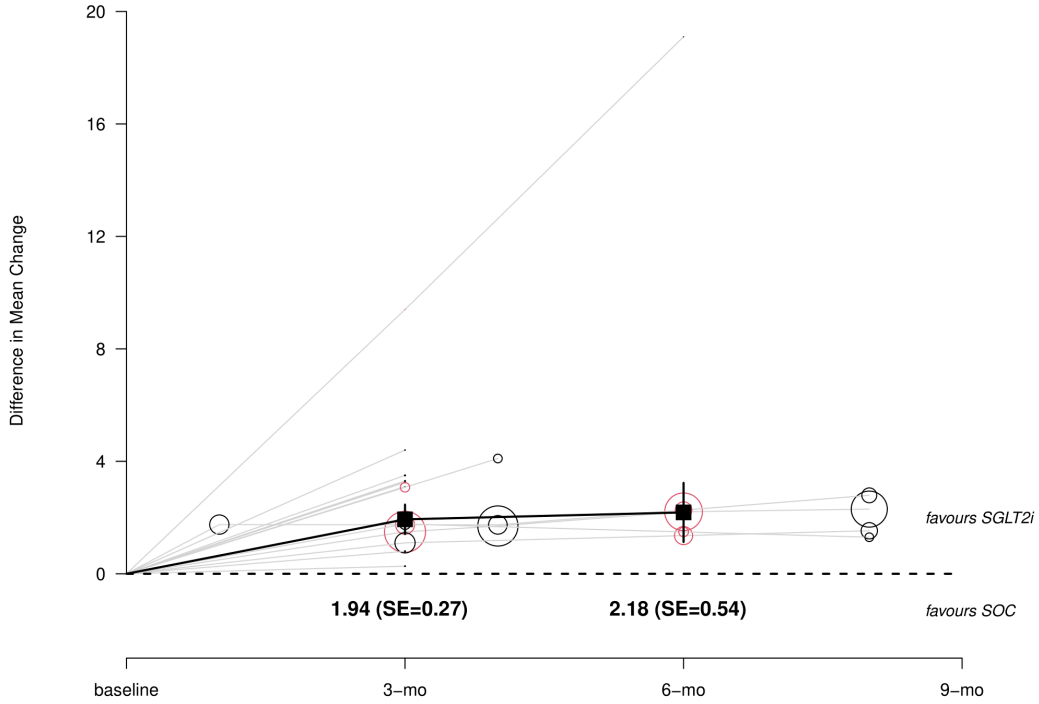
At 6 months, data on KCCQ-OSS were available for 17132 patients. The analysis showed a mean change between 6 months and baseline 2.18 (SE=0.54) points higher for patients treated with SGLT2i compared to those treated with SOC alone **(Figures 3 and 4)**. A significant heterogeneity was observed among single study estimates (I^2 = 84.7%, p-heterogeneity < 0.001).

Figure 3. Forest plot of between-arms difference in mean change of KCCQ-OSS from baseline to 3 months and to 6 months according to experimental treatment groups.



The figure shows the between-arms difference in mean change of KCCQ-OSS assessed from baseline to 3 months (left panel) or 6 months of follow-up (right panel), for patients assigned to SGLT2i-containing arms compared with those assigned to SOC arms. Each square represents the study-specific mean change difference, with values above 0 indicating superiority of the SGLT2i-containing arm over SOC arm. The size of the square is proportional to the precision of the estimate (i.e. the inverse of the variance). Horizontal lines represent the 95% confidence interval (95% CI). Diamonds represent the pooled mean change differences of KCCQ-OSS between treatment arms, calculated at 3 and 6 months of follow-up, with their corresponding 95% CI. The dashed vertical lines indicate the pooled differences in mean change, while the solid vertical line represents a mean change difference of 0, which is the null-hypothesis value (i.e. no difference between treatment arms).

Figure 4. Trajectories over time of between-arms difference in KCCQ-OSS mean change assessed in each trial and pooled estimates according to experimental treatment groups.



The figure shows difference in mean change of KCCQ-OSS over time, for each treatment comparison (grey lines) along with the meta-analytic pooled estimates according to experimental treatment groups with their corresponding 95% CI. Each grey line represents a single treatment comparison, with the size of each circle indicating the precision of the corresponding effect. For trials where the comparisons at 3 and 6 months of follow-up were not reported or derivable (red circles), these were estimated using available information from the previous and subsequent timepoints. Additionally, the numerical pooled estimates at 3 and 6 months, along with their standard errors, are reported.

3.5.2 Pooled estimates based on pseudo-IPD

The results from the pseudo-IPD meta-analysis align closely with those obtained from the one based on AD.

Using the two-stage approach, the pooled estimate for the mean change in KCCQ-OSS from baseline to 3 months was 2.06 (SE=0.29) in favor of the SGLT2i-treated group, which is very similar to the estimate obtained from the AD. At 6 months, the treatment effect is still confirmed, though slightly less pronounced than what observed with the AD analysis, with a pooled estimate

for the mean change of 1.70 (SE=0.41). Both SEs closely match those obtained in the AD analysis.

In the one-stage model, the difference between the follow-up score (at 3 or 6 months) and the baseline score was used as dependent variable. The parameter estimates from this analysis are summarized in **Table 3**. The results from the one-stage model are consistent with those obtained from the previous approaches, while offering additional insights into the time-dependent effects of the treatments. At 3 months, the pooled estimate for the mean change in KCCQ-OSS was 1.68 (SE=0.16), and at 6 months, it was 1.81 (SE=0.18).

Beyond confirming the previous findings, the one-stage model indicates a statistically significant improvement in QoL over time for all patients, regardless of whether they received SOC alone or in combination with SGLT2i (time coefficient=0.59, p=0.0007).

While patients in the SGLT2i group showed a tendency toward a faster rate of improvement (positive interaction coefficient = 0.13), the interaction term was not statistically significant (p=0.6). This suggests that the time-dependent effect of SGLT2i does not differ significantly from that of SOC alone, indicating a consistent benefit over time without a significant amplification of effect as the months progress.

The analysis also revealed a negative association between the baseline score and the follow-up score (baseline score coefficient= -0.26), consistent with the phenomenon of regression to the mean, where patients with lower baseline scores tend to show greater improvement.

The one-stage LMM analysis yielded smaller SEs for the estimated treatment effects at both 3 and 6 months compared to both the two-stage approach and the AD analysis.

Table 3. Parameter estimates from the one-stage LMM

Parameter	Level/Units	Coefficient	SE	p-value
Intercept		6.17	1.59	0.002
Baseline score	+1	-0.26	0.003	<.0001
Treatment	SGLT2i vs SOC	1.68	0.16	<.0001
Months	6 vs 3	0.59	0.18	0.0007
Months*treatment	6 vs 3 SGLT2i vs SOC	0.13	0.24	0.6

Abbreviations: SGLT2i, sodium glucose co-transporter 2 inhibitors; SOC, standard of care;

3.5.3 Subgroup and sensitivity analysis

The leave-one-out sensitivity analysis, conducted on AD, showed that the pooled estimates remained consistent and statistically significant after the exclusion of each individual study (**Table 4**).

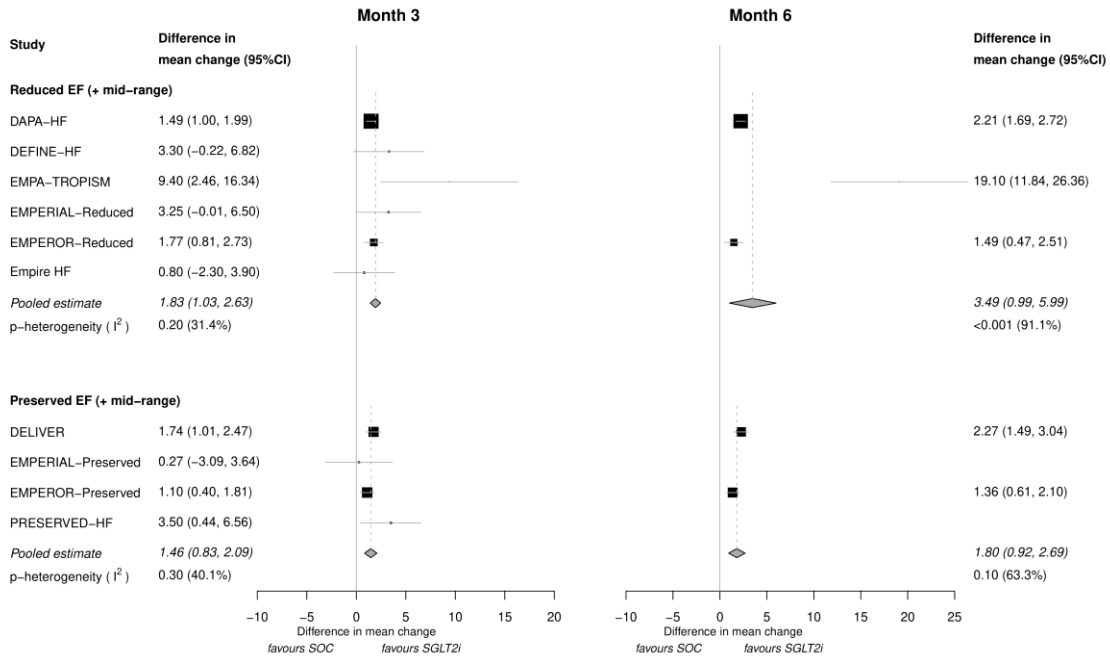
Table 4. Leave-One-Out analysis of pooled estimates.

Excluded study	Difference in mean change at 3 months (95%CI)	Difference in mean change at 6 months (95%CI)
-	1.94 (1.41, 2.46)	2.18 (1.13, 3.24)
CHIEF-HF	1.91 (1.37, 2.44)	
DAPA-HF	2.12 (1.45, 2.78)	2.47 (0.84, 4.10)
DEFINE-HF	1.91 (1.37, 2.45)	
DELIVER	2.04 (1.39, 2.68)	2.33 (0.87, 3.79)
EMBRACE-HF	1.94 (1.40, 2.49)	
EMPA-TROPISM	1.84 (1.38, 2.30)	1.90 (1.44, 2.36)
EMPERIAL-Preserved	1.98 (1.44, 2.52)	
EMPERIAL-Reduced	1.91 (1.37, 2.44)	
EMPEROR-Preserved	2.11 (1.54, 2.68)	2.58 (1.18, 3.98)
EMPEROR-Reduced	2.01 (1.39, 2.62)	2.47 (1.14, 3.80)
Empire HF	1.98 (1.43, 2.53)	
EMPULSE	1.89 (1.37, 2.41)	
PRESERVED-HF	1.89 (1.36, 2.42)	
SOLOIST-WHF	1.66 (1.23, 2.08)	

Role of LVEF

The positive effects of SGLT2i on QoL were consistent across the entire spectrum of LVEF, with similar results when considering trials enrolling patients with reduced (or at most mildly reduced) LVEF and patients with preserved (or at most mildly reduced) LVEF. Differences in mean change at 3 months were 1.83 (95% CI, 1.03–2.63) and 1.46 (95% CI, 0.83–2.09), respectively, with no significant heterogeneity between the pooled estimates (p-heterogeneity =0.58). The pooled estimates at 6 months, stratified by LVEF, also show no significant differences between groups (p-heterogeneity = 0.44) (**Figure 5**).

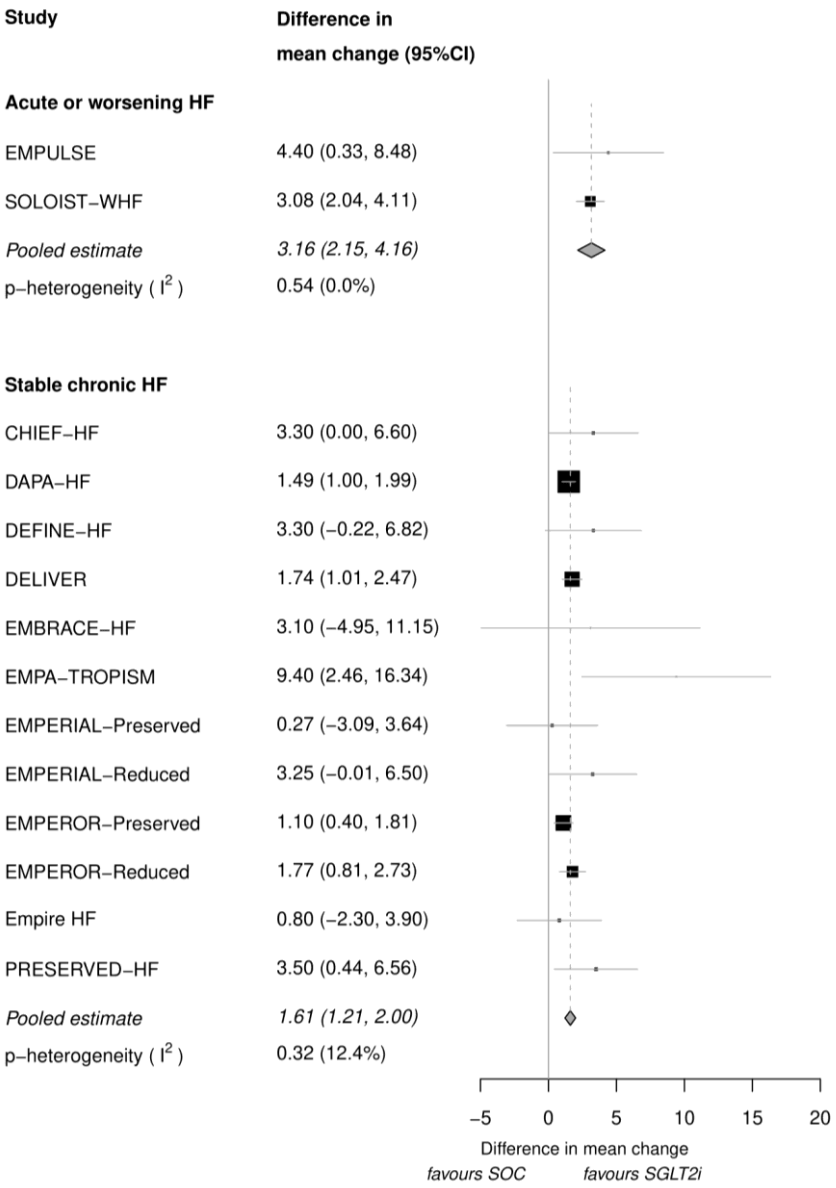
Figure 5. Forest plot of between-arms difference in mean change of KCCQ-OSS from baseline to 3 months and to 6 months according to experimental treatment groups stratified by LVEF.



The figure shows the between-arms difference in mean change of KCCQ-OSS assessed from baseline to 3 months (left panel) or 6 months of follow-up (right panel), for patients assigned to SGLT2i-containing arms compared with those assigned to SOC arms. Studies are grouped according to LVEF (i.e. reduced or at most mildly reduced, preserved or at most mildly reduced). Each square represents the study-specific mean change difference, with values above 0 indicating superiority of the SGLT2i-containing arm over SOC arm. The size of the square is proportional to the precision of the estimate (i.e. the inverse of the variance). Horizontal lines represent the 95% confidence interval (95% CI). Diamonds represent the pooled mean change differences of KCCQ-OSS between treatment arms, calculated at 3 and 6 months of follow-up, with their corresponding 95% CI. The dashed vertical lines indicate the pooled differences in mean change, while the solid vertical line represents a mean change difference of 0, which is the null-hypothesis value (i.e. no difference between treatment arms).

Stable vs worsening HF

Figure 6. Forest plot of between-arms difference in mean change of KCCQ-OSS from baseline to 3 months according to experimental treatment groups stratified by stable chronic HF vs. acute or worsening HF.



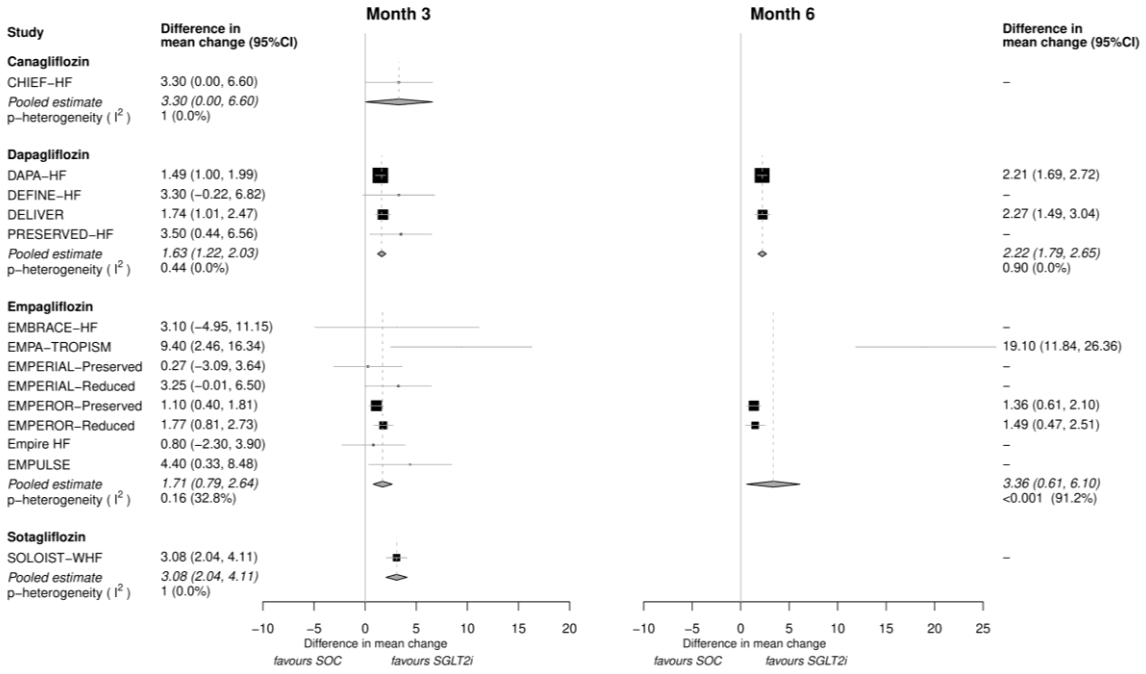
The figure shows the between-arms difference in mean change of KCCQ-OSS assessed from baseline to 3 months, for patients assigned to SGLT2i-containing arms compared with those assigned to SOC arms. Studies are grouped according to stable chronic HF vs. acute or worsening HF. Each square represents the study-specific mean change difference, with values above 0 indicating superiority of the SGLT2i-containing arm over SOC arm. The size of the square is proportional to the precision of the estimate (i.e. the inverse of the variance). Horizontal lines represent the 95% confidence interval (95% CI). Diamonds represent the pooled mean change differences of KCCQ-OSS between treatment arms, calculated at 3 months of follow-up, with their corresponding 95% CI. The dashed vertical lines indicate the pooled differences in mean change, while the solid vertical line represents a mean change difference of 0, which is the null-hypothesis value (i.e. no difference between treatment arms).

This subgroup analysis was feasible only at 3 months since the two trials including patients with acute or worsening HF (i.e. EMPULSE and SOLOIST-WHF) did not extend beyond this timeframe. In both groups, patients treated with SGLT2i in addition to SOC experienced a positive impact on their QoL. Specifically, the difference in mean change was 3.16 (95% CI, 2.15–4.16) for patients with an ongoing or a recent hospitalization due to acute HF and 1.61 (95% CI, 1.21–2.00) for those with chronic stable HF (**Figure 6**). A significant heterogeneity was found between the pooled estimates (p-heterogeneity = 0.006).

SGLT2i type

The stratified analysis based on the type of SGLT2i administered included canagliflozin, dapagliflozin, empagliflozin, and sotagliflozin at 3 months, whereas only dapagliflozin and empagliflozin at 6 months as some studies did not extend beyond the 3-month follow-up. Across all SGLT2i types, when added to the SOC, a positive impact on patients' QoL was consistently observed. Pooled estimates at 3 and 6 months, stratified by SGLT2i type, are reported in **Figure 7**. No statistically significant heterogeneity was found among different drugs.

Figure 7. Forest plot of between-arms difference in mean change of KCCQ-OSS from baseline to 3 months and to 6 months according to experimental treatment groups stratified by SGLT2i type.



The figure shows the between-arms difference in mean change of KCCQ-OSS assessed from baseline to 3 months (left panel) or 6 months of follow-up (right panel), for patients assigned to SGLT2i-containing arms compared with those assigned to standard of care (SOC) arms. Studies are grouped according to SGLT2i type (i.e., canagliflozin, dapagliflozin, empagliflozin, sotagliflozin). Each square represents the study-specific mean change difference, with values above 0 indicating superiority of the SGLT2i-containing arm over SOC arm. The size of the square is proportional to the precision of the estimate (i.e., the inverse of the variance). Horizontal lines represent the 95% confidence interval (95% CI). Diamonds represent the pooled mean change differences of KCCQ-OSS between treatment arms, calculated at 3 and 6 months of follow-up, with their corresponding 95% CI. The dashed vertical lines indicate the pooled differences in mean change, while the solid vertical line represents a mean change difference of 0, which is the null-hypothesis value (i.e., no difference between treatment arms).

3.6 Discussion

PROs are crucial in healthcare research, particularly for chronic conditions like HF, as they provide direct insights into patients' symptoms, QoL, and overall health. These self-reported measures eliminate interpretation bias, offering a more accurate reflection of the patient's experience. However, the use of PROs in RCTs presents significant challenges for quantitative analysis. The variability in questionnaire designs, scoring systems, and the different interpretation of scales complicates direct comparisons across studies. Additionally, the longitudinal nature of PRO data, with follow-up intervals that often differ between trials, and the data sometimes only graphically reported, further complicates meta-analyses that may overlook this variability by either focusing solely on the final QoL measurement or synthesizing only data reported at a specific timepoint. These issues can lead to significant information loss and introduce bias, ultimately compromising the reliability of meta-analytic findings.

A meta-analysis was conducted to address the challenges associated with analyzing longitudinal QoL data, using SGLT2i as a case study in the context of HF. The primary objective was to evaluate the impact of initiating SGLT2i, in addition to the SOC, on QoL (assessed with the KCCQ-OSS) in patients with HF, irrespective of LVEF and clinical setting. This was achieved through a systematic meta-analysis of RCTs, employing different analytical approaches based on AD and pseudo-IPD to address common challenges in PRO analysis. In addition, linear interpolation techniques were used to estimate missing timepoints in longitudinal data, and information was extracted from graphical illustrations when necessary. By leveraging these methods, the analysis aimed to minimize information loss and provide deeper insights into the expected QoL improvement over time, as well as the interaction between time and treatment, which traditional meta-analyses often overlook.

3.6.1 Evaluation of meta-analytical and screening approaches

The use of methods such as linear interpolation and data extraction from graphical representations, even when working solely with AD, significantly reduced information loss and provided more reliable estimates at specific timepoints. By generating estimates at defined intervals, such as 3 and 6 months, this approach can provide clinicians and patients with precise, actionable information on both the expected magnitude of QoL improvement and the timeframe within which these improvements are likely to occur. This is particularly important in clinical practice, as it enables better planning and management of patient expectations.

When comparing the different approaches, the results based on the reconstruction of pseudo-IPD were similar to those derived from AD, reinforcing the validity of both methods. However, the one-stage model demonstrated clear advantages over the other approaches. It not only confirmed the overall treatment effects but produced smaller SEs compared to both the two-stage approach and the AD meta-analysis, indicating greater precision in the estimates and reinforcing the reliability of the findings regarding the treatment effects over time. Noteworthy, the one-stage model allowed for the identification of a positive coefficient related to time, indicating that patients experienced improvement in QoL over time whether they were treated with SGLT2i or SOC alone. This longitudinal insight, reflecting the evolution of patient health status, may not be captured in a traditional meta-analysis focused solely on mean changes between treatment groups. In such a meta-analysis, a positive mean difference in favor of SGLT2i could be observed, even if both groups experienced a decline in QoL, or if both improved but to different extents. The one-stage model, by contrast, provides a clearer depiction of the QoL trajectory over time in both treatment and control groups. For these reasons, the one-stage approach is recommended as the most robust method for analyzing longitudinal QoL data in future studies, as it offers more comprehensive and nuanced insights into treatment effects over time.

For the screening process, ASReview was tested as a supplementary tool to assess its potential as an efficient alternative to the traditional, time-intensive approach. The tool demonstrated significant efficiency in streamlining the review process, successfully identifying key studies within a reduced screening timeframe while maintaining a good balance between comprehensiveness and speed. Although it ranked one relevant study lower, this study was also cited within meta-analyses previously identified as relevant, indicating it would have been captured anyway. This suggests that ASReview, alongside other AI-driven tools that have not been directly tested in this study, could effectively replace traditional screening methods, particularly when handling large datasets. Given the rapid advancements in technology, it would be worthwhile for researchers to explore these screening options to enhance the efficiency and accuracy of systematic reviews.

3.6.2 Clinical interpretation and implication for clinical practice

In this meta-analysis, involving over 20,000 patients across 14 RCTs, evidence of the positive impact of SGLT2i on QoL in HF patients was confirmed. Despite a modest, albeit significant, improvement in KCCQ-OSS of less than 5 points at 3 and 6 months, this benefit was consistent across the entire spectrum of LVEF and various clinical settings, including both stable chronic HF and acute or worsening HF.

Previous meta-analyses, such as that by Vaduganathan et al., demonstrated the efficacy of SGLT2i in reducing all-cause mortality, cardiovascular mortality, and HF-related hospitalizations.[50] Additionally, improvements in QoL have been observed in multiple domains of KCCQ, as shown in trials like DAPA-HF, DELIVER, EMPEROR-Preserved and EMPEROR-Reduced.[38,44,45,51] In individual patient-level pooled analyses of DAPA-HF and DELIVER, dapagliflozin improved multiple domains of health status as measured by KCCQ, regardless of LVEF.[52] In the EMPULSE trial, initiation of empagliflozin in patients hospitalized for acute HF ameliorated symptoms compared to placebo as early as 15 days and through 3 months follow-up.[47] The current meta-analysis, by

showing early significant improvement in KCCQ-OSS (since the third month), confirms and extends previous results. Interestingly, patients who started SGLT2i treatment during a hospitalization or shortly after an episode of worsening HF showed a greater improvement in QoL compared to those who started the therapy during the chronic phase of HF. This may be related to the steep decline in KCCQ scores that often precedes a HF hospitalization, indicating that SGLT2i could play a critical role in enhancing recovery after significant health deterioration.[53] These findings highlight the potential benefit of initiating SGLT2i early, particularly during or following a hospitalization for acute HF, to optimize clinical outcomes and accelerate improvements in health status.[54,55]

Although the results showed a progressive improvement in QoL at both 3 and 6 months, the observed benefit remained below the standard threshold of 5 points, which is typically considered as relevant in most recent studies.[56,57] However, despite the modest changes in mean KCCQ-OSS scores, individual-level pooled analyses from studies like DAPA-HF and DELIVER revealed a significantly higher proportion of patients treated with dapagliflozin who achieved substantial improvements in health status (≥10 or ≥15 point increases) compared to those receiving placebo. This suggests that, although the mean improvement might appear modest, a notable subset of patients experienced meaningful QoL gains.[52]

The stratified analysis based on the type of SGLT2i administered showed a consistent benefit on QoL, irrespective of the specific drug used, reinforcing the growing consensus of a class effect of SGLT2i in HF management. These benefits were observed across the entire spectrum of LVEF, aligning with recent updates in both European and American guidelines, which recommend SGLT2i for patients with reduced, mildly reduced, and preserved LVEF.[21,22,58–60]

SGLT2i are nowadays recommended in patients with HFrEF as an addition to OMT with renin–angiotensin–aldosterone system inhibitors, beta-blockers, and mineralcorticoid receptor antagonists. In patients with HFpEF, the treatment of comorbidities is also recommended. Importantly, the benefits of

gliflozins were shown to be independent and additive from other treatments.[50,54] Furthermore, the analysis showed that patients receiving SOC therapy, without SGLT2i, also experienced improvements in QoL over time. This could be due to the close follow-up and high-intensity care typically provided in clinical trials, as well as optimization of guideline-recommended medical therapy before enrollment or during hospitalization. However, it is important to highlight that while improvements in QoL were observed in both groups, the magnitude of improvement was significantly higher in patients receiving SGLT2i, underscoring the additional benefit of these medications when added to SOC.

3.6.3 Limitations

This analysis has several limitations. First, it was based on published AD rather than IPD. However, this limitation was partially mitigated by the use of reconstructed IPD, which allowed for a more comprehensive analysis and helped overcome some of the inherent restrictions of working with AD.

Second, linear interpolation was employed to estimate the KCCQ-OSS at 3 or 6 months when data were missing. Although this method is widely accepted in such cases, it introduces a potential source of error. Nevertheless, it offers a more reliable solution compared to excluding studies that did not report data at specific timepoints or relying solely on final estimates from studies with varying follow-up durations. Ignoring these discrepancies could lead to information loss or bias, so interpolation allowed for a more consistent comparison across studies at key intervals.

Third, the focus was on studies that assessed QoL using the KCCQ-OSS. Different domains of the KCCQ or the use of alternative questionnaires, such as the MLHFQ, may yield different results. While significant differences are not anticipated, this variability in QoL assessments could influence outcomes. However, aside from the KCCQ-QoL subscale, which specifically measures QoL but was reported in very few studies, the KCCQ-OSS was the only tool that

incorporates these specific QoL items. For this reason, it was considered the most practical and widely applicable choice for this analysis.

Additionally, the analysis included studies that assessed QoL either as a primary or secondary endpoint. Nonetheless, there is a high level of confidence in the absence of publication bias, as no statistically significant differences were observed between studies that focused on QoL as the primary endpoint (e.g., CHIEF-HF, PRESERVED-HF) and those that assessed it as a secondary outcome (data not shown).

Another limitation relates to the heterogeneity observed at the 6-month follow-up between individual study estimates. This variability could stem from differences in study populations, treatment protocols, or baseline characteristics. Random-effects models were employed to account for these variations, thus enhancing the robustness of the findings.

From a clinical perspective, a notable limitation is the lack of detailed information on functional capacity measures such as cardiopulmonary exercise testing, 6-MWT, or NYHA class. These measures could have provided a more comprehensive evaluation of health status in HF patients. However, the use of PROs, such as the KCCQ, has been shown to provide reliable, patient-centered insights into disease progression and treatment effects. Future research should aim to integrate functional measures alongside PROs to further refine the understanding of QoL in HF patients.

Notwithstanding these limitations, the analysis provided valuable evidence regarding the effect of SGLT2i on QoL in individuals with HF. Further research, preferably using IPD and comprehensive outcome measures, would be beneficial to validate and expand upon these findings.

4 Joint modeling of survival and QoL data

4.1 Challenges introduced by PROs and survival data

Most studies, including both RCTs and non-randomized studies, still tend to analyze PROs and survival outcomes separately.[19]

Continuous longitudinal outcomes, such as PROs, are usually analyzed using a LMM[61], which can be expressed as:

$$Y_i(t) = \boldsymbol{X_i^T}(t)\boldsymbol{\beta} + \boldsymbol{Z_i^T}(t)\boldsymbol{b_i} + \varepsilon_i(t), \qquad \varepsilon_i(t) \sim N(0, \sigma^2)$$

where $Y_i(t)$ represents the longitudinal continuous outcome (e.g., the KCCQ score describing QoL) observed for individual i at time t. $X_i^T(t)$ is a vector of covariates associated with fixed effects, modeled through the parameter β. These represent the general impact of observable variables common to all individuals on the value of $Y_i(t)$ (e.g., treatment, age, or other demographic factors). $Z_i^T(t)$ is a vector of covariates describing individual and temporal variations related to random effects. The random effects, b_i, reflect the individual variation from the population mean, i.e., the unexplained variability among patients in terms of individual differences in disease progression or treatment response. $\varepsilon_i(t)$ is the error associated with the outcome measurement for individual i at time t, and it is normally distributed.

Time-to-event outcomes, such as death, are generally modeled using a proportional hazards model[61], expressed as:

$$h_i(t) = h_0(t) \exp(\varphi^T \boldsymbol{v_i})$$

where $h_i(t)$ denotes the hazard rate for the i-th patient at time t, obtained as the product of a baseline hazard $h_0(t)$, a baseline risk common to all individuals at time t, which can be modeled parametrically, flexibly (using splines), or left unspecified (as in the Cox model) and a risk factor

exp ($\varphi^T \boldsymbol{v_i}$) associated with the i-th individual, where $\boldsymbol{v_i}$ is a vector of covariates that may influence the event risk, and φ^T is a vector of coefficients quantifying their impact. Each element of φ represents the logarithm of the relative risk (hazard ratio, HR) for each of the covariates in $\boldsymbol{v_i}$. A positive value of φ_j implies an increased risk associated with the covariate v_{ij}, while a negative value indicates a reduction in risk. In general, if exp ($\varphi^T \boldsymbol{v_i}$) >1, the individual has a higher event risk compared to the baseline risk, while if exp ($\varphi^T \boldsymbol{v_i}$) < 1, the risk is lower.

However, if the processes of QoL and survival are interconnected, analyzing them separately does not account for how one may influence the other, leading to crucial information loss and potentially biased estimates.
The issue can be viewed from two different perspectives, depending on the primary outcome of interest. Key questions to reflect on are:

1. Does a patient with lower QoL have a higher risk of death? If so, does this relationship influence QoL trajectory estimates over time?

2. Does a change in the QoL trajectory impact the risk of death?

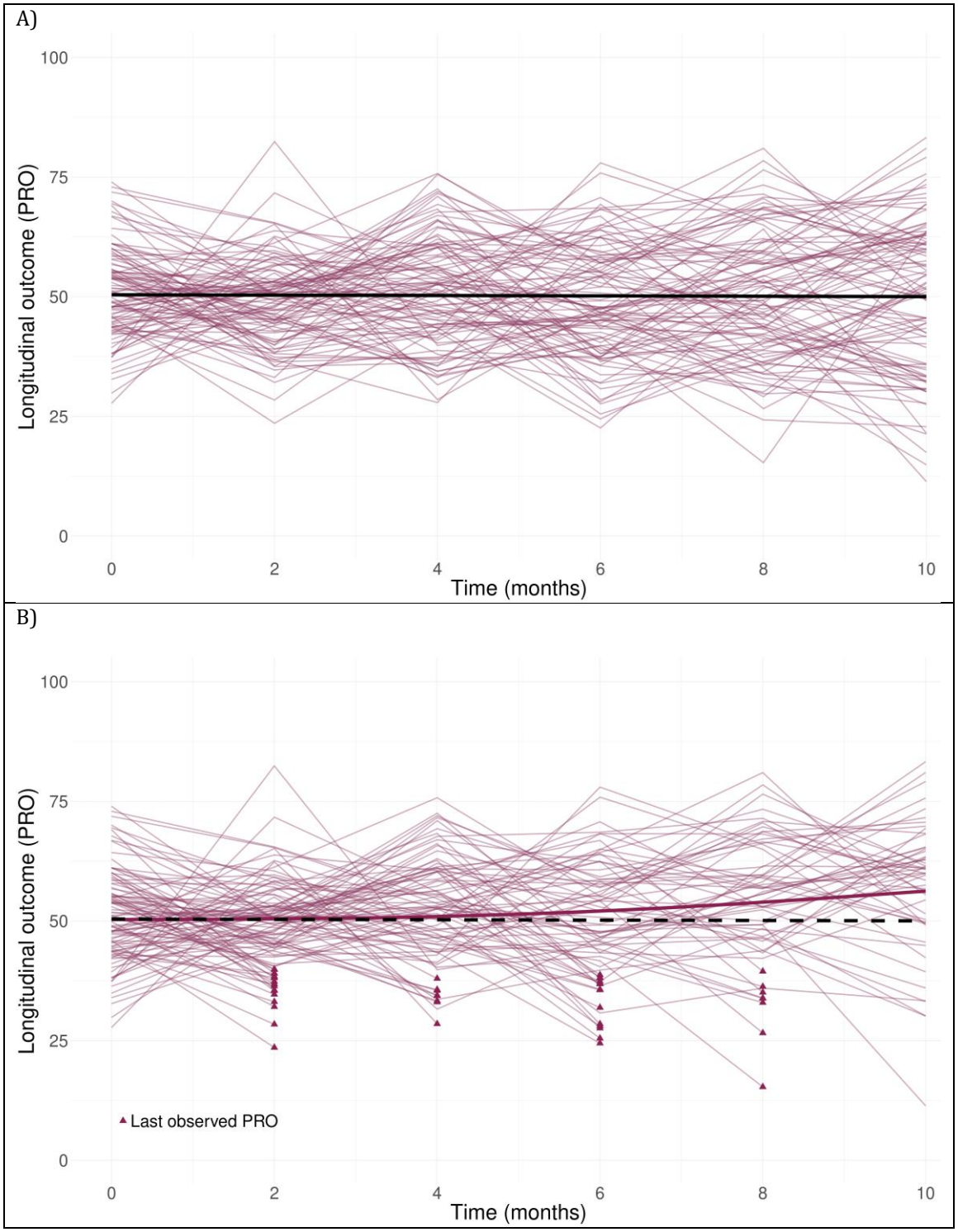
1. In longitudinal trials, informative dropout (e.g., patient death due to lower QoL scores) can introduce significant bias. Since PROs are collected only from patients still alive at the time of questionnaire administration, the surviving population tends to be composed of individuals with generally better health. This leads to an overestimation of overall QoL because patients with lower scores tend to die and therefore exit the study early. This distortion is immediately evident graphically: consider a situation where observed QoL in a patient group remains, on average, constant over a period of 10 months. If all patients were observed until the end of the study, the observed QoL trajectory would reflect the real one (**Figure 8**, panel A). In contrast, if patients with lower QoL scores had a higher risk of death and progressively left the study, the observed mean QoL would be

skewed as the more vulnerable patients are excluded from measurements (**Figure 8**, panel B). In a hypothetical extreme scenario where treatment and control have no direct effect on an individual's QoL, but the treatment keeps all patients alive while the control group allows those with poorer health to die, an analysis of PRO scores alone might lead to the incorrect conclusion that the control provides a better quality of life, albeit for a shorter time. Relying on this misleading evidence, a patient who values QoL over extended survival might choose the control, clearly making the wrong choice.[19]

2. In a survival model that only considers the baseline PRO value, both statistical and clinical valuable information is lost. The interest indeed may lie specifically in analyzing score changes over time to understand if such changes are associated with prognosis improvement or worsening. It could also be worthwhile to explore whether this change might represent a surrogate indicator of prognosis.

   In certain contexts, attempts have been made to address this issue by including time-dependent covariates in survival models, such as in the extended Cox model. However, this solution does not account for measurement error or the intrinsic individual variability that might influence the longitudinal variable's trajectory between follow-up visits (even when not measured) and, consequently, its relationship with the risk of death.[61]

Figure 8. Impact of informative dropout on QoL measurement in longitudinal studies.



Panel A: Simulation of a longitudinal study where QoL remains constant over time with no dropout.
Panel B: Simulation of the same study with dropout (due to death) of patients with lower QoL scores. The solid lines represent the QoL trend in patients still in the study. The triangle indicates the last observation for a deceased and censored patient before the next follow-up. The black line (both solid and dashed) represents the true mean QoL estimate, while the colored solid line indicates the distorted mean QoL estimate due to informative dropout.

4.2 Joint models

A more advanced solution to the above-mentioned problems is provided by joint models, which combine the analysis of repeated QoL measurements and the event of interest (such as death), accounting for measurement error. These models enable comprehensive use of all longitudinal observations, taking into account their endogenous nature, thereby reducing bias and improving the efficiency of estimates.

Introduced in 1996 in the context of AIDS studies, they were subsequently developed in oncology and, over the last 20 years, have seen increasing use.[62] However, despite their ability to enhance the analysis of complex data, their application remains limited due to their inherent statistical complexity.

Joint models essentially consist of two integrated components[61]:

- A **longitudinal component**, modeled using a LMM, which specifically takes the following form:

$$Y_i(t) = m_i(t) + \varepsilon_i(t), \qquad \varepsilon_i(t) \sim N(0, \sigma^2)$$

$$m_i(t) = \boldsymbol{X_i^T}(t)\boldsymbol{\beta} + \boldsymbol{Z_i^T}(t)\boldsymbol{b_i}, \qquad \boldsymbol{b_i} \sim N(0, \Sigma)$$

  where $Y_i(t)$ represents the continuous longitudinal outcome (PRO), and $\boldsymbol{X_i^T}(t)$ is a vector of covariates associated with fixed effects, modeled through the parameter $\boldsymbol{\beta}$. $\boldsymbol{Z_i^T}(t)$ represents covariates associated with random effects, while $\boldsymbol{b_i}$ denotes the unexplained individual variability. Here, $m_i(t)$ represents the true unobserved value of the longitudinal outcome for the i-th patient at time t, distinct from $Y_i(t)$, which includes the measurement error $\varepsilon_i(t)$.

- A **survival component**, modeled using a proportional hazards model, which specifically takes the following form:

$$h(t|M_i(t), \boldsymbol{v_i}) = h_0(t) \exp[\varphi^T \boldsymbol{v_i} + \alpha m_i(t)]$$

where $h_0(t)$ is the baseline hazard, and $\boldsymbol{v_i}$ is a vector of time-independent covariates at baseline (such as the patient's comorbidities or assigned treatment) with the corresponding vector of log-hazard ratios, φ^T. $M_i(t) = \{m_i(s), 0 \leq s \leq t]\}$ indicates the history of the true unobserved longitudinal process up to time t.

In joint models, it is recommended to explicitly specify the baseline hazard, as leaving this component unspecified can lead to an underestimation of the SEs in parameter estimates. Using a parametric hazard function, based on known distributions (e.g., Weibull or log-normal), or a more flexible specification of the baseline hazard, helps to improve the accuracy of estimates and reduce the risk of error.[61]

These two components are then connected through shared random-effect parameters. Specifically, the parameter α quantifies the effect of the underlying longitudinal outcome (QoL) on the risk of death. $\exp(\alpha)$ indicates the relative increase in the risk of an event at time t resulting from a one-unit increase in $m_i(t)$ at that same time point. In other words, the term $\alpha m_i(t)$ is incorporated into the linear predictor of the survival model, acting as a form of parameterization that reflects the current value of the longitudinal outcome. This approach is particularly useful because it establishes a direct link between the expected QoL value and the risk of death. Moreover, it takes into account that, at the time of the event, the QoL measurement may not be available, making the inclusion of the expected value a strategic method to improve risk estimates.

4.2.1 JM Package

The JM package in R[63] is one of the primary tools for implementing joint models. This package has been designed to fit a broad range of joint models for continuous longitudinal responses and time-to-event data using maximum likelihood estimation. Its main function, *jointModel(),* integrates a LMM (created via *lme()* from the *nlme* package) and a survival model generated by *coxph()* or *survreg()* from the survival package.

The flexibility of *jointModel()* lies in its variety of survival submodel options, which are specified through the method argument. Each method provides distinct formulations for the baseline hazard function, including the following options:

- *weibull-AFT-GH*: fits a Weibull model under the accelerated failure time (AFT) structure;
- *weibull-PH-GH*: fits a Weibull model under a proportional hazards framework;
- *piecewise-PH-GH*: fits a piecewise-constant baseline risk model, allowing control over the intervals of the time scale and estimating the hazard over these intervals;
- *Cox-PH-GH*: fits a proportional hazards model with an unspecified baseline;
- *spline-PH-GH*: uses spline functions to approximate the log baseline hazard function, offering flexibility with the number and position of internal knots for precise modeling.

For each option, *jointModel()* leverages Gauss-Hermite quadrature to approximate complex integrals, but in high-dimensional random effects structures, the ch-Laplace method can be selected for improved computational efficiency.

Following a random-effects approach, the JM package restricts models to linear mixed-effects submodels with independent and identically distributed errors and no serial correlation, so users should avoid specifying additional

correlation structures or variance functions when calling *lme().* The model-fitting procedure combines the Expectation-Maximization (EM) algorithm with a quasi-Newton optimizer, such as *BFGS* (from *optim()*) or *nlminb(),* depending on convergence needs). For more computationally demanding models, especially those using a Cox-PH structure, only the EM algorithm is applied, making the approach flexible yet computationally feasible.

Beyond model fitting, JM provides several functions for extracting model summaries, coefficient estimates, and empirical Bayes predictions, among other diagnostics. The flexibility and comprehensiveness of JM have made it a valuable tool for complex longitudinal data analysis, even though its statistical complexity may present challenges for broader adoption.

4.3 Data simulation

As described in Chapter 2, the MITRADVANCE-HF trial, which initiated this line of research due to a practical challenge, is still in the recruitment phase. Consequently, joint data on survival and QoL were simulated to evaluate the joint models.

Several scenarios were proposed, such as cases where the treatment could either positively or negatively impact QoL without influencing survival, or conversely, improve survival without affecting QoL, among other possible scenarios.

In detail, the following steps were taken to simulate the data:

Step 1

To conduct the simulation in a clear manner, the follow-up period was set at 5 years and divided into monthly intervals. In each interval, the QoL score for patient i in group j at time k was simulated using a LMM with the following specification:

$$QoL_{ijk} = \alpha + u_i + (\beta_{time} + u_{i\_time})time_k + \beta_{trt} * trt_j + \beta_{time*trt}(time_k * trt_j) + \varepsilon_{ijk}$$

The parameters were set as follows:

- α: represents the average baseline QoL, fixed at 50.
- u_i: random intercept effect for patient i, which introduces individual variability in baseline QoL. This was generated from a normal distribution with a mean of 0 and a standard deviation of 5.
- β_{time}: fixed at 0, implying that controls maintain a constant QoL over time.
- u_{i_time}: patient-specific random slope for time, representing individual variations in the QoL trajectory over time. This random effect was generated from a normal distribution with a mean of 0 and standard deviation of 0.1, allowing each patient to have a unique QoL trend over time.

- β_{trt}: fixed at 0, ensuring that treated and control patients have the same baseline QoL value (as expected in a RCT due to randomization).
- $\beta_{time*trt}$: indicates the linear change in the QoL of treated patients over time. Three different values were assigned to this parameter based on the simulated scenario. Specifically, $\beta_{time*trt}$ was set to 0 to simulate the case where the experimental treatment has no effect on QoL; it was set to -0.333 to hypothesize a worsening of QoL compared to the control group; and it was set to 0.333 to simulate an improvement in QoL over time associated with the treatment.
- ε_{ijk} : random error generated from a normal distribution with a mean of 0 and a standard deviation of 5.

Subsequently, the probability of experiencing the event (death) within each monthly interval for each i-th subject p_{jk} was estimated using the following formula:

$$p_{jk} = montly\_event\_rate * \exp[\beta_{QoL_k}(QoL_{ijk} - 50) + \beta_{trt} * trt_j]$$

Where:

- $montly\_event\_rate$: fixed at 0.005, represents the baseline monthly probability of event for the control group (trt=0), which approximates the monthly event rate.
- β_{trt}: represents the treatment effect, set to log(0.5) to simulate the case where the experimental treatment halves the risk of death compared to the control group, and set to 0 (i.e. log(1)) to simulate the case where the treatment has no effect on survival.
- β_{QoL_k}: set to log(0.96), assumes that for each increase of 1 point in QoL, the event hazard decreases by 4%. The effect of QoL is assumed to be the same for both groups.

For simplicity, in simulating the monthly event probability p_{jk}, individual patient variability was not included. In other words, it was assumed that the

event probability depends solely on the treatment trt_j and the QoL value at time k, without accounting for unobserved differences between individual patients. As a result, patients with the same treatment assignment and QoL score had the same probability of experiencing the event. This simplification was made to create a more manageable model and to focus on the combined effect of treatment and QoL, without introducing additional complexities related to individual variability.

It should be noted that the monthly mortality rate accounts for both the treatment effect and the current QoL (i.e., QoL_{ijk}, measured at the beginning of each month). A Bernoulli distribution was used with probability p_{jk} to determine whether the event (death) occurred within each considered monthly interval.

Step 2

Within each sample, for every subject who experienced the event, the first time interval in which the event occurred was selected, and subsequent intervals were excluded due to the subject's death. Varying observation times were simulated for each subject, with informative censoring introduced by modeling a lower probability of observation during intervals when subjects had lower QoL scores. Consequently, the probability of missing a QoL measurement was recalculated at each time interval based on the subject's current QoL. This probability varied with a base probability of 0.8, adjusted using a logistic regression model to account for differences in QoL. As a result of this approach, although each subject had the potential for up to 60 QoL observations over the 5-year follow-up, the median number of measurements per patient was 14 (IQR: 11–17), with all subjects having a baseline QoL measurement available.

Using this strategy, 6 different scenarios were simulated, each consisting of 500 samples, with each sample containing 1,500 patients: 750 assigned to the treatment group and 750 to the control group.

Table 5. Excerpt of the simulated dataset.

HR_trt	b_time_trt	Sample_ID	Unique_id	Trt	From	To	Event	Qol_obs
0.5	-0.333	1	1	0	0	14	1	48
0.5	-0.333	1	2	0	0	1	0	42
0.5	-0.333	1	2	0	1	2	0	34
0.5	-0.333	1	2	0	2	15	0	41
0.5	-0.333	1	2	0	15	21	0	40
0.5	-0.333	1	2	0	21	28	0	42
0.5	-0.333	1	2	0	28	43	0	27
…	…	…	…	…	…	…	…	…
0.5	-0.333	1	10013	1	0	1	0	56
0.5	-0.333	1	10013	1	1	13	0	59
0.5	-0.333	1	10013	1	13	34	1	56

Table 5 presents a portion of the simulated dataset, specifically for the first sample ("Sample_ID"=1) from the first simulated scenario, in which the treatment halves the event hazard (HR_trt=0.5) but worsens QoL (b_time_trt=-0.333). Each row represents a time interval for an individual patient, uniquely identified by the variable "Unique_ID." The variable "Trt" indicates whether the patient belongs to the treatment or control group. The "From" variable represents the start of the observation interval (in months). For all patients, the first observation begins at time 0; subsequently, "From" takes on the value corresponding to the moment when QoL was measured for each patient, marking the start of a new observation interval. The "To" variable indicates the end of the time interval, either coinciding with a follow-up timepoint (to record a new QoL measurement) or with the patient's exit from the study, either due to death (patients 1 and 10013) or censoring (patient 2). The "Event" variable distinguishes between these situations, indicating if the event of death occurred at the end of the time interval. For instance, patient 1 died after 14 months, while patient 2 was censored at 43 months. "Qol_obs" represents the QoL score observed at the beginning of each interval. In the case of patient 1, baseline QoL was 48, and no additional QoL measurements were recorded before death occurred at 14 months.

4.4 Statistical analysis methods

For each sample, the effect of treatment (HR trt) and the effect of QoL (HR QoL) on survival were estimated using the following statistical models:

- **Univariate Cox model**, in which only treatment was included as explanatory variable. Consequently, this model did not allow for estimating the effect of QoL on survival.
- **Extended Cox model with time-dependent covariate**, where, in addition to the treatment variable, QoL was included as a variable that changed during the patient follow-up.
- **Joint model**, as a combination of a survival model (to estimate HR trt) with a LMM (to model QoL over time). The longitudinal component included a random intercept only, to avoid overcomplicating the model and to assess whether this simpler structure could sufficiently capture the QoL variation over time. The Weibull-PH distribution was chosen, which parameterizes survival in terms of proportional hazard. This distribution allows for greater flexibility in modeling time-varying hazards, fitting better to the simulated scenarios where the effects of treatment and QoL interact with each other. The Gaussian Hermite (GH) method was used for numerical integration.[63]

To compare the models described above in terms of accuracy, for each scenario, the following was calculated:

- **Mean estimate of the parameter of interest** (HR trt and HR QoL) for each sample.

- **Bias**, defined as the mean difference between the log of the estimated HR in each sample s ($\widehat{ln\theta_s}$) and its true log HR value ($ln\theta$) set in the simulated data:

$$Bias = \frac{1}{500}\sum_{s=1}^{500}(\widehat{ln\theta_s} - ln\theta)$$

- **Standard error of estimates**, defined as the square root of the variance of the log HRs across the 500 samples:

$$SE = \sqrt{\frac{1}{499}\sum_{s=1}^{500}(\widehat{ln\theta_s} - \overline{ln\theta})^2}$$

where $\widehat{ln\theta_s}$ is the log of the estimated HR for sample s, and $\overline{ln\theta}$ is the average log HR across the 500 samples.

- **95% Coverage Probability (95% CP)**, defined as the percentage of times the 95% confidence interval of the estimates included the true value of HR:

$$95\%\ CP = \frac{1}{500}\sum_{s=1}^{500} I(\theta \in [LCL_s, UCL_s])$$

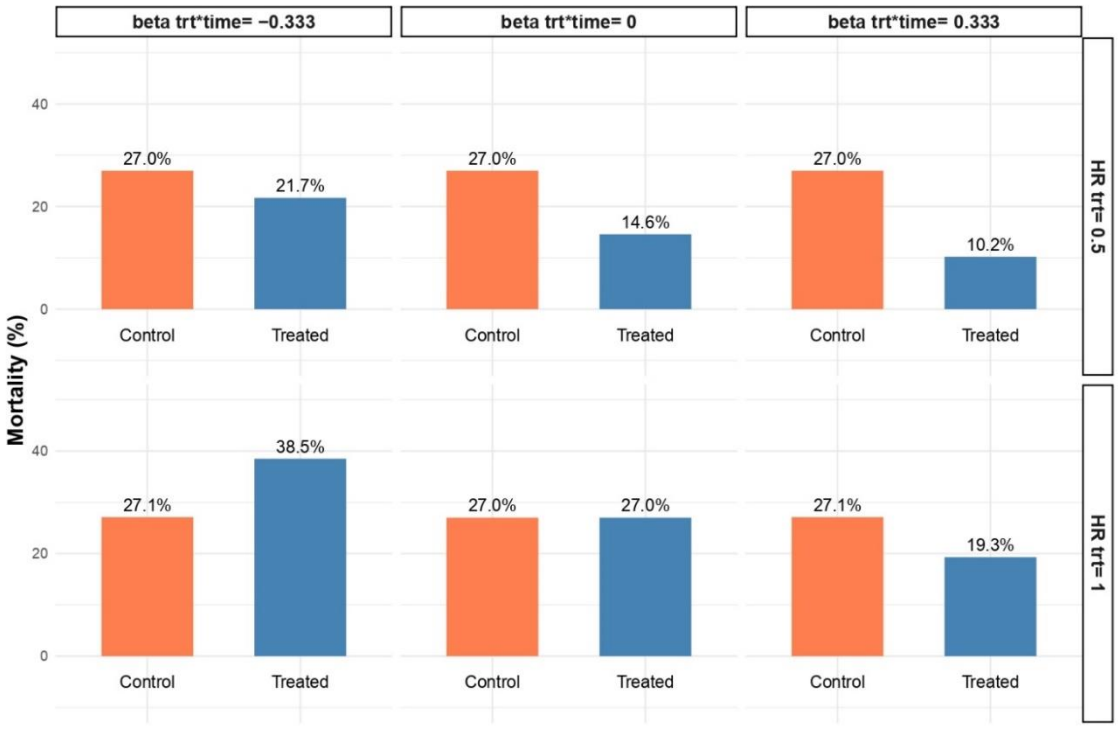
where I(·) is an indicator function that equals 1 if θ is within the interval defined by LCL_s (lower confidence limit) and UCL_s (upper confidence limit), and 0 otherwise.

The JM package by Rizopoulos[63] was used to fit joint models in R (version 4.3.1), while SAS (version 9.4) was employed for generating simulated data, fitting the Cox models, and conducting subsequent analyses.

4.5 Results

The simulated dataset includes 30,384,085 observations for treated patients and 28,467,716 for controls, divided across the 6 simulated scenarios, each containing 500 samples of 1,500 patients (750 treated and 750 controls). The median number of measurements per patient was 14 (IQR: 11–18) in the treated group and 13 (IQR: 11-16) in the control group. **Figure 9** displays the simulated mortality percentages within groups based on the parameters set for each scenario.

Figure 9. Mortality percentages by treatment group under the different scenarios.



As expected, when the treatment had no effect on either survival or QoL, the mortality percentage was equal in both groups (27%). Likewise, when the treatment halved the risk of death but had no effect on QoL, the mortality percentage was approximately halved in the treated group (14.6% vs. 27.0%). Interestingly, in the other 4 simulated scenarios where the treatment had either a positive or negative effect on QoL, it indirectly impacted survival. This occurred because, by definition of the joint model, the QoL score at each

interval influenced the probability of the event: values below the mean had a protective effect, whereas values above the mean increased risk.

In the 2 scenarios where treatment negatively impacted QoL, the mortality percentage in the treated group was higher than in the scenario with the same HR but with $\beta_{time*trt}$=0. When treatment had no effect on survival, the mortality in the treated group increased from 27% to 38.5%, due solely to treatment's effect on QoL, which, in turn, influenced survival. Even in the scenario where treatment had a protective effect against mortality (HR = 0.5) but negatively affected QoL, the mortality percentage in treated patients increased from 14.6% to 21.7%. This still showed a protective effect compared to the control group, though it was attenuated.

When treatment positively influenced QoL (causing it to improve over time) and QoL remained a protective factor against death, the observed effect on mortality percentage was amplified. Mortality in the treated group decreased to 19.3%, even when the treatment itself had no direct effect on survival. Mortality decreased further to 10.2% when the treatment had a positive effect on both outcomes, enhancing survival both directly and indirectly through its beneficial impact on QoL.

The joint model performed significantly better than the Cox models, especially in more complex situations where QoL played an important role in influencing survival. The main differences are highlighted in **Tables 6-8**, which show the estimates of the treatment effect (HR trt) and the effect of QoL (HR QoL) on survival in the 6 simulated scenarios, stratified by the type of model used for estimation.

Table 6. Estimation of treatment effect on survival (HR trt) across Cox models.

Parameter settings			Univariate Cox				Extended Cox			
HR trt	**HR QoL**	**β trt*time**	**HR trt estimate**	**Bias**	**SE**	**95% CP**	**HR trt estimate**	**Bias**	**SE**	**95% CP**
0.5	0.96	-0.333	0.85	0.525	0.106	0	0.74	0.381	0.107	0.104
0.5	0.96	0	0.50	-0.002	0.128	0.936	0.74	0.382	0.101	0.086
0.5	0.96	0.333	0.33	-0.416	0.133	0.112	1.12	0.799	0.127	0
1	0.96	-0.333	1.65	0.498	0.093	0	1.04	0.037	0.089	0.976
1	0.96	0	1.01	0.001	0.098	0.960	1.00	-0.004	0.076	0.992
1	0.96	0.333	0.67	-0.410	0.113	0.026	1.39	0.327	0.099	0.214

Abbreviations: CP, coverage probability; SE, mean standard error.

Table 7. Joint model estimation of treatment effect (HR_trt) on survival.

Parameter settings			Joint model			
HR trt	**HR QoL**	**β trt*time**	**HR trt estimate**	**Bias**	**SE**	**95% CP**
0.5	0.96	-0.333	0.50	-0.011	0.157	0.946
0.5	0.96	0	0.50	-0.012	0.130	0.922
0.5	0.96	0.333	0.54	0.055	0.163	0.948
1	0.96	-0.333	1.02	0.011	0.139	0.926
1	0.96	0	1.00	-0.005	0.100	0.960
1	0.96	0.333	1.08	0.063	0.139	0.940

Abbreviations: CP, coverage probability; SE, mean standard error.

Table 8. Comparison of the estimation of the QoL effect on survival (HR QoL).

Parameter settings			Extended Cox				Joint model			
HR trt	**HR QoL**	**β trt*time**	**HR QoL estimate**	**Bias**	**SE**	**95% CP**	**HR QoL estimate**	**Bias**	**SE**	**95% CP**
0.5	0.96	-0.333	1.03	0.071	0.005	0	0.96	-0.003	0.009	0.936
0.5	0.96	0	1.02	0.057	0.006	0	0.96	-0.002	0.010	0.964
0.5	0.96	0.333	1.01	0.045	0.007	0	0.95	-0.006	0.010	0.924
1	0.96	-0.333	1.03	0.068	0.004	0	0.96	0.0003	0.008	0.946
1	0.96	0	1.02	0.056	0.005	0	0.96	-0.002	0.010	0.950
1	0.96	0.333	1.01	0.047	0.006	0	0.95	-0.006	0.009	0.904

Abbreviations: CP, coverage probability; SD, standard deviation; SE, standard error.

When treatment had no impact on QoL (scenarios on rows 2 and 5 in the tables), the univariate Cox model accurately estimated the treatment effect (**Table 6**). The extended Cox model performed well in the scenario where the treatment had no effect on survival either, but introduced substantial bias in scenarios where the treatment halved the risk of death, underestimating its protective effect (HR=0.74 instead of HR=0.5). In contrast, the joint model demonstrated excellent performance in both cases, comparable with those of

the univariate Cox model (**Table 7**). Additionally, it provided reliable estimates for the protective effect of QoL on survival with minimal bias and a CP equal or greater than 95% (**Table 8**).

When the treatment worsened QoL but QoL was itself associated with greater survival (scenarios on rows 1 and 4), the Cox models revealed clear limitations (**Table 6**). Both models inaccurately estimated the treatment effect on survival in Scenario 1, making it appear less effective than it was. In Scenario 4, only the univariate model misinterpreted the effect, suggesting a potentially detrimental impact when there was none (HR=1.65 instead of HR=1). Conversely, the extended Cox model performed well in estimating the treatment effect in scenario 4: although it slightly overestimated the HR value (HR=1.04 instead of the true HR=1), its CP exceeded 95%. Nonetheless, in this same scenario, it entirely failed to capture the protective effect of QoL on survival (**Table 8**). Conversely, the joint model provided accurate estimates for both the treatment effect (HR_trt) and QoL effect (HR_QoL) on survival (**Table 7-8**), achieving a CP level greater than 93% in all cases.

In scenarios where the treatment improved QoL and QoL was linked to better survival (scenarios on rows 3 and 6 in the tables), further differences among models emerged. The univariate Cox model in both scenarios tended to overestimate the protective effect of the treatment on survival, making it appear more protective than it actually was in scenario 3 (HR=0.33 instead of HR=0.5), and falsely protective in scenario 6 when it had no true effect (HR=0.67 instead of HR=1). By contrast, the extended Cox model suggested a detrimental treatment effect in both scenarios, even when the treatment was actually protective (**Table 6**), showing substantial bias. Additionally, it failed to identify the protective influence of QoL on survival (**Table 8**). Once again, the joint model was markedly superior, although it showed a slightly higher bias in estimating the treatment effect in these scenarios (making it appear less protective than it was in scenario 3 and suggesting a detrimental effect in scenario 6 when no effect was present). The CP however, still was very high (**Table 7**). Regarding the impact of QoL on survival, the joint model maintained a high level of accuracy, with a CP above 90% (**Table 8**).

4.6 Discussion

Clinical studies, both randomized and not, frequently face significant challenges when it comes to analyzing PROs alongside survival data. A fundamental issue arises because PROs are assessed only for patients who are still alive at the time of questionnaire administration. This practice introduces bias in estimating treatment effects, as those with lower QoL are at a higher risk of mortality and thus may be underrepresented in the data. Conducting separate analyses of PROs and survival outcomes can compromise the integrity of the analyses, as it overlooks the interplay between these two critical dimensions of patient health. For instance, since patients experiencing deterioration in their QoL may not be captured in the analysis, this could lead to an inflated perception of the treatment's overall QoL benefits. Moreover, conventional survival models that rely solely on baseline PRO values neglect critical statistical and clinical insights, failing to account for the temporal dynamics of QoL changes over time. This oversight limits the understanding of how these changes relate to prognosis in the patient population under study.

The aim was to explore statistical models for the joint analysis of survival outcomes and QoL, driven by the need to identify suitable methodologies for analyzing data from the MITRADVANCE-HF trial. This multicenter RCT evaluates the effectiveness of percutaneous treatment for mitral regurgitation in patients with advanced HF and addresses both survival and QoL outcomes. Currently, the study is still in the recruitment phase and has not yet accumulated sufficient data for analysis. Therefore, simulated joint survival and QoL data were generated to explore various scenarios and test statistical methods that will be applicable once real data become available.

To comprehensively evaluate the effectiveness of joint model in analyzing survival and QoL data, a comparison was made with traditional statistical methods that included univariate Cox model and extended Cox model with time-dependent covariates. The univariate Cox model focused solely on treatment effects without accounting for changes in QoL. By contrast, the extended Cox model allowed for QoL to vary over time during patient follow-

up, offering a more dynamic view of how QoL impacts survival. However, this method still struggled to capture the inherent relationships between longitudinal QoL data and survival outcomes. The joint model, on the other hand, integrated the longitudinal component for QoL with the survival analysis, enabling a more nuanced understanding of how QoL changes influenced mortality risk. Key metrics such as mean estimates, mean bias, mean SE, and 95% CP were calculated across all the 6 different scenarios to compare model performances, ensuring robust analytical frameworks for the trial's eventual data analysis.

The simulated dataset, comprising over 50 million observations, provided valuable insights into the interconnection between QoL and survival across the various scenarios. The findings indicated that even a modest association between QoL and survival, exemplified by an HR QoL of 0.96, resulted in significant effects on mortality rates. When the treatment adversely affected QoL, a concerning increase in mortality rates was observed among treated patients, underscoring how detrimental changes in QoL could lead to poorer survival outcomes. Conversely, when treatment positively influenced QoL there was a significant reduction in mortality. This dual treatment impact (directly on survival and indirectly through QoL) highlighted the necessity of employing joint modeling approaches to effectively capture these relationships, even when the association between QoL and survival seems not particularly strong.

The joint model significantly outperformed both the univariate and extended Cox models, particularly in complex scenarios where QoL had a substantial influence on survival. The joint model's ability to integrate longitudinal QoL data with survival analysis proved crucial in accurately estimating both the treatment and the QoL effect on survival.

The univariate Cox model, while accurate in scenarios where treatment had no impact on QoL, tended to overestimate the protective effect of treatment on survival when it was positively associated with QoL, sometimes even indicating a benefit where none existed. Conversely, when treatment negatively affected QoL, the model either underestimated its protective effect

on survival or, in cases where there was no true effect, incorrectly suggested a harmful effect. This pattern highlighted the potential for misinterpretation in clinical settings, where changes in QoL might be mistakenly attributed to treatment effects. The extended Cox model showed some improvement over the univariate model in certain scenarios but introduced substantial bias, particularly when QoL effects were present. Moreover, it consistently failed to accurately estimate the protective effect of QoL on survival, as seen by its inability to capture the real HR QoL value across all scenarios. This limitation suggests that the model may lead to inaccurate conclusions, therefore its use in such settings should be discouraged. On the other hand, the joint model's estimates showed a reliable accuracy with a CP close to 95%, reflecting its robustness in varied scenarios. It not only accounted for the direct effects of treatment on survival but also captured the nuances of QoL's impact, even with a simplified approach that modeled QoL with only a random intercept. This suggests it is a better choice for analyzing complex clinical trial data, effectively balancing model complexity and interpretability.

In all simulated scenarios, the joint model produced a larger SE for the estimate compared to Cox models, likely due to its accounting for uncertainty from both the longitudinal and survival sub-models. However, some examples in the literature[62] has shown that joint models can yield smaller SEs. This can occur when the longitudinal trajectory is highly informative for survival, allowing the joint model to leverage this dependency and thereby reduce the variance of the treatment effect estimate. In these situations, the increased efficiency can enhance precision and potentially lower sample size requirements.

For the future analysis of data from the MITRADVANCE-HF study, the tested joint model appears to be well-suited, as it enables accurate estimation of treatment effects by accounting for both QoL and survival. This simulation study also highlights critical implications for clinical research and practice, especially in trials targeting chronic conditions like HF, where PROs play a crucial role. Integrating QoL measures with survival data in statistical analyses

can provide a more accurate and comprehensive understanding of treatment impacts.

4.4.1 Limitations

One significant limitation of this study is that the data utilized were simulated rather than derived from actual patient outcomes. While simulation allows for the examination of the performance of statistical models under controlled conditions and different scenarios, it may not fully capture the complexities and variability present in real-world data. The simulated dataset included random effects to account for individual variability in responses, which introduces an element of realism; however, it may still lack the inherent noise and unpredictable factors found in clinical populations. This could affect the generalizability of the results to actual clinical practice. Despite these limitations, the findings indicate that joint models appear to be a suitable approach for analyzing the interrelationships between QoL and survival in the context of the MITRADVANCE-HF study. They demonstrate a robust capacity to provide accurate estimates, highlighting their potential utility in clinical research and practice.

5 Conclusion

This project underscores the importance of advanced methodological approaches in clinical research to address the complexities introduced by PROs in both meta-analyses and clinical studies.

The first part of the project demonstrated that advanced meta-analytic techniques, such as pseudo IPD reconstruction and the one-stage LMM, enhance the accuracy of pooled estimates in longitudinal QoL data, enabling a clearer understanding of treatment effects over time. This approach allowed for the detection of subtle trends that may have been missed with AD, further enhancing the precision of the analysis. The use of methods to minimize information loss, either through graphical extraction or linear interpolation at specific timepoints, was crucial in ensuring that results were standardized across studies with varying follow-up intervals. These refinements ensure that healthcare providers and patients can gain clearer insights into both the magnitude and timing of QoL improvements, supporting more informed clinical decision-making. Therefore, employing such techniques is highly recommended to enhance the accuracy and clinical relevance of meta-analyses. Additionally, recent advancements in AI tools, such as ASReview, have demonstrated potential in optimizing the literature screening process, providing efficient alternatives to traditional methods. Employing AI-powered tools in the screening process is recommended to expedite systematic reviews and meta-analyses, enabling quicker access to high-quality findings and facilitating prompt clinical application of the latest evidence.

In parallel, the second part of the project reinforced the utility of the joint model for analyzing complex datasets that include both longitudinal QoL data and survival outcomes. The joint model demonstrated superior accuracy in estimating treatment effects and in capturing the protective influence of QoL on survival, even when simpler models showed substantial bias or failed to reflect these relationships. This accuracy, along with the consistently high CP, highlights the joint model's robustness, particularly in scenarios where QoL plays a significant role in patient outcomes. Its capacity to integrate multiple

dimensions of patient data makes it valuable for analyzing chronic conditions where PROs and survival are interdependent. The use of joint models is advisable even when the link between the longitudinal and survival components seems weak, as they can capture interdependencies that might otherwise be missed, potentially impacting outcomes. These results suggest that joint models should be prioritized in future studies, such as the MITRADVANCE-HF trial, to ensure that nuanced, multidimensional effects are accurately estimated.

Together, the findings from these two parts illustrate the value of integrating sophisticated analytical techniques for QoL data across different study designs. These methods not only strengthen evidence-based conclusions but also promote a more comprehensive understanding of treatment impacts on both patient QoL and survival, ultimately supporting a more holistic approach to patient care.

Appendix

Case report form of the KCCQ

The following questions refer to your heart failure and how it may affect your life. Please read and complete the following questions. There are no right or wrong answers. Please mark the answer that best applies to you.

1. **Heart failure** affects different people in different ways. Some feel shortness of breath while others feel fatigue. Please indicate how much you are limited by heart failure (shortness of breath or fatigue) in your ability to do the following activities over the past 2 weeks.

Place an **X** in one box on each line

Activity	**Extremely limited**	**Quite a bit limited**	**Moderately limited**	**Slightly limited**	**Not at all limited**	**Limited for other reasons or did not do the activity**
Dressing yourself	□	□	□	□	□	□
Showering/Bathing	□	□	□	□	□	□
Walking 1 block on level ground	□	□	□	□	□	□
Doing yardwork, housework or carrying groceries	□	□	□	□	□	□
Climbing a flight of stairs without stopping	□	□	□	□	□	□
Hurrying or jogging (as if to catch a bus)	□	□	□	□	□	□

2. Compared with 2 weeks ago, have your symptoms of **heart failure** (shortness of breath, fatigue or ankle swelling) changed?

My symptoms of heart failure have become...

Much worse	Slightly worse	Not changed	Slightly better	Much better	I've had no symptoms over the last 2 weeks
□	□	□	□	□	□

3. Over the past 2 weeks, how many times did you have **swelling** in your feet, ankles or legs when you woke up in the morning?

Every morning	3 or more times a week, but not every day	1-2 times a week	Less than once a week	Never over the past 2 weeks
□	□	□	□	□

4. Over the past 2 weeks, how much has **swelling** in your feet, ankles or legs bothered you?

It has been…

Extremely bothersome	Quite a bit bothersome	Moderately bothersome	Slightly bothersome	Not at all bothersome	I've had no swelling
□	□	□	□	□	□

5. Over the past 2 weeks, on average, how many times has **fatigue** limited your ability to do what you want?

All the time	Several times a day	At least once a day	3 or more times a week, but not every day	1-2 times a week	Less than once a week	Never over the past 2 weeks
□	□	□	□	□	□	□

6. Over the past 2 weeks, how much has your **fatigue** bothered you?

It has been…

Extremely bothersome	Quite a bit bothersome	Moderately bothersome	Slightly bothersome	Not at all bothersome	I've had no fatigue
□	□	□	□	□	□

7. Over the past 2 weeks, on average, how many times has **shortness of breath** limited your ability to do what you wanted?

All the time	Several times a day	At least once a day	3 or more times a week, but not every day	1-2 times a week	Less than once a week	Never over the past 2 weeks
□	□	□	□	□	□	□

8. Over the past 2 weeks, how much has your **shortness of breath** bothered you?

It has been…

Extremely bothersome	Quite a bit bothersome	Moderately bothersome	Slightly bothersome	Not at all bothersome	I've had no shortness of breath
□	□	□	□	□	□

9. Over the past 2 weeks, on average, how many times have you been forced to sleep sitting up in a chair or with at least 3 pillows to prop you up because of **shortness of breath**?

Every night	3 or more times a week, but not every day	1-2 times a week	Less than once a week	Never over the past 2 weeks
□	□	□	□	□

10. Heart failure symptoms can worsen for a number of reasons. How sure are you that you know what to do, or whom to call, if your **heart failure** gets worse?

Not at all sure	Not very sure	Somewhat sure	Mostly sure	Completely sure
□	□	□	□	□

11. How well do you understand what things you are able to do to keep your heart failure symptoms from getting worse? (for example, weighing yourself, eating a low salt diet, etc.)

Do not understand at all	Do not understand very well	Somewhat understand	Mostly understand	Completely understand
□	□	□	□	□

12. Over the past 2 weeks, how much has your **heart failure** limited your enjoyment of life?

It has extremely limited my enjoyment of life	It has limited my enjoyment of life quite a bit	It has moderately limited my enjoyment of life	It has slightly limited my enjoyment of life	It has not limited my enjoyment of life at all
□	□	□	□	□

13. If you had to spend the rest of your life with your **heart failure** the way it is right now, how would you feel about this?

Not at all satisfied	Mostly dissatisfied	Somewhat satisfied	Mostly satisfied	Completely satisfied
□	□	□	□	□

14. Over the past 2 weeks, how often have you felt discouraged or down in the dumps because of your **heart failure**?

I felt that way all the time	I felt that way most of the time	I occasionally felt that way	I rarely felt that way	I never felt that way
□	□	□	□	□

15. How much does your heart failure affect your lifestyle? Please indicate how your **heart failure** may have limited your participation in the following activities over the past 2 weeks.

Place an **X** in one box on each line

Activity	**Severely limited**	**Quite a bit limited**	**Moderately limited**	**Slightly limited**	**Not at all limited**	**Limited for other reasons or did not do the activity**
Hobbies, recreational activities	□	□	□	□	□	□
Working or doing household chores	□	□	□	□	□	□
Visiting family or friends out of your home	□	□	□	□	□	□
Intimate relationships with loved ones	□	□	□	□	□	□

List of abbreviations

6-MWT	6-minute walking test
AD	Aggregate data
AI	Artificial intelligence
ANCOVA	Analysis of covariance
CI	Confidence interval
CP	Coverage probability
EQ-5D	EuroQoL-5 dimensions
FDA	Food and Drug Administration
HF	Heart failure
HFmrEF	Heart failure with mildly reduced ejection fraction
HFpEF	Heart failure with preserved ejection fraction
HFrEF	Heart failure with reduced ejection fraction
HR	Hazard ratio
IPD	Individual patient data
IQR	Interquartile range
KCCQ	Kansas City cardiomyopathy questionnaire
KCCQ-CSS	Kansas City cardiomyopathy questionnaire – clinical summary score
KCCQ-OSS	Kansas City cardiomyopathy questionnaire – overall summary score
KCCQ-QoL	Kansas City cardiomyopathy questionnaire – quality of life score
KCCQ-TSS	Kansas City cardiomyopathy questionnaire – total symptoms score
LMM	Linear mixed effect model
LVEF	Left ventricular ejection fraction
MLHFQ	Minnesota living with heart failure questionnaire
NYHA	New York Heart Association
OMT	Optimal medical therapy
PRO	Patient's reported outcome
QoL	Quality of life
RCT	Randomized control trial
SD	Standard deviation
SE	Standard error
SGLT2i	Sodium-glucose co-transporter-2 inhibitors
SMR	Secondary mitral regurgitation
SOC	Standard of care

References

1. Rivera SC, Kyte DG, Aiyegbusi OL, Slade AL, McMullan C, Calvert MJ. The impact of patient-reported outcome (PRO) data from clinical trials: A systematic review and critical analysis. *Health Qual Life Outcomes*. 2019;17(1):1-19. doi:10.1186/s12955-019-1220-z
2. Balestroni G, Bertolotti G. EuroQol-5D (EQ-5D): an instrument for measuring quality of life. *Monaldi Arch Chest Dis*. 2012;78:155-159.
3. Tsao CW, Aday AW, Almarzooq ZI, et al. *Heart Disease and Stroke Statistics - 2023 Update: A Report from the American Heart Association*. Vol 147.; 2023. doi:10.1161/CIR.0000000000001123
4. Bilbao A, Escobar A, García-Perez L, Navarro G, Quirós R. The Minnesota living with heart failure questionnaire: Comparison of different factor structures. *Health Qual Life Outcomes*. 2016;14(1):1-11. doi:10.1186/s12955-016-0425-7
5. Stogios N, Fezza G, Wong J V., Ross HJ, Farkouh ME, Nolan RP. Current challenges for using the Kansas City Cardiomyopathy Questionnaire to obtain a standardized patient-reported health status outcome. *Eur J Heart Fail*. 2021;23(2):205-207. doi:10.1002/EJHF.2139
6. Miani D, Rozbowsky P, Gregori D, et al. The Kansas City Cardiomyopathy Questionnaire: Italian translation and validation. *Ital Hear J*. 2003;4(9):620-626.
7. Kularatna S, Byrnes J, Chan YK, Carrington MJ, Stewart S, Scuffham PA. Comparison of contemporaneous responses for EQ-5D-3L and Minnesota Living with Heart Failure; a case for disease specific multiattribute utility instrument in cardiovascular conditions. *Int J Cardiol*. 2017;227:172-176. doi:10.1016/j.ijcard.2016.11.030
8. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol*. 2000;35(5):1245-1255. doi:10.1016/S0735-1097(00)00531-3
9. Rector T. FDA Medical Device Development Tool (MDDT) Qualification

Package for the Minnesota Living with Heart Failure Questionnaire (MLHFQ) Thomas S . Rector , PhD Principal Proponent November 2017 Table of Contents. 2017;(November):40.

10. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol*. 2000;35(5):1245-1255. doi:10.1016/S0735-1097(00)00531-3
11. Spertus JA, Jones PG, Sandhu AT, Arnold S V. Interpreting the Kansas City Cardiomyopathy Questionnaire in Clinical Trials and Clinical Care: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2020;76(20):2379-2390. doi:10.1016/j.jacc.2020.09.542
12. Anker SD, Agewall S, Borggrefe M, et al. The importance of patient-reported outcomes: a call for their comprehensive integration in cardiovascular clinical trials. *Eur Heart J*. 2014;35(30):2001-2009. doi:10.1093/EURHEARTJ/EHU205
13. Nassif M, Fine JT, Dolan C, et al. Validation of the Kansas City Cardiomyopathy Questionnaire in Symptomatic Obstructive Hypertrophic Cardiomyopathy. *JACC Hear Fail*. 2022;10(8):531-539. doi:10.1016/J.JCHF.2022.03.002
14. Yee D, Novak E, Platts A, Nassif ME, LaRue SJ, Vader JM. Comparison of the Kansas City Cardiomyopathy Questionnaire and Minnesota Living With Heart Failure Questionnaire in Predicting Heart Failure Outcomes. *Am J Cardiol*. 2019;123(5):807-812. doi:10.1016/J.AMJCARD.2018.11.037
15. Gomes-Neto M, Durães AR, Conceição LSR, et al. Effect of Aerobic Exercise on Peak Oxygen Consumption, VE/VCO2 Slope, and Health-Related Quality of Life in Patients with Heart Failure with Preserved Left Ventricular Ejection Fraction: a Systematic Review and Meta-Analysis. *Curr Atheroscler Rep*. 2019;21(11). doi:10.1007/s11883-019-0806-6
16. Gomes-Neto M, Durães AR, Conceição LSR, et al. Effect of combined aerobic and resistance training on peak oxygen consumption, muscle strength and health-related quality of life in patients with heart failure

with reduced left ventricular ejection fraction: a systematic review and meta-analysis. *Int J Cardiol*. 2019;293:165-175. doi:10.1016/j.ijcard.2019.02.050

17. Kyriakou M, Middleton N, Ktisti S, Philippou K, Lambrinou E. Supportive Care Interventions to Promote Health-Related Quality of Life in Patients Living With Heart Failure: A Systematic Review and Meta-Analysis. *Hear Lung Circ*. 2020;29(11):1633-1647. doi:10.1016/j.hlc.2020.04.019
18. Turgeon RD, Barry AR, Hawkins NM, Ellis UM. Pharmacotherapy for heart failure with reduced ejection fraction and health-related quality of life: a systematic review and meta-analysis. *Eur J Heart Fail*. 2021;23(4):578-589. doi:10.1002/ejhf.2141
19. Spertus J V., Hatfield LA, Cohen DJ, et al. Integrating quality of life and survival outcomes in cardiovascular clinical trials: Results from the PARTNER trial. *Circ Cardiovasc Qual Outcomes*. 2019;12(6):1-8. doi:10.1161/CIRCOUTCOMES.118.005420
20. Braunwald E. Gliflozins in the Management of Cardiovascular Disease. *N Engl J Med*. 2022;386(21):2024-2034. doi:10.1056/NEJMRA2115011/SUPPL_FILE/NEJMRA2115011_DISCLOSURES.PDF
21. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failureDeveloped by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J*. 2021;42(36):3599-3726. doi:10.1093/EURHEARTJ/EHAB368
22. Heidenreich PA, Bozkurt B, Aguilar D, et al. *2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines*. Vol 145.; 2022. doi:10.1161/CIR.0000000000001063
23. Aguilar-Gallardo JS, Correa A, Contreras JP, Wiener Cardiovascular

Institute MA. Cardio-renal benefits of sodium-glucose co-transporter 2 inhibitors in heart failure with reduced ejection fraction: mechanisms and clinical evidence. *Eur Hear Journal-Cardiovascular Pharmacother*. 2022;8:311-321. doi:10.1093/ehjcvp/pvab056

24. Usman MS, Siddiqi TJ, Anker SD, et al. Effect of SGLT2 Inhibitors on Cardiovascular Outcomes Across Various Patient Populations. *J Am Coll Cardiol*. 2023;81(25):2377-2387. doi:10.1016/J.JACC.2023.04.034
25. Asgar AW, Mack MJ, Stone GW. Secondary mitral regurgitation in heart failure: pathophysiology, prognosis, and therapeutic considerations. *J Am Coll Cardiol*. 2015;65(12):1231-1248. doi:10.1016/J.JACC.2015.02.009
26. Oriecuia C, Tomasoni D, Sala I, et al. Sodium glucose co-transporter 2 inhibitors and quality of life in patients with heart failure: a comprehensive systematic review and meta-analysis of randomized controlled trials. *Eur Hear journal Cardiovasc Pharmacother*. Published online November 2023. doi:10.1093/ehjcvp/pvad088
27. Deeks JJ, Altman DG, Bradburn MJ. Statistical Methods for Examining Heterogeneity and Combining Results from Several Studies in Meta-Analysis. *Syst Rev Heal Care Meta-Analysis Context Second Ed*. Published online March 17, 2008:285-312. doi:10.1002/9780470693926.CH15
28. Uman LS. Systematic Reviews and Meta-Analyses. *J Can Acad Child Adolesc Psychiatry*. 2011;20(1):57. doi:10.1016/j.revmed.2014.05.011
29. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177-188. doi:10.1016/0197-2456(86)90046-2
30. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *J Chinese Integr Med*. 2009;7(9):889-896. doi:10.3736/jcim20090918
31. Coens C, Pe M, Dueck AC, et al. International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *Lancet Oncol*. 2020;21(2):e83-e96. doi:10.1016/S1470-2045(19)30790-9

32. van de Schoot R, de Bruin J, Schram R, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell 2021 32*. 2021;3(2):125-133. doi:10.1038/S42256-020-00287-7
33. ASReview LAB - A tool for AI-assisted systematic reviews. doi:10.5281/ZENODO.13629181
34. Higgins JP, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343. doi:10.1136/bmj.d5928
35. Drevon D, Fursa SR, Malcolm AL. Intercoder Reliability and Validity of WebPlotDigitizer in Extracting Graphed Data. *Behav Modif*. 2017;41(2):323-339. doi:10.1177/0145445516673998
36. Papadimitropoulou K, Stijnen T, Riley RD, Dekkers OM, le Cessie S. Meta-analysis of continuous outcomes: Using pseudo IPD created from aggregate data to adjust for baseline imbalance and assess treatment-by-baseline modification. *Res Synth Methods*. 2020;11(6):780-794. doi:10.1002/jrsm.1434
37. Spertus JA, Birmingham MC, Nassif M, et al. The SGLT2 inhibitor canagliflozin in heart failure: the CHIEF-HF remote, patient-centered randomized trial. *Nat Med*. 2022;28(4):809-813. doi:10.1038/s41591-022-01703-8
38. Kosiborod MN, Jhund PS, Docherty KF, et al. Effects of Dapagliflozin on Symptoms, Function, and Quality of Life in Patients With Heart Failure and Reduced Ejection Fraction. *Circulation*. 2020;141(2):90-99. doi:10.1161/CIRCULATIONAHA.119.044138
39. Nassif ME, Windsor S, Tang F, et al. Dapagliflozin Effects on Biomarkers, Symptoms, and Functional Status in Patients With Heart Failure With Reduced Ejection Fraction. *Circulation*. 2019;140(18). doi:10.1161/CIRCULATIONAHA.119.042929
40. Kosiborod MN, Bhatt AS, Claggett BL, et al. Effect of Dapagliflozin on Health Status in Patients With Preserved or Mildly Reduced Ejection Fraction. Published online 2023. doi:10.1016/j.jacc.2022.11.006

41. Nassif ME, Qintar M, Windsor SL, et al. Empagliflozin Effects on Pulmonary Artery Pressure in Patients With Heart Failure. *Circulation*. 2021;143(17):1673-1686. doi:10.1161/CIRCULATIONAHA.120.052503
42. Santos-Gallego CG, Vargas-Delgado AP, Requena-Ibanez JA, et al. Randomized Trial of Empagliflozin in Nondiabetic Patients With Heart Failure and Reduced Ejection Fraction. *J Am Coll Cardiol*. 2021;77(3):243-255. doi:10.1016/j.jacc.2020.11.008
43. Abraham WT, Lindenfeld JA, Ponikowski P, et al. Effect of empagliflozin on exercise ability and symptoms in heart failure patients with reduced and preserved ejection fraction, with and without type 2 diabetes. *Eur Heart J*. 2021;42(6):700-710. doi:10.1093/EURHEARTJ/EHAA943
44. Butler J, Filippatos G, Jamal Siddiqi T, et al. Empagliflozin, Health Status, and Quality of Life in Patients With Heart Failure and Preserved Ejection Fraction: The EMPEROR-Preserved Trial. *Circulation*. 2022;145(3):184-193. doi:10.1161/CIRCULATIONAHA.121.057812
45. Butler J, Anker SD, Filippatos G, et al. Empagliflozin and health-related quality of life outcomes in patients with heart failure with reduced ejection fraction: the EMPEROR-Reduced trial. *Eur Heart J*. 2021;42(13):1203-1212. doi:10.1093/EURHEARTJ/EHAA1007
46. Jensen J, Omar M, Kistorp C, et al. Twelve weeks of treatment with empagliflozin in patients with heart failure and reduced ejection fraction: A double-blinded, randomized, and placebo-controlled trial. *Am Heart J*. 2020;228:47-56. doi:10.1016/J.AHJ.2020.07.011
47. Kosiborod MN, Angermann CE, Collins SP, et al. Effects of Empagliflozin on Symptoms, Physical Limitations, and Quality of Life in Patients Hospitalized for Acute Heart Failure: Results From the EMPULSE Trial. *Circulation*. 2022;146(4):279-288. doi:10.1161/CIRCULATIONAHA.122.059725
48. Nassif ME, Windsor SL, Borlaug BA, et al. The SGLT2 inhibitor dapagliflozin in heart failure with preserved ejection fraction: a multicenter randomized trial. *Nat Med 2021 2711*. 2021;27(11):1954-

1960. doi:10.1038/s41591-021-01536-x

49. Bhatt DL, Szarek M, Steg PG, et al. Sotagliflozin in Patients with Diabetes and Recent Worsening Heart Failure. *N Engl J Med*. 2021;384(2):117-128.doi:10.1056/NEJMOA2030183/SUPPL_FILE/NEJMOA2030183_DATA-SHARING.PDF

50. Vaduganathan M, Docherty KF, Claggett BL, et al. SGLT-2 inhibitors in patients with heart failure: a comprehensive meta-analysis of five randomised controlled trials. *Lancet*. 2022;400(10354):757-767. doi:10.1016/S0140-6736(22)01429-5

51. Kosiborod MN, Bhatt AS, Claggett BL, et al. Effect of Dapagliflozin on Health Status in Patients With Preserved or Mildly Reduced Ejection Fraction. *J Am Coll Cardiol*. 2023;81(5):460-473. doi:10.1016/j.jacc.2022.11.006

52. Bhatt AS, Kosiborod MN, Vaduganathan M, et al. Effect of dapagliflozin on health status and quality of life across the spectrum of ejection fraction: Participant-level pooled analysis from the DAPA-HF and DELIVER trials. *Eur J Heart Fail*. Published online 2023. doi:10.1002/EJHF.2909

53. Vaduganathan M, Claggett BL, John ;, Mcmurray J V, Solomon SD. Health Status Trajectories Before and After Hospitalization for Heart Failure. *Circulation*. 1872;145:1872-1874. doi:10.1161/CIRCULATIONAHA.122.059282

54. Tomasoni D, Fonarow GC, Adamo M, et al. Sodium–glucose co-transporter 2 inhibitors as an early, first-line therapy in patients with heart failure and reduced ejection fraction. *Eur J Heart Fail*. 2022;24(3):431-441. doi:10.1002/EJHF.2397

55. Metra M, Adamo M, Tomasoni D, et al. Pre-discharge and early post-discharge management of patients hospitalized for acute heart failure: A scientific statement by the Heart Failure Association of the ESC. *Eur J Heart Fail*. 2023;25(7):1115-1131. doi:10.1002/EJHF.2888

56. Stogios N, Fezza G, Wong J V., Ross HJ, Farkouh ME, Nolan RP. Current challenges for using the Kansas City Cardiomyopathy Questionnaire to

obtain a standardized patient-reported health status outcome. *Eur J Heart Fail*. 2021;23(2):205-207. doi:10.1002/ejhf.2139

57. Butler J, Khan MS, Mori C, et al. Minimal clinically important difference in quality of life scores for patients with heart failure and reduced ejection fraction. *Eur J Heart Fail*. 2020;22(6):999-1005. doi:10.1002/EJHF.1810
58. McDonagh TA, Metra M, Adamo M, et al. 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2023;44(37). doi:10.1093/EURHEARTJ/EHAD195
59. Bayés-Genís A, Aimo A, Metra M, et al. Head-to-head comparison between recommendations by the ESC and ACC/AHA/HFSA heart failure guidelines. *Eur J Heart Fail*. 2022;24(6):916-926. doi:10.1002/EJHF.2542
60. Tomasoni D, Adamo M, Bozkurt B, et al. Aiming at harmony. Comparing and contrasting International HFrEF Guidelines. *Eur Heart J Suppl*. 2022;24(Suppl L):L20. doi:10.1093/EURHEARTJSUPP/SUAC124
61. Rizopoulos Dimitris. - *Joint Models for Longitudinal and Time-to-Event Data. With Applications in R*.; 2012. doi:10.1201/b19109-17
62. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol*. 2010;28(16):2796-2801. doi:10.1200/JCO.2009.25.0654
63. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw*. 2010;35(9):1-33. doi:10.18637/jss.v035.i09