

This is a “preproof” accepted article for *Journal of Clinical and Translational Science*.

This version may be subject to change during the production process.

10.1017/cts.2024.492

Trends in Use of Composite Endpoints in Clinical Trials: A Comparison between Acute Heart Failure Trials and COVID-19 Trials

Lan Shi MS¹, Christopher John Lindsell PhD², Dandan Liu PhD¹

¹Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee

²Department of Biostatistics & Bioinformatics, Duke University, Durham, North Carolina

Address for Correspondence: Dandan Liu, PhD, 2525 West End N, Suite 1100, Nashville, TN 37203

Running Title: Trend in Use of Composite Endpoints in Clinical Trials

Target Journal: Journal of Clinical Translational Science

Themed Issue: Innovations for Enhancing the Informativeness and Quality of Clinical Trials

Funding Acknowledgement:

This study was supported by National Center for Advancing Translational Sciences (NCATS) Clinical Translational Science Award (CTSA) Program, Award Number UL1TR002243, and NCATS Center for Innovative TRIals in ChilDrEN and AdulTs (TRIDENT) Award U24TR001608. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCATS and NIH.

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Abstract:

Composite endpoints can encode multiple pieces of information and are increasingly adopted in clinical trials. Advocacy for using composite endpoints began decades ago in cardiovascular trials, leading to incorporation of patient-oriented outcomes and consideration of a hierarchical ranking system. The use of composite endpoints in COVID-19 trials has evolved similarly. We conducted a literature review to investigate the use of composite endpoints in acute heart failure and COVID-19 clinical trials. The results showed more frequent use of patient-oriented outcomes and ordinal composite endpoints in COVID-19 trials, which might be driven by global consensus on a set of common outcome measures. **INTRODUCTION**

There rarely exists a single measure that encodes sufficient information for efficient clinical and statistical evaluation of efficacy in clinical trials [1]. By encoding multiple pieces of information into single variables, composite endpoints are appealing as they minimize sample size needs, shorten follow-up durations, and cut costs [1–3]. Composite endpoints are increasingly adopted in clinical trials, especially in diseases with complex presentations like acute heart failure (AHF) or coronavirus disease (COVID-19).

The evolution of composite endpoints is mainly driven to increase the ability of clinical trials to detect treatment effects across the full spectrum of disease. Decades ago, composite endpoints were developed to better inform the impact of innovative clinical therapies in randomized clinical trials of AHF [4] [5]. As many as half of cardiovascular trials have adopted a composite endpoint inclusive of morbidity and mortality [2]. Advances in cardiovascular medicine have led to a decline in morbidity and mortality and the subsequent inclusion of nonfatal events like hospitalization in defining a time-to-first event primary composite endpoints [6]. The commonly used analytical approach is under the survival analysis framework [7–9]. Meanwhile, a hierarchical clinical composite evaluated at a fixed time point was developed as an ordinal outcome in chronic heart failure trials [10], and later adapted to AHF trials by incorporating the occurrence of worsening clinical events beyond a fixed time point to fully evaluate the clinical course of patients [11]. Recent FDA guidance clarifies that therapies for treating heart failure can be approved based on their effect on symptoms or physical function, even if they fail to show a favorable effect on survival or hospitalization risk [12]. This further shifts the emphasis from clinical endpoints to more patient-oriented outcomes in AHF clinical trials [5]. Such a shift

naturally motivates advanced statistical analysis methods. For example, the win ratio was proposed to handle the inherent limitation of using time-to-event composite endpoints, where each patient's first event is emphasized over its clinical importance [13]. Another example is the global ranking approach that, similar to hierarchical clinical endpoints, ranks patients using varying aspects of the clinical course based on a prespecified hierarchical ranking system [14] [15]. Regardless of approach, there is a clear trend to studying AHF treatments using composite endpoints as mortality rates decline, and improving how a patient feels and functions becomes the primary motivator.

The evolution of composite endpoints for COVID-19 appears to follow a very similar pattern, though over a much shorter course. Early COVID-19 clinical trials primarily focused on mortality and serious clinical events. After generations of mutation, the severity of COVID-19 has decreased such that clinical trials focused on mortality and serious clinical events are less feasible. Trials testing effectiveness in reducing symptoms, improving patient quality of life and preventing long-COVID have been critical. Morbidity and mortality remain important, leading to endpoints that combine clinical outcomes with patient-reported outcomes (PROs). PROs are defined as “any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else”; and can encompass symptoms, quality of life, and more [11]. The World Health Organization (WHO) COVID-19 ordinal scale for clinical improvement (OSCI) [17] is a clear example that incorporates both clinical endpoints and PROs for consideration as an efficacy endpoint in COVID-19 clinical trials [18].

Given the similar evolution of composite endpoints from an exclusive focus on clinical events to a multi-dimensional evaluation incorporating PROs in both AHF and COVID-19 clinical trials, we conducted a literature review to investigate and compare the adoption of composite endpoints. To the best of our knowledge, this is the first review that conducts annotations of composite endpoints regarding their composition and statistical types for both AHF and COVID-19 trials and explores the trends of their use in an analytical fashion. Our study combined with the observed evolution will shed light on the potential facilitators and barriers of composite endpoints uptake as well as the potential for composite outcomes in other areas.

MATERIALS AND METHODS

Literature Search Strategy

PubMed was searched to identify eligible trials. The final search was conducted on May 4, 2023. No date restrictions were placed on the search. The primary search strategy involved threefold identification pinpointing clinical trials, composite endpoints, and disease types. To identify clinical trials, the checkbox “Clinical Trials” in the PubMed filter of article types was used. To identify AHF and COVID-19 clinical trials using composite endpoints, titles and abstracts were searched for terms implying the use of composite endpoints (“composite endpoint” OR “composite” OR “multiple endpoint” OR “ordinal” OR “-free” OR “win ratio” OR “global ranking”) AND terms indicating disease types (“acute heart failure” OR “COVID-19”). Extra terms “ordinal”, “-free” (e.g., hospital-free or ventilator-free days), “win ratio” and “global ranking” were added to not miss studies using composite endpoints but not explicitly stating that the endpoint was a composite. To avoid missing clinical trial papers not marked as “Clinical Trials” in PubMed and to capture AHF papers not specifically mentioning “acute heart failure”, a supplementary search was conducted not using the “Clinical Trials” checkbox but including “clinical trial” and removing disease types in the keyword search.

Annotation Strategy

Papers were included if they were presenting primary results evaluating efficacy from AHF or COVID-19 clinical trials with at least one composite endpoint as either the primary or secondary outcome. Each publication was independently annotated by two authors (Liu, Faculty; and Shi, Research Assistant), recording exclusion reason, year of publication, number of composite endpoints, statistical type (time-to-event, binary, ordinal, count, continuous, or mixed; mixed mainly denotes the use of win ratio and global ranking approaches), composition type (clinical-only, PRO-only, or both), and number of components in the primary composite endpoint. Annotations were then compared and reconciled. Conflicts were resolved by consensus. If a resolution could not be reached, another faculty member would be involved to make the final decision.

Data Analysis

Descriptive statistics were calculated to characterize the overall use of composite endpoints for AHF and COVID-19 trials. Wilcoxon Rank Sum test or Fisher's Exact test was used for two group comparisons. The proportions of statistical types and compositions of the primary endpoints are plotted by publication year for each clinical field.

RESULTS

Our search identified 946 publications, 419 from the primary search and 527 from the supplementary search. Of these, 227 met inclusion criteria with 46 AHF trials and 181 COVID-19 trials (Figure 1). There are clear differences in the use of composite endpoints between AHF and COVID-19 trials (Table 1). COVID-19 trials were more likely to include multiple composite endpoints than AHF trials (43.5% in AHF vs 59.1% in COVID-19, $p < 0.001$). Among trials where the primary outcome was a composite endpoint, COVID-19 trials were more likely to include PROs in the primary composite endpoints (PRO-only + both: 7% in AHF vs 35.1% in COVID-19, $p < 0.001$), although the number of components in the primary endpoint were similar between diseases ($p = 0.73$). AHF trials predominantly used time-to-event composite endpoints (62.8% in AHF vs 26% in COVID-19), whereas COVID-19 trials were much more likely to use ordinal composite endpoints (2.3% in AHF vs 24.7% in COVID-19) with more diversity in statistical types (26% using time-to-event, 37.7% using binary, and 9.7% using continuous). The use of win ratio and global ranking approaches (i.e., mixed statistical type) in trials of both diseases is low (4.7% in AHF vs 1.9% in COVID-19).

The composition and statistical type of the primary composite endpoints are shown by publication year in Figure 2. For AHF trials, the combined use of clinical outcomes and PROs began to appear after 2011, yet remain used in less than 10% of studies (Figure 2a). In comparison, composites with "both" composition type were heavily used throughout the COVID-19 pandemic, where the proportion slightly dropped from 38.5% in 2020 to 26.5% in 2022, then rose again to 55.6% in 2023. Time-to-event analysis approaches consistently dominate in AHF (Figure 2b). The use of a mixed approach to analysis appeared after 2016, with 1 trial in 2016-2019 and another post-2019. In contrast, the diversity of statistical types remains consistent over time in COVID-19 trials where ordinal, time-to-event, and binary approaches were most frequent in COVID-19 trials.

DISCUSSION

This literature review compared trends in the use, construction, and analysis of composite endpoints in clinical trials for two highly disparate clinical fields. AHF was chosen as one because there is increasing advocacy for using composite endpoints over the past two decades [4,5,10]. This was driven partly by the lack of success in short-term pharmacological therapy since the 1970s, and that multiple domains need to be assessed for safety and efficacy [19]. COVID-19 was chosen for comparison due to its similar evolution patterns, though over a much shorter course of four years. Even though we are still early to observe the use of composite endpoints in COVID-19 trials, this field has caught up with AHF trials in the shift from the exclusive focus on clinical events to multi-dimensional evaluation by additionally incorporating PROs. Our findings suggest that it is critical to understand the current stage of disease management when picking endpoints; as disease transitions from being fatal to treatable, there is a need to measure change in outcomes over the full disease severity, reflecting not just survival but also how the patient functions and feels. Among trials where the primary outcome is a composite endpoint, COVID-19 trials were more likely to include PROs and use ordinal composite endpoints than AHF trials.

The well-recognized benefits of using composite endpoints are accompanied by challenges in their use, partly reflected in our results. While time-to-event composite endpoints only consider time-to-the-first clinical events and ignore their importance, ordinal composite endpoints can be difficult to interpret. Methods such as the win ratio and global ranking rely on ranking patients against one another within the trial, and clinical effect sizes are difficult to extract. Although advocacy for including PROs and using a prespecified hierarchical ranking system for composite endpoints occurred early in AHF trials [4,5,8,11], we posit the lack of global consensus impedes progress and explains the dominant use of clinical and time-to-event composite endpoints in AHF trials. In contrast, during the rapidly evolving COVID-19 outbreak multiple international organizations including the WHO developed common outcome measures for COVID-19 clinical research [17,20], and guidance appeared early for construction of new endpoints. Many COVID-19 trials either used or refined the WHO ordinal composite endpoint, although dichotomizing the scale to improve interpretation appears to have remained common.

There are some limitations to this study. This was not intended as a systematic review but an exploration of trends in the use of composite outcomes in two disease entities – one with a long history of clinical trials and the other an infectious disease causing a global pandemic. This study only searches for clinical trials with composite endpoints, not all clinical trials in AHF and COVID-19, thus we are not able to assess the overall proportion of studies adopting composite endpoints in these clinical fields. The number of studies, especially for AHF, were smaller than expected despite an additional supplementary search. This may be because that many published trials did not explicitly state the use of composite endpoints and were missed in our search. Many papers provided only a partial list of secondary composite endpoints, and some papers put the full list in a separate supplemental material precluding comprehensive annotation. There was no clear consensus on the choice of one secondary endpoint to annotate for composition and statistical type. Therefore, we only annotated the number of composite endpoints including secondary endpoints. Lastly, the overtime comparison on the use of composite endpoints was purely descriptive.

In summary, for AHF and COVID-19 trials, the use of composite endpoints has evolved to include PROs as well as clinical events, and their use commonly ranks multiple events. The change corresponds to the change in purpose of the trials from preventing mortality, then preventing progression, to improving quality of life. Achieving consensus on common outcome measurements and hierarchical rankings is expected to accelerate uptake.

DISCLOSURES

Dr. Lindsell reports receiving grants to the institution from the National Center for Advancing Translational Sciences (NCATS) for the submitted work; grants to the institution from NIH and Department of Defense and contracts to the institution from the CDC, bioMerieux, AstraZeneca, AbbVie, Entegriion Inc, and Endpoint Health outside the submitted work; patents for risk stratification in sepsis and septic shock issued to Cincinnati Children's Hospital Medical Center; service on DSMBs unrelated to the current work; and stock options in Bioscape Digital unrelated to the current work. Other authors have no conflicts of interest to declare.

REFERENCES

1. **Sankoh AJ, Li H, D'Agostino Sr RB.** Use of composite endpoints in clinical trials. *Statistics in Medicine* 2014; **33**(27): 4709–4714.
2. **Ferreira-González I, Busse JW, Heels-Ansdell D, et al.** Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ (Clinical research ed.)* 2007; **334**(7597): 786.
3. **Ferreira-González I, Alonso-Coello P, Solà I, et al.** Composite Endpoints in Clinical Trials. *Revista Española de Cardiología (English Edition)* 2008; **61**(3): 283–290.
4. **Califf RM, Harrelson-Woodlief L, Topol EJ.** Left ventricular ejection fraction may not be useful as an end point of thrombolytic therapy comparative trials. *Circulation* American Heart Association, 1990; **82**(5): 1847–1853.
5. **Braunwald E, Cannon CP, McCabe CH.** Use of composite endpoints in thrombolysis trials of acute myocardial infarction. *The American Journal of Cardiology* 1993; **72**(19): G3–G12.
6. **Hussain A, Misra A, Bozkurt B.** Endpoints in Heart Failure Drug Development. *Cardiac Failure Review* 2022; **8**: e01.
7. **Fox K, Ford I, Steg PG, Tendera M, Ferrari R, BEAUTIFUL Investigators.** Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. *Lancet (London, England)* 2008; **372**(9641): 807–816.
8. **Boden WE, van Gilst WH, Scheldewaert RG, et al.** Diltiazem in acute myocardial infarction treated with thrombolytic agents: a randomised placebo-controlled trial. Incomplete Infarction Trial of European Research Collaborators Evaluating Prognosis post-Thrombolysis (INTERCEPT). *Lancet (London, England)* 2000; **355**(9217): 1751–1756.
9. **Brugaletta S, Gomez-Lara J, Ortega-Paz L, et al.** 10-Year Follow-Up of Patients With Everolimus-Eluting Versus Bare-Metal Stents After ST-Segment Elevation Myocardial Infarction. *Journal of the American College of Cardiology* 2021; **77**(9): 1165–1178.

10. **Packer M.** Proposal for a new clinical end point to evaluate the efficacy of drugs and devices in the treatment of chronic heart failure. *Journal of Cardiac Failure* 2001; **7**(2): 176–182.
11. **Packer M.** Development and Evolution of a Hierarchical Clinical Composite End Point for the Evaluation of Drugs and Devices for Acute and Chronic Heart Failure: A 20-Year Perspective. *Circulation* 2016; **134**(21): 1664–1678.
12. **Food and Drug Administration.** *Treatment for Heart Failure: Endpoints for Drug Development Guidance for Industry.* FDA, 2020 [cited October 4, 2023]. (<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/treatment-heart-failure-endpoints-drug-development-guidance-industry>).
13. **Pocock SJ, Ariti CA, Collier TJ, Wang D.** The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *European Heart Journal* 2012; **33**(2): 176–182.
14. **O’Brien PC.** Procedures for comparing samples with multiple endpoints. *Biometrics* 1984; **40**(4): 1079–1087.
15. **Felker GM, Anstrom KJ, Rogers JG.** A global ranking approach to end points in trials of mechanical circulatory support devices. *Journal of Cardiac Failure* 2008; **14**(5): 368–372.
16. **U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health.** Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health and Quality of Life Outcomes* 2006; **4**(1): 79.
17. **WHO R&D Blueprint novel Coronavirus COVID-19 Therapeutic Trial Synopsis.** [cited October 4, 2023]. (https://www.who.int/docs/default-source/blue-print/covid-19-therapeutic-trial-synopsis.pdf?sfvrsn=44b83344_1&download=true).

18. **Thorlund K, Smith D, Linsell C, *et al.*** The importance of appropriate selection of clinical endpoints in outpatient COVID-19 clinical trials. *Communications Medicine* Nature Publishing Group, 2023; **3**(1): 1–5.
19. **Felker GM, Pang PS, Adams KF, *et al.*** Clinical Trials of Pharmacological Therapies in Acute Heart Failure Syndromes. *Circulation: Heart Failure* American Heart Association, 2010; **3**(2): 314–325.
20. **WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection.** A minimal common outcome measure set for COVID-19 clinical research. *The Lancet. Infectious Diseases* 2020; **20**(8): e192–e197.

Table 1. Descriptive statistics of the use of composite endpoints in Acute Heart Failure (AHF) and COVID-19 trials. In the table, "Q1", "Q3", "min", "max", "cat.", "No.", and "PRO" are abbreviations for the first and third quartiles, minimum, maximum, "categorical", "the number of", and "patient-reported outcome".

	AHF (N=46)	COVID-19 (N=181)	p-value*
Publication Year			< 0.001
median (Q1, Q3)	2016 (2012, 2020)	2022 (2021, 2022)	
min ; max	2000 - 2023	2020 - 2023	
Publication Year (cat.)^{†‡}			
≤2010	8 (17.4%)	0 (0.0%)	
2011-2015	11 (23.9%)	0 (0.0%)	
2016-2019	15 (32.6%)	0 (0.0%)	
≥2020	12 (26.1%)	181 (100.0%)	
2020	2 (4.3%)	15 (8.3%)	
2021	5 (10.9%)	63 (34.8%)	
2022	3 (6.5%)	80 (44.2%)	
2023	2 (4.3%)	23 (12.7%)	
No. Composite Endpoints			
1	26 (56.5%)	74 (40.9%)	< 0.001
2	16 (34.8%)	42 (23.2%)	
>2	4 (8.7%)	65 (35.9%)	
Primary is Composite			
no	3 (6.5%)	27 (14.9%)	0.152
yes	43 (93.5%)	154 (85.1%)	
Primary Composition Type[§]			
clinical-only	40 (93.0%)	100 (64.9%)	< 0.001
PRO-only	0 (0.0%)	5 (3.2%)	

	AHF (N=46)	COVID-19 (N=181)	p-value*
both	3 (7.0%)	49 (31.8%)	
Primary Statistical Type[§]			
time-to-event	27 (62.8%)	40 (26.0%)	< 0.001
binary	10 (23.3%)	58 (37.7%)	
ordinal	1 (2.3%)	38 (24.7%)	
count	2 (4.7%)	0 (0.0%)	
continuous	1 (2.3%)	15 (9.7%)	
mixed	2 (4.7%)	3 (1.9%)	
Primary No. Components			0.727
median (Q1, Q3)	3 (2, 4)	3 (2, 7)	
min ; max	2 - 10	2 - 12	

* The p-values are based on the Wilcoxon Rank Sum test for continuous variables and Fisher's Exact test for categorical variables.

† For continuous variable "Publication Year", categorized variables was derived as "Publication Year (cat.)" to provide more detailed information about the distribution. Note, this additional categorical variable would not be used for hypothesis testing.

‡ The categories of publication years of AHF trials were derived based on quartiles; while for COVID-19, all eligible papers were published in or after 2020, so the number of trials for each year of 2020-2023 was listed.

§ For proportions related to the attributes of primary composite endpoints, the denominators are the total numbers of papers with primary endpoints being composite, which were summarized in "Primary is Composite - Yes".

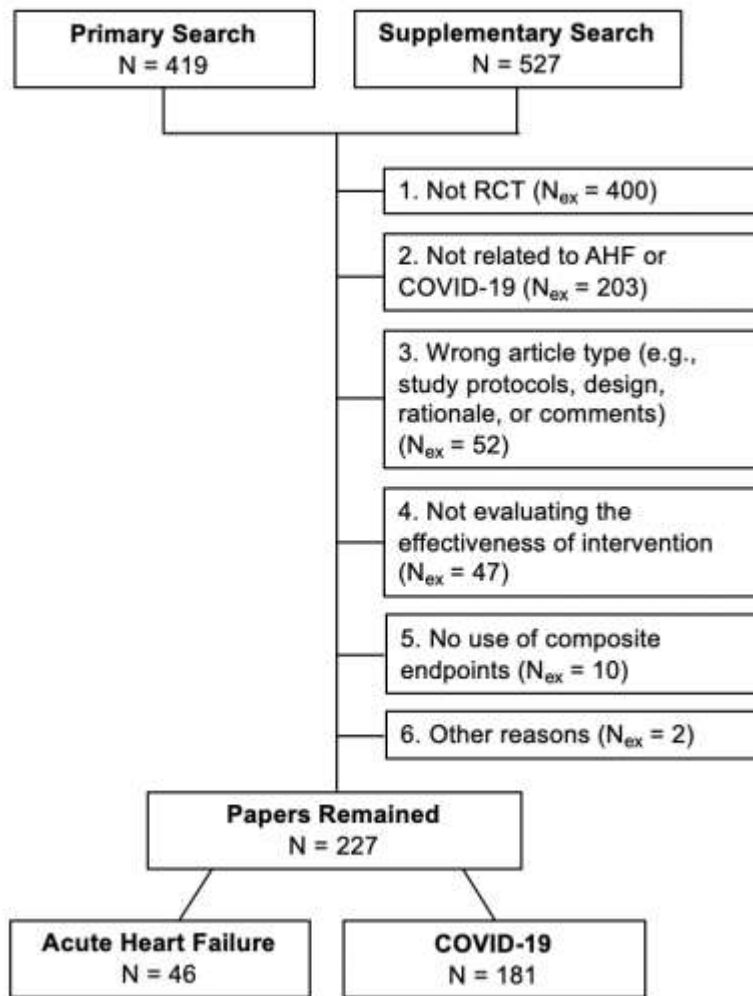
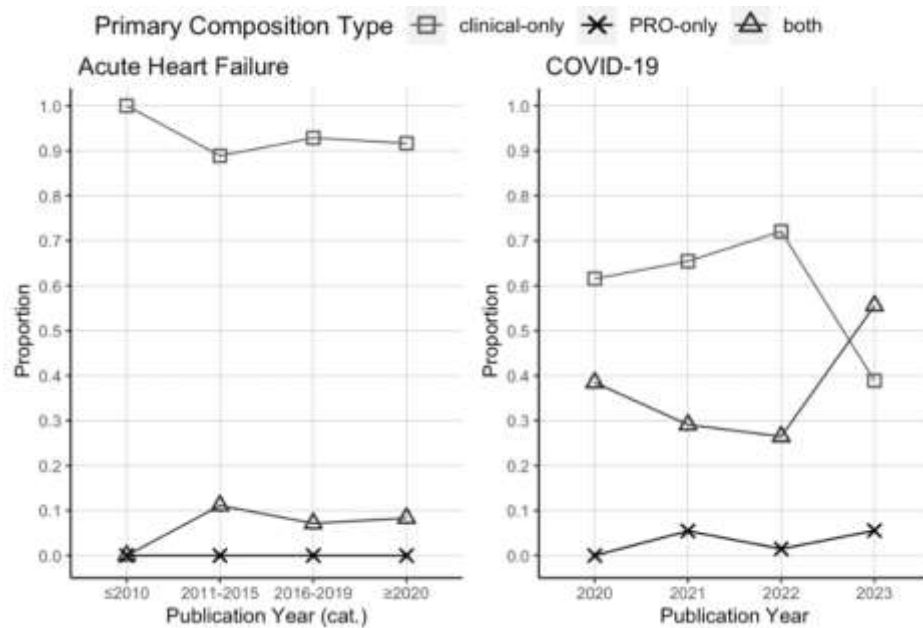
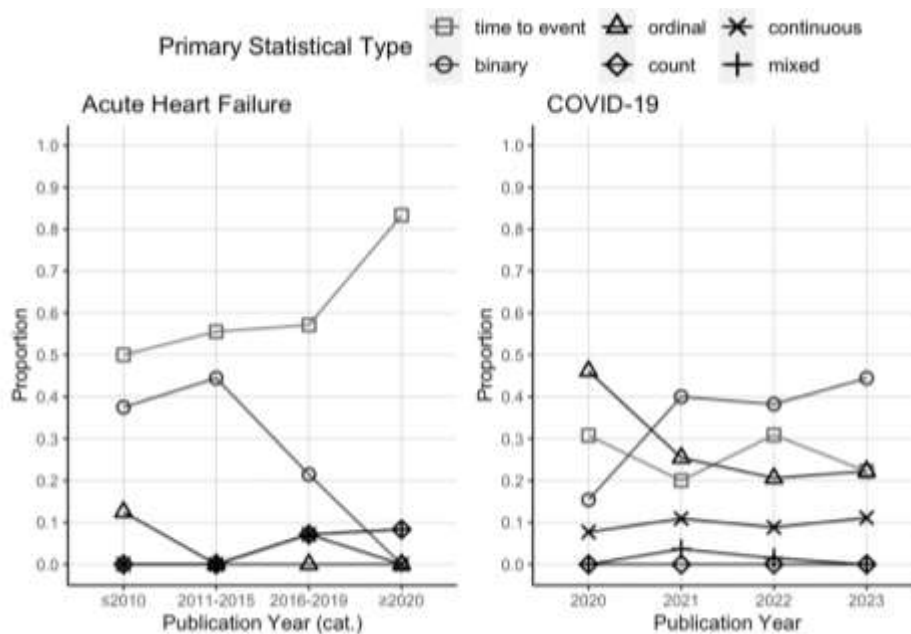


Figure 1. Flow diagram for the inclusion/exclusion of clinical trial publications in our study. There are 946 publications in total with 419 from the primary search and 527 from the supplementary search. The 6 exclusion reasons are listed in a hierarchical order, with 1>2>...>6. That is, if a paper is neither a randomized clinical trial (RCT; reason 1) nor related to Acute Heart Failure (AHF) or COVID-19 (reason 2), it will be classified as "1. Not RCT". The number of papers excluded due to a specific reason is denoted as " N_{ex} " presented in the parentheses. In the end, 227 papers met inclusion criteria with 46 from AHF trials and 181 from COVID-19 trials.



(a)



(b)

Figure 2. Distribution of (a) composition type and (b) statistical type of the primary composite endpoints summarized over publication years for Acute Heart Failure (AHF) trials (≤ 2010 , 2011-2015, 2016-2019 and ≥ 2020) and COVID-19 trials (2020-2023). In the figure, "cat." and "PRO" are abbreviations for "categorical" and "patient-reported outcome".