Subgroup analyses of randomized clinical trials in heart failure: facts and numbers

Andrija Vidic1, John T. Chibnall2, Niharika Goparaju1 and Paul J. Hauptman1*

1Department of Medicine and 2Department of Psychiatry, Saint Louis University School of Medicine, St. Louis, MO, USA

Abstract

Subgroup analyses of major randomized clinical trials in heart failure are published frequently, but their impact on medical knowledge and practice guidelines has not been previously reported. In a novel analysis, we determined number of citations, impact factors, number of authors, and citations in guidelines of both parent trials and sub-studies; we also qualitatively assessed whether the analyses were described as post-hoc and non-pre-specified. A total of 229 sub-studies evaluating outcomes in patient subgroups were published (median 6, range 0–36 per trial). The number of subjects in the parent trials positively correlated with number of sub-studies (rho = 0.51, P = 0.009). The subgroups are frequently not pre-specified. The impact factors of sub-studies were lower in comparison to the parent trials as were the number of citations two years after the publication date; in addition, parent trials were cited more frequently in European and American professional guidelines compared with the sub-studies. We maintain that the sub-studies derived from major heart failure trials are frequently published, but their contribution to clinical guidelines and medical knowledge are highly debatable.

Keywords Heart failure; Sub-studies; Randomized clinical trials

*Correspondence to: Paul J. Hauptman, Saint Louis University Hospital, 3635 Vista Avenue, St. Louis, MO 63110, USA. Tel: 314-268-5293; Fax: 314-577-8861. Email: hauptmpj@slu.edu

Major advances in the management of heart failure have been established in large, double-blind, randomized placebo-controlled clinical trials (RCTs) of both device and drug interventions. Data derived from these trials have influenced clinical practice guidelines, quality metrics, and patient care. Following the dissemination of the results of RCTs, additional analyses evaluating the efficacy and/or safety of the particular intervention in specific patient subgroups are published frequently. However, it is not clear whether these sub-studies add meaningfully to general medical knowledge, in large part because the analyses are often not pre-specified and the specific patient subgroups were not included in the randomization schema.1,2

Wittes has stated that ‘if reporting on subgroups is tempting but treacherous, failing to report on them seems unscientific and incurious’.3 This tension was also highlighted by Feinstein who termed subgroup analysis a ‘clinic-statistical tragedy’; that is, statisticians and clinicians approach subgroups from different perspectives. Broadly stated, he summarized the challenge of placing subgroup analysis in context by area of expertise: ‘The statisticians are right in denouncing subgroups that are formed post hoc from exercises in pure data dredging. The clinicians are also right, in insisting that a subgroup is respectable and worthwhile when established a priori from pathophysiological principles’.4

In light of this ongoing controversy and the plethora of subgroup analyses that populate the medical literature following reporting of the results of the parent trial, we sought to characterize the types of sub-studies that appear after the initial RCT report and to critically analyse their impact. The major goal was to evaluate the number and scope of these sub-studies and their contribution to clinical practice guidelines, specifically the 2013 American College of Cardiology Foundation (ACCF)/American Heart Association (AHA) Guideline for the Management of Heart Failure and the 2012 European Society of Cardiology Guidelines for the Diagnosis and Treatment of Acute and Chronic Heart Failure.5,6 We were also interested in the proportion of sub-studies that directly referenced an interaction effect as represented in the forest plot of the parent trial publication.2

© 2016 The Authors. ESC Heart Failure published by John Wiley & Sons Ltd on behalf of the European Society of Cardiology. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
A method to assess the impact of subgroup analyses

We employed a widely used CHF Trials application\(^7\) to identify major RCTs published in the last two decades that evaluated classes of medications and devices, many of which are accepted as guideline-directed therapies for the treatment of congestive heart failure (CHF). Only trials with more than 500 participants were included in five major therapy groups: angiotensin receptor blockers/angiotensin converting enzyme inhibitors (ARB/ACEI), beta-blockers, aldosterone antagonists, implantable cardioverter defibrillators (ICDs), and cardiac resynchronization therapy (CRT). The three studies from the Candesartan in Heart failure—Assessment of Mortality and Morbidity (CHARM) programme (CHARM-Added, CHARM-Alternative and CHARM-Preserved) were considered to be a single trial.

To focus the analysis on sub-studies that involved comparison of treatment effect centred on baseline patient characteristics, we refined our sample by excluding meta-analyses and sub-studies that focused on biomarkers or imaging, mode of death, risk models, cost-analyses, registries, and evaluation of quality-adjusted life years (QALYs) (Table 2). These latter analyses can provide insight into mechanism of action and pathophysiology and may use outcome measures that are not part of the primary or secondary endpoints (e.g. impact of an intervention on a biomarker).

The Web of Science Database was used to identify a comprehensive list of all English language sub-studies pertaining to the parent RCTs.\(^8\) Using the 'Times Cited' option on Web of Science, we identified all sub-studies that cited the original article. Similarly, we analysed the number of citations of sub-studies over a two-year window following the publication date of the sub-study. Therefore, sub-studies published after November 2012 were excluded. The impact factors for the journals in which the parent trials and sub-studies were published were obtained from Journal Citation Reports (JCR) for the specific year of publication.\(^9\) Sub-studies published in 2014 were excluded from this part of the analysis as the impact factors were not known.

The endpoints used in the sub-studies were examined for whether or not they were the primary endpoint(s) or secondary endpoint(s) in the parent RCT. We also examined (1) whether parent trial characteristics (size, impact factor, number of authors) predicted the publication of subgroup analyses and (2) whether the parent trial and sub-studies were referenced in the 2013 ACCF/AHA Guidelines by searching on author name, title of study and name of drug or device. Studies published in 2013, 2014 or 2015 were excluded from this part of the analysis as they were published too late to be referenced. Similarly we evaluated the 2012 European Society of Cardiology Guidelines for the Diagnosis and Treatment of Acute and Chronic Heart Failure, excluding studies published in 2012 to 2015 inclusive.

Descriptive statistics were calculated for all variables. Means and standard deviations were used where applicable; in the case of extreme skewness, median and range were used. Spearman rho correlation coefficients were calculated between some sub-study and parent study variables.

The data on subgroup analyses: large numbers, uncertain impact

Our findings raise issues about the value of subgroup analyses. Of 25 major clinical trials relevant to guideline-directed optimal medical therapy in heart failure published between 1996 and 2013 with greater than 500 participants, a total of 477 publications were identified that included analyses of data derived from the parent RCT. As shown in Figure 1, we subsequently removed 248 papers that met our exclusion criteria.

The median number of study patients was 3043 (range 571–14 703). The mean numbers of primary and secondary endpoints in the RCTs were 1.4 ± 0.6 and 3.0 ± 2.3, respectively. One or more pre-specified primary endpoints were met in 88% (22/25) of RCTs. The number of subgroups mentioned in the parent study (in text or forest plots) ranged from 1 to 19 (mean 8.9 ± 5.0). All but one of the RCTs was financially supported by industry (pharmaceutical or device).

The median number of sub-studies per RCT was 6.0 (range 0–36). The number of subjects in the main study was positively correlated with the number of sub-studies (rho = 0.51, P = 0.009). The median number of new endpoints introduced in the sub-studies was 3.5 (range 0–45). The median number of months between publication of the parent RCT and publication of the sub-studies was 48 (range 4–178); thus, 50% of sub-studies were published more than 4 years after publication of the parent study (Figure 2). Excluding three RCTs where the author byline indicated a consortium rather than individual authors, the numbers of authors on the parent RCTs and the sub-studies were comparable, with means of 10.9 ± 3.6 and 9.2 ± 2.8, respectively.

Sub-study analyses of subgroups that were not pre-specified or part of the randomization scheme in the parent

---

Table 1 Inclusion and Exclusion Criteria

| Inclusion criteria | Exclusion criteria |
|--------------------|--------------------|
| >500 study subjects | Meta analyses |
| Trial listed in the application found at www.imedicalapps.com for the key therapeutic classes: ACEI/ARB | Biomarker substudies |
| Beta-blocker | Imaging substudies |
| Mineralocorticoid receptor antagonist | Mode of death analyses |
| ICD/CRT device | Risk model development |
| Trial dates between 1996 and 2013 | Cost and QALY analyses |

---

DOI: 10.1002/ehf2.12093
RCT were common (186/229 or 81.2% and 183/229 or 79.9%, respectively). The number of subjects in the parent RCTs was positively correlated with the number of sub-studies with subgroups that were not pre-specified (rho = 0.44, P = 0.03).

The impact factor of the parent RCT publication (mean 31.6 ± 14.0) was much higher than the sub-study impact factors (mean 6.5 ± 4.3), as were the number of citations (main study mean 233.7 ± 102.2 vs. sub-study mean 10.8 ± 10.0) (Table 2). The parent study impact factor was negatively correlated with median months to publication of sub-studies (rho = −0.41, P = 0.04): higher parent RCT impact factor was associated with faster publication of sub-studies. In addition, higher impact factor parent RCTs generated higher average impact factor sub-studies (rho = 0.44, P = 0.03).

The parent RCT was much more likely to be referenced in the ACCF/AHA and ESC Guidelines (87.5% and 83.3%, respectively) compared with the sub-studies (5.2% and 1.8%, respectively) (Table 3). Although a slight majority (141/229, 61.6%) of sub-studies tested an interaction effect (e.g., moderation of parent trial efficacy as a function of a subject characteristic variable), only one-third examined an interaction effect represented in the parent trial forest plot.

Prior evaluations of the value of subgroup analyses derived from HF Trials

An extensive literature exists about the limitations of subgroup analyses derived from RCTs.1-4,10,11 While these analyses, in theory, permit evaluation of treatment effects in select patient subgroups defined by baseline characteristics, there are multiple statistical pitfalls that can lead to inappropriate conclusions and, by extension, misguided clinical decision-making. Assman and colleagues reviewed original RCT reports published in four prominent journals in 1997 and found a numerical range of subgroups from 1 to 24, often accompanied...
by improper statistical testing. In a subsequent analysis by Sun et al., 207 of 469 published RCTs were found to include subgroup analyses; the number of subgroup analyses was correlated with higher impact factors and a larger number of patients in the clinical trial. Subgroup analyses without statistical significance were more likely to be published from parent studies that were industry funded. Hernandez et al. noted subgroup analyses in 39 of 63 RCTs from the cardiovascular discipline, but with only 11 reporting tests of interaction between average and subgroup treatment effects; larger studies were more likely to include subgroup analyses. None of these studies critically evaluated the publication of separate papers that involved subgroup analyses and followed the publication of the parent RCT.

We noted that the impact factor of the journals and citation index regarding sub-studies are much lower compared with parent RCTs. From our perspective, we believe that there is a general lack of clarity in defining the endpoints as post hoc and/or not pre-specified, consistent with a prior report. Importantly, very few sub-studies are referenced in the major professional guidelines, suggesting that they may be seen at best as hypothesis generating. As such, they have relatively limited relevance and/or do not meet a more rigorous threshold upon which formal recommendations for clinical care can be made. Lastly, sub-study investigations of interaction effects originated from parent trial interaction forest plots in only about one-third of the cases, indicating that two-thirds of interaction sub-studies were post hoc in nature.

---

**Table 2** Comparisons of parent RCT and sub-study impact factors and citation indices

| Drug/device class       | Impact factor of parent RCT | Citation index of parent RCT | Impact factor of sub-studies | Citation index of sub-studies |
|-------------------------|-----------------------------|------------------------------|-----------------------------|-------------------------------|
| Aldosterone Antagonist  | 39.0 (12.7), 3              | 293.7 (57.1), 3              | 8.0 (4.7), 22               | 12.4 (11.4), 18              |
| ARB/ACEI                | 32.6 (11.5), 5              | 165.2 (79.9), 5              | 7.2 (4.4), 59               | 11.8 (10.3), 57             |
| Beta-Blocker            | 16.5 (8.1), 7               | 226.4 (114.8), 7             | 5.4 (3.1), 42               | 11.1 (9.4), 40              |
| CRT                     | 41.5 (11.9), 7              | 244.8 (109.2), 6             | 7.9 (4.9), 44               | 13.0 (12.7), 27             |
| ICD                     | 34.5 (8.5), 3               | 283.0 (115.6), 3             | 4.8 (3.1), 48               | 7.2 (6.7), 44               |
| Total                   | 31.6 (14.0), 25             | 233.7 (102.2), 24            | 6.5 (4.3), 215              | 10.8 (10.0), 186            |

ARB/ACEI, angiotensin receptor blocker/angiotensin converting enzyme inhibitor; CRT, cardiac resynchronization therapy; ICD, implantable cardioverter defibrillator.

RCTs on aldosterone inhibitors are the following: EMPHASIS-HF, EPHESUS, RALES. The RCTs of ARB/ACEI include the following: VALIANT, CHARM, HEAAL, I-PRESERVED, Val-HeFT. The RCTs under beta-blocker include the following: CAPRICORN, CIBIS-II, COMET, COPERNICUS, MERIT-HF, SENIORS, US-Carvedilol Studies. The RCTs on CRT are as follows: BLOCK-HF, CARE-HF, COMPANION, MADIT-CRT, MIRACLE, MIRACLE-ICD, RAFT. The RCTs on ICD are as follows: AVID, MADIT-II, SCD-HeFT.

Data are mean (SD), N.
caused harm to patients of erroneous interpretation of subgroup analyses that have

References

Despite this, papers that include subgroup analyses continue to be published for years following the publication of the parent RCT (median 4 years). In our analysis, the size of the study was positively correlated with the number of sub-studies, and the higher the impact factor of the parent RCT, the faster the sub-studies were published. Nevertheless, because it is generally not feasible to critically assess the impact of subgroup analyses on clinical care, we can ask: do the results of these analyses impact clinician decision-making? If so, how and to what degree? Well-designed focus groups with clinicians could provide some insight into the extent to which results from these studies are incorporated into daily practice. Whether they should in fact influence decisions about patient selection for specific therapy is less clear.

Concluding thoughts

Although a guideline exists that supports the publication of subgroup analysis, this is far from universally accepted. In a frequently cited example, the authors of the definitive paper from the ISIS-2 study were asked by the journal editors to include patient astrological sign in one of the tables in order to highlight the ‘trap’ associated with subgroup analysis. Peter Sleight has also commented that there are ‘...examples of erroneous interpretation of subgroup analyses that have caused harm to patients.’

Subgroups have the potential to generate hypotheses for further prospective investigation; there is one such example in the heart failure discipline in which the combination of hydralazine and nitrates was determined to be effective in African Americans in the Vasodilator-Heart Failure Trial (V-HeFT), and subsequently this group was examined in the African–American Heart Failure Trial (A-HeFT). However there are also a significant number of examples in which subgroups led to additional negative studies, such as amiodarone in the elderly and amlodipine in patients with non-ischemic cardiomyopathy or more commonly to no studies at all.

In summary, subgroup analyses are frequently published, vary in their transparency about the nature of the statistics (in particular whether the subgroups were pre-specified), and infrequently contribute to recommendations in published clinical guidelines. We believe that a uniform approach and greater degree of rigour may be required in assessing the value of these studies prior to publication, incorporation into clinical practice guidelines, and by extension, clinical practice itself.

Conflicts of interest

None declared.

References

1. Assman SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis) uses of baseline data in clinical trials. Lancet 2000; 355: 1064–1069.
2. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in Medicine — reporting of subgroup analyses in clinical trials. N Engl J Med 2007; 357: 2189–2194.
3. Wittes J. On looking at subgroups. Circulation 2009; 119: 912–915.
4. Feinstein AR. The problem of cogent subgroups: a clinico-statistical tragedy. J Clin Epidemiol 1998; 51: 297–299.
5. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJV, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WHH, Tsai EJ, Wilkoff BL. 2013 ACCF/AHA Guideline for the Management of Heart Failure: A report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines. J Am Coll Cardiol 2013; 62: e147–e239.
6. McMurray JJV, Adamopoulos S, Anker SD, Auricchio A, Bohl M, Dickstein K, Falk V, Filippatos G, Fonseca C, Gomez-Sanchez MA, Jaarsma T, Kober L,

Table 3 Representation of parent RCTs and sub-studies in guideline statements

| Drug/device class | ACCF/AHA guidelines | ESC guidelines |
|-------------------|---------------------|---------------|
|                   | Parent RCT | Sub-studies | Parent RCT | Sub-studies |
| Aldosterone Antagonist | 100% (3/3) | 0% (0/20) | 100% (3/3) | 0% (0/12) |
| ARB/ACEI | 80% (4/5) | 5% (3/58) | 80% (4/5) | 5% (3/55) |
| Beta-blocker | 86% (6/7) | 9% (4/42) | 100% (7/7) | 0% (0/38) |
| CRT | 100% (6/6) | 7% (2/27) | 67% (4/6) | 0% (0/21) |
| ICD | 67% (2/3) | 2% (1/46) | 67% (2/3) | 0% (0/43) |
| Total | 87% (21/24) | 5% (10/193) | 83% (20/24) | 2% (3/169) |

ACCF/AHA, American College of Cardiology Foundation/American Heart Association; ARB/ACEI, angiotensin receptor blocker/angiotensin converting enzyme inhibitor; CRT, cardiac resynchronization therapy; ESC, European Society of Cardiology; ICD, implantable cardioverter defibrillator; RCT, randomized controlled clinical trial.

Data are % (n/N).

*Reference 5.
**Reference 6.
