Effect of Study Design on the Reported Effect of Cardiac Resynchronization Therapy (CRT) on Quantitative Physiological Measures: Stratified Meta-Analysis in Narrow-QRS Heart Failure and Implications for Planning Future Studies

Richard J. Jabbour, MRCP; Matthew J. Shun-Shin, MRCP; Judith A. Finegold, MA, MRCP; S. M. Afzal Sohaib, MRCP; Christopher Cook, MRCP; Sukhjinder S. Nijjer, MRCP; Zachary I. Whinnett, PhD, MRCP; Charlotte H. Manisty, PhD, MRCP; Josep Brugada, MD; Darrel P. Francis, MA, MD, FRCP

Background—Biventricular pacing (CRT) shows clear benefits in heart failure with wide QRS, but results in narrow QRS have appeared conflicting. We tested the hypothesis that study design might have influenced findings.

Method and Results—We identified all reports of CRT-P/D therapy in subjects with narrow QRS reporting effects on continuous physiological variables. Twelve studies (2074 patients) met these criteria. Studies were stratified by presence of bias-resistance steps: the presence of a randomized control arm over a single arm, and blinded outcome measurement. Change in each endpoint was quantified using a standardized effect size (Cohen’s d). We conducted separate meta-analyses for each variable in turn, stratified by trial quality. In non-randomized, non-blinded studies, the majority of variables (10 of 12, 83%) showed significant improvement, ranging from a standardized mean effect size of +1.57 (95%CI +0.43 to +2.7) for ejection fraction to +2.87 (+1.78 to +3.95) for NYHA class. In the randomized, non-blinded study, only 3 out of 6 variables (50%) showed improvement. For the randomized blinded studies, 0 out of 9 variables (0%) showed benefit, ranging from −0.04 (−0.31 to +0.22) for ejection fraction to −0.1 (−0.73 to +0.53) for 6-minute walk test.

Conclusions—Differences in degrees of resistance to bias, rather than choice of endpoint, explain the variation between studies of CRT in narrow-QRS heart failure addressing physiological variables. When bias-resistance features are implemented, it becomes clear that these patients do not improve in any tested physiological variable. Guidance from studies without careful planning to resist bias may be far less useful than commonly perceived. (J Am Heart Assoc. 2015;4:e000896 doi: 10.1161/JAHA.114.000896)

Key Words: cardiac resynchronization therapy • heart failure • narrow QRS

Cardiac resynchronization therapy (CRT) undoubtedly provides both symptomatic and prognostic benefit in patients with heart failure and a wide QRS complex.1–3 Whether it is effective in patients with narrow QRS complexes has appeared contentious. Studies addressing this have implemented bias-resistance steps (such as the inclusion of a randomized control arm and blinding of endpoint assessment) to varying degrees, and have addressed a variety of endpoints. While these studies have been reviewed in the past, no meta-analysis has focused on trial design as a potential explanatory variable for the differing results.4

We formally assessed the effect of CRT in patients with narrow QRS, to identify whether the conflict between different study results was an effect of trial design. To make it possible to compare effects on different endpoints, we calculated for each the standardized effect size (Cohen’s d).

Methods

Eligibility and Search Strategy
We identified all reports of studies of heart failure patients with narrow QRS (<130 ms) and had either CRT pacemaker


Correspondence to: Richard J. Jabbour, MRCP, International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, 59-61 North Wharf Road, London W2 1LA, United Kingdom. E-mail: rjj03@ic.ac.uk

Received July 4, 2014; accepted October 24, 2014.

© 2015 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley Blackwell. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
or CRT defibrillator implantation (CRT-P or CRT-D) inserted.

MEDLINE (1946–September 2013), EMBASE (1974–September 2013), the Cochrane central register of controlled trials (Cochrane Library 2011, Issue 4), and www.controlled-trials.com (a meta-registry of randomized controlled clinical trials that includes the ISRCTN register) were searched using appropriate terms in the online appendix. Reference lists of the retrieved articles were hand-searched for additional publications. Conference presentations of the reported trials were used if they provided incremental information.

Effect Sizes

The primary aim of this meta-analysis was to assess whether bias-resistance elements of study design affect study results. For each study we included all reported measured variables of functional status provided on a continuous scale and common left ventricular function measurements. For randomized controlled trials we defined the effect size as the difference between the change scores in each arm. For single arm studies we defined the effect size as the reported change score in the intervention arm. To allow measurements of different physiological quantities to be compared on a common scale we calculated for each the standardized effect size (Cohen’s d) by dividing by the standard deviation of that variable in the patients before CRT implantation.

Classification of Studies by Presence of Bias-Resistance Features

We stratified the studies into 3 broad groups depending on the number of bias-resistance features:

- 0 bias-resistance features (Neither an equivalent control group nor blinding of measurements);
- 1 bias-resistance feature (Randomization to a control group or intervention, but without blinding of patient and echocardiographic operator);
- 2 bias-resistance features (Randomization with blinding of patient and echocardiographic operator).

The bias-resistance features were only considered valid if the results presented used them. For example, if a study had a randomized control arm but only presented data from the intervention arm then we were obliged to consider the report to be of a single arm study.

We further assessed all studies using the Cochrane “risk of bias” tool to qualitatively identify if there were any additional sources of biases that could have affected the results.

Data Abstraction

Data was abstracted in duplicate by 2 authors (RJ and CC). Disagreements were resolved by a third author (MJS).

Data Analysis

We summarized the data and tested for inequality between the groups using a random-effects meta-analysis using the statistical environment “R” with the “metafor” package. We stratified by trial quality along with end-point. Data were graphically presented using the package “ggplot2.”

Results

Search Results and Classification of Studies by Bias-Resistance of Design

Three hundred eighty-two articles met the initial search criteria, of which 131 were duplicates and 51 were excluded at the abstract stage. From 136 full-text articles screened, 12 studies (2074 patients) met the inclusion criteria (Figure 1). Of these 12 studies, 9 enrolled patients with QRS durations <120 ms, and 3 enrolled patients with QRS <130 ms. The characteristics and classification of studies are presented in Table 1.

Across the 12 studies, various echocardiographic left ventricular functional and size parameters, and functional measured physiological variables were reported including: ejection fraction, end systolic volume, end diastolic volume, end systolic diameter, end diastolic diameter, sub-maximal exercise duration, quality of life score, NYHA class change, 6-minute walk distance, myocardial performance index, peak VO₂, and peak VE/VO₂ slope [Table 2].

![Figure 1. PRISMA flow diagram of studies](http://jaha.ahajournals.org/)

DOI: 10.1161/JAHA.114.000896
Table 1. Selected Baseline Characteristics of Studies Involving CRT and Narrow QRS

<table>
<thead>
<tr>
<th>Study</th>
<th>Blinded</th>
<th>Random</th>
<th>Arms</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>Enrollment Period</th>
<th>Intervention Group</th>
<th>Control/Comparison Group</th>
<th>Outcomes Reported</th>
<th>Duration of Follow-up</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECHO-CRT(^1)</td>
<td>Yes</td>
<td>Yes</td>
<td>Two</td>
<td>NYHA III/V</td>
<td>Dyssynchrony(^1) EF≤35% QRS&lt;130 ms</td>
<td>AF: Acute heart failure</td>
<td>August 2008 to March 2013</td>
<td>CRT-D n=404</td>
<td>D-ICD n=405</td>
<td>Death NYHA QOL</td>
<td>6 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LESSER-EARTH(^12)</td>
<td>Yes</td>
<td>Yes</td>
<td>Two</td>
<td>ICD indication</td>
<td>EF≤35% QRS&lt;120 ms</td>
<td>Permanent AF ACS≤6 weeks Previous CRT device</td>
<td>October 2003 to January 2011</td>
<td>CRT-D n=44</td>
<td>D-ICD n=41</td>
<td>Sub Ex 6MWT EF, LVESV NYHA, QOL</td>
<td>12 months</td>
</tr>
<tr>
<td>RethinQ(^17)</td>
<td>Yes</td>
<td>Yes</td>
<td>Two</td>
<td>NYHA III QRS&lt;130 ms EF≤35%</td>
<td>Previous CRT device</td>
<td>August 2005 to January 2007</td>
<td>CRT-D n=87</td>
<td>D-ICD n=85</td>
<td>QOL, 6MWT Peak oxygen consumption</td>
<td>6 months</td>
<td>Double blind</td>
</tr>
<tr>
<td>NARROW-CRT(^10)</td>
<td>No</td>
<td>Yes</td>
<td>Two</td>
<td>EF≤35% Dyssynchrony(^1) Max. medical therapy QRS&lt;120 ms</td>
<td>Conventional indication for pacing Persistent AF</td>
<td>January 2008 to May 2010</td>
<td>CRT-D n=60</td>
<td>D-ICD n=60</td>
<td>HF clinical composite score HF hospitalization HF death LVEF LVESV/DV</td>
<td>12 months</td>
<td>No change in echo parameters stated in control arm</td>
</tr>
<tr>
<td>RESPOND(^8)</td>
<td>No</td>
<td>Yes</td>
<td>Two</td>
<td>EF≤35% NYHA III/V QRS&lt;120 ms Chronic HF Max. medical therapy</td>
<td>Conventional indication for pacing Recent MI</td>
<td>August 2007 to September 2009</td>
<td>CRT-D n=29</td>
<td>OPT n=31</td>
<td>NYHA LVEF QOL 6MWT LVESV/DV</td>
<td>6 months</td>
<td>Maximum medical therapy control arm</td>
</tr>
<tr>
<td>Achilli et al(^16)</td>
<td>No</td>
<td>No</td>
<td>Two</td>
<td>EF≤35% Dyssynchrony(^1) Chronic HF NYHA III/V</td>
<td>Permanent AF Valvular disease(^*) ACS≤3 months Severe COPD</td>
<td>February 2000 to March 2002</td>
<td>CRT n=14</td>
<td>CRT n=38</td>
<td>QRS&gt;120 ms NYHA 6MWT LVEF, LVESV/DD</td>
<td>45 months</td>
<td>Patients divided into two groups based on QRS duration</td>
</tr>
<tr>
<td>Bleaker et al(^18)</td>
<td>No</td>
<td>No</td>
<td>Two</td>
<td>EF≤35% Dyssynchrony(^1) Chronic HF NYHA III/V QRS&lt;120 ms</td>
<td>ACS&lt;3 months Decompensated HF</td>
<td>Insufficient information</td>
<td>CRT-D/P n=33</td>
<td>CRT-D/P n=33</td>
<td>NYHA LVEF LVESV/DV</td>
<td>6 months</td>
<td>Broad QRS comparison group</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Study</th>
<th>Blinded</th>
<th>Random</th>
<th>Arms</th>
<th>Inclusion Criteria</th>
<th>Exclusion Criteria</th>
<th>Enrollment Period</th>
<th>Intervention Group</th>
<th>Control/Comparison Group</th>
<th>Outcomes Reported</th>
<th>Duration of Follow-up</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESTEEM-CRT7</td>
<td>No</td>
<td>No</td>
<td>Single</td>
<td>EF ≤ 35% Dysynchrony†</td>
<td>Persistent AF Sustained VT COPD Bradycardia pacing</td>
<td>June 2005 to December 2007</td>
<td>CRT-D n=68</td>
<td>NA</td>
<td>NYHA, QOL, LVEF/LVESD/LVED</td>
<td>12 months</td>
<td>Single arm; Multi-centre</td>
</tr>
<tr>
<td>Gasparini et al8</td>
<td>No</td>
<td>No</td>
<td>Two</td>
<td>NYHA II/IV Chronic HF EF ≤ 35% QRS &gt; 120 ms</td>
<td>Insufficient information</td>
<td>October 1999 to April 2005</td>
<td>CRT n=45</td>
<td>CRT n=331 QRS &gt; 120 ms</td>
<td>NYHA EF, LVEF/LVESV + 6MWT</td>
<td>28 months (Range: 6–68 months)</td>
<td>Broad QRS control group</td>
</tr>
<tr>
<td>PROSPECT14</td>
<td>No</td>
<td>No</td>
<td>Single</td>
<td>EF ≤ 35% Dysynchrony† NYHA III/IV QRS ≤ 130 ms</td>
<td>Insufficient information</td>
<td>March 2004 to November 2005</td>
<td>CRT-P/D n=41</td>
<td>NA</td>
<td>NYHA, QOL 6MWT, CCS LVEF, LVESV/SV MPI</td>
<td>6 months</td>
<td>Single arm sub-study; multi-centre</td>
</tr>
<tr>
<td>van Bommel13</td>
<td>No</td>
<td>No</td>
<td>Single</td>
<td>EF ≤ 35% Dysynchrony† Chronic HF QRS ≤ 120 ms</td>
<td>Insufficient information</td>
<td>Insufficient information</td>
<td>CRT-P/D n=123</td>
<td>NA</td>
<td>LVEF LVESV/DV NYHA</td>
<td>6 months</td>
<td>Single arm; multi-centre</td>
</tr>
<tr>
<td>Yu et al15</td>
<td>No</td>
<td>No</td>
<td>Two</td>
<td>NYHA II/IV Dysynchrony† QRS ≤ 120 ms</td>
<td>Insufficient information</td>
<td>Insufficient information</td>
<td>CRT n=51</td>
<td>CRT n=51 QRS &gt;120 ms</td>
<td>NYHA, QOL 6MWT, Ex Mitral regurgitation LVEF, MPI LVESV/LVED</td>
<td>3 months</td>
<td>Broad QRS comparison group</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; CCS, clinical composite score; D-ICD, dual chamber ICD; EDD, left ventricular end diastolic diameter; EDV, left ventricular end diastolic volume; EF, ejection fraction; ESD, left ventricular end systolic diameter; ESV, left ventricular end systolic volume; Ex, exercise capacity, metabolic equivalent; MPI, myocardial performance index; OPT, optimal medical therapy; QOL, quality of life questionnaire; Sub Ex, exercise duration at submaximal level.

*Surgically correctable significant valvular disease.
†Echocardiographic evidence of interventricular/intraventricular asynchrony.
Classification of Studies by Bias-Resistance Features

Eight studies, including 435 patients, had neither a randomized controlled arm nor blinding. Three of these were single armed studies. In four of these studies comparison data was presented, but arose from patients with a QRS above the threshold and so were analysed as single armed. One of these studies was carried out as a randomized trial, but only presented continuous variable data from the intervention arm, and hence was analysed as a single armed study.

One study had a randomized control arm but neither patients nor echocardiographers were blinded to whether CRT was active, enrolling 60 patients. Three studies had both a randomized control arm and blinding of patients and sonographers totaling 1066 patients. In these trials the patients received a CRT-D device, with the control patients having the CRT function inactive. The baseline characteristics and effects of CRT in the intervention arm are available in Table 3 and the Cochrane risk of bias assessment tool in Table 4.

Is it Bias-Resistance of Study Design or Choice of End-Point That Leads to Unintentionally False-Positive Results?

A series of meta-analyses, one for each end-point, stratified by the presence of bias-resistance features, is shown in Figure 2.

The green diamonds show the meta-analysis summary results of the randomized, blinded studies. None (0/9, 0%) of these showed a significant effect of CRT on its end-point. However, the trials with fewer bias-resistance features showed a different pattern.

The orange diamonds show the results of the randomized, unblinded study. Half (3/6, 50%) of the end-points showed statistically significant improvement.

The red diamonds show the meta-analysis summary results of the studies with neither a randomized controlled arm nor blinding. Most (10/12, 83%) end-points showed statistically significant improvement.

Danger of Viewing Multiple Positive End-Points in a Trial as Strong Evidence

Commentators sometimes remark on the multiplicity of positive end-points within a single arm study as though their great number might somehow overcome the weakness of the study in lacking blinding, or even lacking a control arm. This is not wise because this counts the same patients on multiple occasions as though they were separate.

Figure 3 illustrates this danger of viewing multiple positive endpoints as strong evidence. It shows 1 point for each endpoint reported in each of the studies. The number of patients (and bias-resistance of the design) is given by the horizontal position, and the apparent standardized effect size is given by the vertical position. The left-hand group of studies, with only a single arm, and therefore the least bias-resistance, give the
Figure 2. Meta-analyses of effects on physiological variables, with studies stratified by number of bias-resistance features in the study design. For each variable a meta-analysis was conducted stratified by the presence of bias-resistance features. The majority (10 out of 12) of variables reported in studies without randomization and blinding (red symbols) favored intervention to a statistically significant degree. In contrast, all 9 of the outcome variables reported by randomized, blinded trials (green symbols) were neutral. The 6 variables reported by studies with randomization but not blinding (orange symbols) were equally divided between suggesting significant response to intervention (3 variables) and not (3 variables).
impression of efficacy. The right-hand group of studies, which have randomization and blinding, do not.

The key factor is the bias-resistance of the study design, rather than the choice or number of endpoints.

Discussion

The conflict between study reports on the efficacy of biventricular pacing (CRT) on physiological variables in heart failure with narrow QRS, seems to originate not in the choice of physiological endpoint, but in the design of the study. There may have been unintentional bias introduced when study design did not possess bias-resistance features. Measurements made for routine clinical purposes do not have the correct properties for drawing reliable scientific inferences. Readers may not realize that such data are not equivalent to data from a scientific experiment carefully designed to answer a question reliably.

In the 3 studies implementing both a randomized control arm and blinding, the effect of CRT on endpoints is neutral on these physiological variables. Two of these trials stopped early due to futility because of detrimental signals in event rates.11,12

Rationale for CRT Implantation in Relation to QRS Complex Width

Broad QRS complexes were the defining characteristic of the early patients receiving CRT from the very first case reports through to the pivotal physiological studies and landmark trials.2,20-23 The powerful symptomatic and morbidity/mortality reduction were a strong stimulus for attempting expansion into patients with narrow QRS.3,24

One rationale for such expansion has been the umbrella concept of dyssynchrony. It was conjectured that CRT might, even in the absence of electrical dyssynchrony (wide QRS), alleviate isolated mechanical (echocardiographic) dyssynchrony. More recently however, it has emerged that the apparent predictive power of mechanical (echocardiographic) dyssynchrony for benefit from CRT exists only when studies
## Table 3. Baseline Characteristics and Effects of CRT in Narrow QRS Arm of Each Study

<table>
<thead>
<tr>
<th>Baseline Characteristics</th>
<th>ECHO-CRT</th>
<th>LESSER-EARTH</th>
<th>RethinQ</th>
<th>NARROW-CRT</th>
<th>RESPOND</th>
<th>Achilli et al</th>
<th>Beeker et al</th>
<th>ESTEM-CRT</th>
<th>Gasparini et al</th>
<th>PROSPECT</th>
<th>van Bommel</th>
<th>Yu et al</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Double Blind</td>
<td>Unblinded</td>
<td>Two Arms</td>
<td>Single Arm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Patients (n)</td>
<td>404</td>
<td>44</td>
<td>87</td>
<td>60</td>
<td>29</td>
<td>14</td>
<td>33</td>
<td>68</td>
<td>45</td>
<td>41</td>
<td>123</td>
<td>51</td>
</tr>
<tr>
<td>Age (years)</td>
<td>58±13</td>
<td>62±10</td>
<td>60±12</td>
<td>65±9</td>
<td>67±8</td>
<td>68.3±8</td>
<td>63±11</td>
<td>58±14</td>
<td>68±9</td>
<td>64±13</td>
<td>61±11</td>
<td>63±11</td>
</tr>
<tr>
<td>Male (%)</td>
<td>73</td>
<td>64</td>
<td>71</td>
<td>88</td>
<td>86</td>
<td>71</td>
<td>85</td>
<td>68</td>
<td>84.4</td>
<td>79</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>Ischaemics (%)</td>
<td>54</td>
<td>73</td>
<td>54</td>
<td>NA</td>
<td>76</td>
<td>29</td>
<td>70</td>
<td>60</td>
<td>NA</td>
<td>49</td>
<td>61</td>
<td>49</td>
</tr>
<tr>
<td>Ejection Fraction (%)</td>
<td>27±6</td>
<td>25±6</td>
<td>25±5</td>
<td>28±5</td>
<td>22±8</td>
<td>25±5</td>
<td>22±6</td>
<td>25±7</td>
<td>29±4</td>
<td>25±6</td>
<td>27±7</td>
<td>28±7</td>
</tr>
<tr>
<td>Intervention CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
<td>CRT</td>
</tr>
<tr>
<td>Control</td>
<td>D-ICD</td>
<td>D-ICD</td>
<td>D-ICD</td>
<td>D-ICD</td>
<td>OMT</td>
<td>NA</td>
<td>CRT**</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>QRS (ms)</td>
<td>105±13</td>
<td>105±10</td>
<td>107±12</td>
<td>107±14</td>
<td>92±11</td>
<td>110±11</td>
<td>110±8</td>
<td>102±10</td>
<td>109±9</td>
<td>108±14</td>
<td>106</td>
<td>(98–114)</td>
</tr>
</tbody>
</table>

### Effects of CRT in Narrow QRS arm

- **Reduction in NYHA class**: NA NA NA NA 1.1±1.5 1.6±1.8 0.9±1.4 NA NA 0.8±1.2 NA 0.7±1.5
- **Reduction in QOL score**: 12±21 NA 8 (10 to 1)‡ NA 12.2±25 NA 14±22 24±21* NA 17±26 NA 8±28
- **Improvement in 6MWT (m)**: NA –11.3 (–31.7 to 9.7)† 26 (0 to 46)‡ NA 103±186 94±136 96±152 NA 182±286 48±97 NA 46±122
- **Reduction in LVEDV (ml)**: NA NA 16 (29 to 8)‡ 18 (47 to 6)† 15.4 (44.6)§ NA 27±43 1±35* NA 8±64 12±40 14±29
- **Reduction in LVESV (ml)**: NA NA –6.4 (–18.8 to 5.9)§ 19 (34 to 12)‡ 30 (48 to 11)† 26.2 (66.3)§ NA 40±64 1±26* 71.8±113 9±37 17±56 19±39
- **Reduction in LVEDD (mm)**: NA NA 0 (2 to 0)‡ 3 (5 to 0)† NA 6.2±12 NA 2±5* NA 4±9 NA 0.3±1
- **Reduction in LVESD (mm)**: NA NA 1 (3 to 0)‡ 5 (7 to 2)† NA 5.8±11 NA 0±10* NA 4±8 NA 0.5±1
- **Improvement in LVEF (%)**: NA 3.3 (0.7 to 6)† 1.2 (0.4 to 4.4)‡ 12 (10 to 13)† 6.7 (18)§ 9±10 8±13 0±7* 23.3±37 2±9 6±20 7±15

QMT indicates optimal medical therapy; D-ICD, dual chamber ICD.

* Data shown from European Society of Cardiology Conference presentation (2008).
† Mean (95% confidence interval).
‡ Median (95% confidence interval).
§ Median (interquartile range).
* Effectively non-comparable control group as broad QRS.
do not have formal enrollment and blinding of measurements.\textsuperscript{25}

Our findings are concordant with a recent large meta-analysis of RCTs of CRT, which showed through a spline-based regression analysis a direct relationship between QRS width and prognostic benefit from CRT, with no statistically significant benefit from CRT once the QRS duration falls below approximately 130 ms.\textsuperscript{1}

**Implications for Research**

Highlighting the importance of bias-resistance steps such as randomization and blinding is not novel, having been introduced in 1948 with Hill’s randomized trial of streptomycin in tuberculosis.\textsuperscript{26} The published reports of observational, and incompletely blinded, studies of therapy produce effect estimates that tend to show exaggerated benefits, and can even be in the opposite direction to thoroughly blinded, randomized controlled trials.\textsuperscript{27} Nevertheless, uncontrolled, non-randomized, and unblinded routine clinical data are widely available in every hospital and it is inevitable that such data will enter the literature. Selective enthusiasm to report (and review favorably, and publish) positive data, together called “publication bias,” can further distort the literature towards positivity, unhelpfully.

**Study Limitation?**

The lesson to learn may be that lack of a suitable control arm with randomization and blinding in a study of treatment outcomes reporting measurements acquired through routine clinical practice should not be considered merely a minor “study limitation.” The potential for this bias is so large that a published study in this class cannot be trusted to give even approximate guidance.

Rarely are readers explicitly warned that the measurement (and its associated confidence interval) from non-bias-resistant studies can be so misleading as to get the direction of effect completely backwards. Table 5 lists, for each study lacking bias resistance steps, what it says on the subject.
The weakness of lack of bias-resistance is far worse than the weakness of small size of RCTs. Small RCTs suffer from noise, which is an equal tendency to over- or under-estimate effect sizes. As sample size grows, or as RCTs are synthesized by meta-analysis, the nature of noise is to progressively subside. Bias, on other hand, is in a consistent direction. Thus a non-bias-resistant study cannot safely be used to give a "rough idea" of what a bias-resistant study would find. The roughness of the idea can dominate any genuine effect, and is easily underestimated by authors and reader alike. Moreover, the larger the biased study, the more likely its confidence interval does not contain the true value.

Meta-analysis of the early observational studies used in our present analysis, when taken in aggregate, point us towards physiological benefit. When the question is restudied with bias removed, we see physiological neutrality alongside event-based evidence of harm. This phenomenon has also been seen in observational study designs for balloon pump therapy, which have consistently generated a wide range of results which, if read superficially, may be misleading. Nevertheless it should not be forgotten that, while the impact of such noise shrinks with sample size, the effect of bias does not. In an irony underappreciated by many of us, bias-vulnerable studies are more likely to be falsely statistically significant if large than small.

**Impact on Publication and Interpretation of Unblinded Data**

This analysis is not a criticism of the conduct of the studies listed. One had the primary endpoint of mortality and therefore did not require blinding to measure their primary endpoint without bias. Instead, our study puts a spotlight on what might be the consequences of drawing inferences from data that inevitably become available from studies that are unblinded or do not have a control arm. The problem is not unique to any study's team, but is common to all of us, perhaps as the result of our inevitable conditioning by normal clinical practice, where it may be good practice to portray to patients a favorable picture of their response to intervention, in order to maximize the overall symptomatic improvement in that individual (only part of which is directly mediated by the device).

Our challenge is to build a community understanding that when addressing mechanistic questions we should not rely on unblinded clinical data, originally obtained for individual-patient clinical purposes, to be a suitable bias-resistant basis for correct evaluation of physiological benefit. In our analysis, what the non-bias-resistant clinical data gave was not a feeble and uncertain pointer, but multiple clear, consistent, and statistically significant pointers to benefit, but when tested in a bias-resistant manner these turned out to be wrong.

If the magnitude of unintentional bias can be so large, we should think carefully before reviewing such data as evidence in a scientific forum.

**Hypothesis Generating?**

Uncontrolled, unblinded, non-randomized data are widely available and often examined for features that might suggest an RCT of therapeutic intervention. However, although the sophisticated authors of such an article might fully understand that an observational study is not a recommendation on which to base therapeutic decisions but a highlighting of an interesting area to trial, many readers cannot resist making this intellectual jump.

The particular hypothesis that CRT would be "beneficial in heart failure with narrow QRS" could have been generated directly from the hypotheses in the positive trials for wide QRS, by simply changing one word. Such a method of generation would have produced the hypothesis:

1. in less time;
2. at less cost;
3. without obscuring the fact that it was a hypothesis, and
4. preserving the clear understanding (from broad-QRS trials) of how to test it.

---

**Table 5. Analysis of Single Arm Trials and Acknowledgement of Limitations Associated With Them**

<table>
<thead>
<tr>
<th></th>
<th>Abstract</th>
<th>Methods</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achilli et al</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Bleeke et al</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>ESTEEM-CRT</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Gasparini et al</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>PROSPECT</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>van Bommel</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
<tr>
<td>Yu et al</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
<td>Not mentioned</td>
</tr>
</tbody>
</table>
It is a fallacy that unblinded, non-randomized studies are a useful step before blinded RCTs for assessing treatment efficacy. In the case of CRT:

1. In broad-QRS heart failure, the data that was not designed to resist bias suggested positive benefits. The subsequent blinded, randomized trials indicated benefit.
2. In narrow-QRS heart failure, the data that was not designed to resist bias suggested positive benefits. The subsequent blinded, randomized trials indicated harm.

Thus, the appearance of positive benefits in studies not designed to resist bias, in 1 group predicted genuine benefit in survival, and in the other group genuine harm to survival.

Clinical Implications

When practicing clinical medicine we may find it difficult to resist looking at unblinded non-randomized studies especially when they are numerous or large or both. Nevertheless, we should remember the many instances in which they have been seriously misleading.\(^2\)\(^8\), \(^3\)\(^0\)-\(^3\)\(^2\) We should focus on studies that incorporate vigorous steps to avoid bias, where such studies exist. Where there are no such studies, we should identify this and focus our energies on designing and implementing trials that have these characteristics. They need not be expensive, if we do not load them with compulsory features beyond inexpensive online randomization and a simple incontrovertible endpoint such as all-cause mortality. If the endpoint is a quantitative physiological marker, then expenditure on measuring this without bias is not a waste of resources, but a necessity for preventing the entire trial being a waste of resources.

Study Limitations

Our search strategy might have missed some studies, although we attempted to be comprehensive. Secondly, we focused on physiological variables rather than event endpoints.\(^7\),\(^1\)\(^0\),\(^1\)\(^1\),\(^1\)\(^4\) This was needed to allow examination of all 3 classes of study, since uncontrolled studies are unable to state effects on event rates. Thirdly, some data from the control arm of 1 randomized trial was missing and could not be obtained from the corresponding author, and therefore this trial had to be interpreted as a single arm study.\(^1\)\(^0\) Finally, we could only use the variables provided by the authors. This is valid as long as the authors did not selectively present variables that showed an improvement; however, this susceptibility to bias exists for any reader of such reports. In addition, if this is an explanation for their discrepancy from the randomized trials then it further underlines the importance of preferentially using the prospectively specified endpoints from blinded randomized controlled trials as guidance for clinical decision-making.

Conclusion

Patients with heart failure and narrow QRS complexes appear to show a physiological improvement with CRT, but only in unblinded studies or those without a randomized control arm. When blinding and randomization are implemented the CRT effect on physiological markers is neutral.

This experience of CRT in narrow-QRS heart failure is a particularly clear illustration of the need throughout cardiology to take elaborate steps to prevent inadvertent bias. The standard method of resisting bias is randomization and blinding. This requires planning and may appear to increase cost. However, in retrospect it might have been preferable to try harder at the outset to acquire bias-resistant data.

Sources of Funding

Nijjer is supported by the MRC (G1100443); Francis (FS/10/038), Finegold (FS/14/25/30676), Sohaib (FS/11/92/29122), Shun-Shin (FS/14/27/30752) and Whinnett (FS/13/44/30291) are supported by the British Heart Foundation.

Acknowledgement

The authors are grateful for infrastructural support from the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London.

Disclosures

None.

References

Trial Design and Narrow QRS CRT

Jabbour et al


6. ggplot2: elegant graphics for data analysis. New York: Springer; 2009. [computer program].


Effect of Study Design on the Reported Effect of Cardiac Resynchronization Therapy (CRT) on Quantitative Physiological Measures: Stratified Meta-Analysis in Narrow-QRS Heart Failure and Implications for Planning Future Studies


J Am Heart Assoc. 2015;4:e000896; originally published January 6, 2015;
doi: 10.1161/JAHA.114.000896

The Journal of the American Heart Association is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Online ISSN: 2047-9980

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://jaha.ahajournals.org/content/4/1/e000896

Subscriptions, Permissions, and Reprints: The Journal of the American Heart Association is an online only Open Access publication. Visit the Journal at http://jaha.ahajournals.org for more information.