Multinational evaluation of the interpretability of the iterative method of optimisation of AV delay for CRT

Claire E. Raphael⁎, Andreas Kyriacou, Siana Jones, Punam Pabari, Graham Cole, Resham Baruah, Alun D. Hughes, Darrel P. Francis

International Centre for Circulatory Health, Imperial College London, United Kingdom

Imperial College NHS Trust, United Kingdom

Abstract

Article history:
Received 12 November 2011
Received in revised form 12 September 2012
Accepted 16 September 2012
Available online 13 October 2012

Keywords:
CRT
Cardiac resynchronisation
Biventricular pacemaker
AV optimisation
Iterative
SMART-AV

Background: AV delay optimisation of biventricular pacing devices (cardiac resynchronisation therapy, CRT) is performed in trials and recommended by current guidelines. The Doppler echocardiographic iterative method is the most commonly recommended. Yet whether it can be executed reliably has never been tested formally.

Methods: 36 multinational specialists, familiar with using the echocardiographic iterative method of CRT optimisation, were shown 20–40 sets of transmitral Doppler traces at 6–8 AV settings and asked to select the optimal AV delay. Unknown to the specialists, some Doppler datasets appeared in duplicate, allowing assessment of both inter and intra-specialist interpretation.

Results: On the Kappa scale of agreement (1 = perfect agreement, 0 = chance alone), the agreement regarding optimal AV delay between specialists was poor (kappa=0.12±0.08). More importantly, agreement of specialists with themselves (i.e. viewing identical sets of traces, twice) was also poor, with Kappa = 0.23 ± 0.07 and mean absolute difference in optimum AV delay of 83 ms between first and second viewing of the same traces.

Conclusion: Iterative AV optimisation is not executed reliably by experts, even in an artificially simplified context where biological variability and variation in image acquisition are eliminated by use of identical traces. This cannot be blamed on insufficient skills of some experts or discordant methods of selecting the optimum, because operators also showed poor agreement with themselves when assessing the same trace. Instead, guidelines should retract any recommendation for this algorithm. Guideline-development processes might usefully begin with a rudimentary check on proposed algorithms, to establish at least minimal credibility.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Background

Atrioventricular optimisation of CRT devices following implantation is mandated in the current American [1] and European [2] Society of Cardiology guidelines. The most strongly recommended technique for optimisation remains the iterative echocardiographic method [1,3,4] (Fig. 2). This method has been used in clinical trials assessing morbidity and mortality benefits of cardiac resynchronisation therapy such as CARE-HF [5]. While the physiological explanation for this method and the steps taken to execute it are undoubtedly rational, there are no data under blinded conditions confirming that the decision making steps of the recommended method can be interpreted consistently by clinicians in practice.

Contention has arisen over the use of echocardiographic iterative optimisation following the publication of SMART-AV, a large, well conducted, externally-monitored trial which has shown no improvement in left ventricular parameters following iterative optimisation of AV delay [6]. While one interpretation is that it does not matter what delays are programmed, such a view strains credulity since the only direct physiological effect of CRT – which undoubtedly improves symptoms and survival – is to change intracardiac timings.

1.1. Rationale for recommending the use of the iterative method of AV optimisation

The rationale behind the iterative method is to maximise diastolic filling time by changing the AV delay settings of the CRT device. An AV delay which is long leads to fusion of the E and A waves and results in reduction in blood flow across the mitral valve and consequently LV filling. In addition, LV filling is further compromised by presystolic...
mitral regurgitation, a consequence of the long AV interval. If the AV delay is too short, this will lead to interruption or even absence of active atrial filling (A wave) of the ventricle due to premature mechanical activation of the ventricle. An optimal filling time lies between the two extremes.

The iterative technique involves examination of the transmitral pulsed wave Doppler trace at a range of AV delays and aims to identify AV delay at which maximal mitral valve inflow is observed (Fig. 1).

1.2. The missing study

Calculating the test–retest reproducibility of the iterative optimum faces multiple difficulties, including the possibility of biological variability that can change the optimum between patient sessions [7]. However, a remarkably unexplored aspect of this process is the ability of specialists to agree on which Doppler trace is actually optimal, even when all other sources of variability are abolished.

In this study, we formally tested this fundamental aspect, namely the ability of specialists to execute the decision making process of the iterative process to select the (same) optimum transmitral Doppler trace. By using already-acquired transmital Doppler images, we eradicated the possibility that different specialists were using subtly different probe positions, or that the patient was in a different clinical state. We could assess purely whether specialists agreed on which transmitral Doppler trace was optimal according to the iterative criteria.

Secondly, we tested whether any discrepancy between specialists’ chosen optimal settings was due to differences between their method of implementing the algorithm, or due to the algorithm being intrinsically unworkable for reliable decision making. We did this by presenting each patient’s data to each specialist twice, without alerting them to the duplication.

2. Methods

2.1. Images

The images were anonymised data acquired from serial heart failure patients undergoing AV optimisation at St Mary’s Hospital using a ProSound SSD-5500SV system (Aloka, Tokyo, Japan) with continuous ECG recording. 30 patients with CRT devices underwent echocardiographic optimisation using the iterative method. Their clinical characteristics are shown in Table 1. Images were taken with the patient positioned in the dorsal decubitus or left lateral decubitus position, at passive end expiration. The AV delay was set starting at 40 ms and sequentially lengthened in 40 ms intervals up to 320 ms or until intrinsic conduction was reached. The mitral inflow pattern was recorded at each setting, using pulsed wave Doppler with the probe positioned at the level of the mitral annulus. For each patient, representative 2-beat segments of transmital Doppler data were assembled into a sequence labelled “A”, “B”, “C”, etc., with “A” being the shortest AV delay (40 ms). The traces from one patient are shown in Fig. 2. Depending on the AV delay at which intrinsic conduction was reached, each patient’s dataset was comprised of 6–8 traces at 40 ms intervals.

2.2. Recruitment of specialists

36 specialists with experience of AV optimisation using the iterative method were recruited from the echocardiography sections of two large, international cardiology conferences (n = 22) and from regional echocardiography specialist meetings (n = 14). All were specialists in echocardiography and were experienced in its application to biventricular pacing devices. All were experienced in selecting the AV optimum by the iterative method.

2.3. Iterative protocol

Each specialist was asked to independently assess a series of optimisation image-sets taken from patients. Each set comprised 6–8 transmital inflow patterns taken at progressively longer AV delays. There were 40 traces in total which they assumed to be from 40 patients but were in fact from 30 patients, with 10 patient datasets shown twice. The instructions given were to choose the trace for each patient that the specialist felt represented the most optimal AV setting. Before beginning, each specialist was reminded of the principle of iterative optimisation. As is common in teaching, a trace with an extremely long AV delay and a trace with a too-short AV delay demonstrating truncation were illustrated to refresh the rationale behind the iterative technique (Fig. 1).

Specialists were asked to assess the images as they would in clinical practice and were given unlimited time to assess each of the image-sets. They were allowed to freely review and modify any decision prior to returning their questionnaire. No conferring between specialists was allowed.

All 36 specialists assessed what they believed were datasets from 20 patients. Unknown to these specialists, however, there were not 20 different sets of images but rather only 10 sets which were shown twice (e.g. Patient “1” was exactly the same set of images as Patient “13”), allowing the study to test intra as well as inter specialist agreement.

Of these 36 specialists, a subset of 10 specialists assessed a further 20 traces, from a fresh group of 20 different patients, to test whether the results were similar in a larger patient cohort.

2.4. Statistical analysis

Statistical analysis was performed using SPSS Version 17. Inter and intra specialist agreement was assessed using Cohen’s Kappa test. A kappa value of less than 0.2 is considered to represent poor agreement and above 0.8 to represent good agreement [8]. Kappa values were calculated for each of the pairs of patient images and the mean and standard deviation across all patient images was calculated.

The authors of this manuscript have certified that they comply with the Principles of Ethical Publishing in the International Journal of Cardiology.

<table>
<thead>
<tr>
<th>Table 1 Patient Characteristics.</th>
<th>All patients</th>
<th>Intra-operator substudy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>67 ± 10</td>
<td>66 ± 11</td>
</tr>
<tr>
<td>% Male</td>
<td>77%</td>
<td>50%</td>
</tr>
<tr>
<td>Aetiology: ischaemic/non ischaemic</td>
<td>53%/47%</td>
<td>60%/40%</td>
</tr>
<tr>
<td>NYHA</td>
<td>2.5 ± 0.5</td>
<td>2.4 ± 0.5</td>
</tr>
<tr>
<td>LVEF</td>
<td>35 ± 8</td>
<td>37 ± 7</td>
</tr>
<tr>
<td>Beta blockers</td>
<td>87%</td>
<td>90%</td>
</tr>
<tr>
<td>ACE inhibitor</td>
<td>70%</td>
<td>90%</td>
</tr>
<tr>
<td>ARB</td>
<td>20%</td>
<td>10%</td>
</tr>
<tr>
<td>Spironolactone/Eplerenone</td>
<td>83%</td>
<td>90%</td>
</tr>
<tr>
<td>Warfarin</td>
<td>7%</td>
<td>10%</td>
</tr>
<tr>
<td>Aspirin</td>
<td>60%</td>
<td>50%</td>
</tr>
<tr>
<td>Statin</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>Digoxin</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>Diuretics</td>
<td>83%</td>
<td>90%</td>
</tr>
</tbody>
</table>

![Fig. 1](image_url)

Fig. 1. The iterative method of echocardiographic AV optimisation in the idealised cartoon form often used in teaching. The guidelines recommend selecting an AV delay where the E and the A wave are clearly separated without truncation of the A wave, allowing maximal diastolic filling time. In (1), the E and the A waves are merged, indicating that the AV delay is too long. In (3), the A waves have become truncated, indicating an AV delay that is too short.
3. Results

3.1. Recruitment of specialists

36 cardiologists from 11 countries and 14 institutions were recruited for the study (Table 2). Each completed the questionnaire in full.

3.2. Agreement between specialists

Each specialist was able to identify an optimum AV delay setting for each patient presented that they would be willing to recommend in clinical practice. Some specialists reported that it was easier to make the assessment in this research environment, with no time pressure and extremely easy side-by-side comparisons between traces for each patient. No specialist indicated that choosing a trace in this research environment was more difficult than in clinical practice. Despite having the facility to change previous selections, there were only two cases (of 1000 patient–specialist combinations) where a specialist changed their selection after they moved on to the next patient. The agreement between specialists regarding the optimal AV delay was poor. On the Kappa scale of agreement (1 = perfect agreement, 0 = chance alone), the average kappa value was 0.12 ± 0.08 (Figs. 2, 4, and 5).

3.3. Agreement of each specialist with themselves

While completing the survey, no specialists noticed that the patient datasets labelled 11–20 (which they assumed to be patients 11–20) were familiar, namely exactly the same as datasets 1–10. Therefore each specialist unknowingly viewed each patient’s dataset twice, allowing assessment of intra-observer agreement (Figs. 3–5). The average intra-specialist agreement was poor (kappa 0.23 ± 0.07) with the kappa values and plots for each patient shown in Fig. 5. The mean absolute difference in milliseconds between the first and second viewing of the trace was 83 ± 33 ms (Table 3).

To test whether these results were peculiar to the 10 patient traces selected, a subset of 10 of the specialists assessed a further set of 20 genuinely distinct patient datasets: thus this group of specialists saw a total of 30 genuinely distinct patient datasets. Results in this substudy were similar: the inter-specialist agreement was poor (kappa = 0.06 ± 0.10) as was the intra-specialist agreement (kappa = 0.22 ± 0.14).

![Fig. 2. Doppler traces of an example patient and the selections of which trace is considered optimal by 36 specialists. All 7 Doppler traces of this patient are shown in the upper panel. For this patient, 1 specialist chose delay A (40 ms) as optimal. 4 chose B (80 ms), 10 C (120 ms), 13 D (160 ms), 5 E (200 ms), and 1 F (240 ms). There was a variety of opinions as to what construed an optimised patient trace, with a wide variety of AV delays selected as optimum.]

![Fig. 3. Intra-specialist reproducibility for duplicate images of the same patient: The same sets of images for a single patient were shown to each specialist twice for a total of 10 patients. This allowed assessment of between specialist consistency of implementing the iterative algorithm. The arrows indicate each individual specialist and their choice of optimum AV setting for the two sets of identical images.]
4. Discussion

In this study we have found that, first, specialists familiar with iterative optimisation of AV delay using echocardiography do not agree with each other’s decisions on what the optimum AV delay is, even when they are looking at identical acquired Doppler strips. Second, more crucially, they do not agree with their own decisions when shown identical sets of Doppler traces a few minutes later, even when they are allowed to freely review their decisions with no time constraints. Third, they only very rarely showed uncertainty in the form of changing their mind (0.2%), suggesting that as echocardiographers we have little insight into how difficult and uncertain the task of iterative optimisation truly is.

4.1. Use of the iterative method of AV optimisation

The iterative process involves progressively shortening the AV delay while measuring the transmitral Doppler flow pattern for each stage [1,3,4] (Fig. 1). The decision making instruction is to select the optimal transmitral Doppler flow pattern amongst a set of acquired data may be intrinsically impossible to execute reliably. This study tested this question alone.

4.2. Limitations in following the iterative technique

While the iterative technique has been described in several different guidelines and papers [1,2,4], methods for determining the optimum setting are not clear. Publications show diagrams of an extremely long AV interval with a merged E/A pattern and of a too-short interval with truncation of the A wave (Fig. 1). However in real life practice several AV delay settings may show separation of E and A, without A wave truncation. This may be an obstacle to consistent identification of an optimum.

4.3. Use of Cohen’s kappa test

Cohen’s kappa test allows an assessment of agreement between (or within) operators while adjusting for chance. If two persons are asked to choose between two options and they each select by simply tossing a coin, they would in raw terms agree in 50% of cases; thus any valid measure of agreement must only look at the territory between raw agreements of 50% and 100% which represents the full spectrum from pure chance to perfectly skilful agreement. Cohen’s kappa value adjusts for the likelihood of agreement through chance by scaling that range (from 50% to 100%) into a kappa range that runs from 0 signifying simply chance to 1 signifying perfect agreement.

Fig. 4. Within-specialist variability for 36 operators assessing one patient: Each operator viewed the same image twice. The first specialist (“S1”) thought option B (80 ms) was optimum on the first and second viewings of the trace. Specialist 2 (“S2”) thought B (80 ms) was optimum on first viewing and that E (200 ms) was optimum on second viewing, whereas Specialist 3 (“S3”) thought that D (160 ms) was optimal on first viewing and B (80 ms) was optimal on the second viewing. This plot demonstrates both the intra specialist agreement and the spread of selections of the “optimum trace” for this patient.

Fig. 5. Within-operator consistency of implementing iterative decision making algorithm. Each panel represents a single patient dataset. The same dataset was shown to each specialist twice for a total of 10 patients to test consistency in assessment of the same dataset on the same day. The optimum reported on the first viewing is shown on the X axis, and from the second viewing on the Y axis. The kappa value describes the amount of agreement of specialists with themselves (0 = agreement expected by chance alone, 0.2 or below = poor agreement, 0.8 or above = good agreement).
Table 3
Mean difference between first and second viewings of the same image by the same specialist.

<table>
<thead>
<tr>
<th>Question</th>
<th>Absolute mean difference between 1st and 2nd viewings/ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>103±40</td>
</tr>
<tr>
<td>2</td>
<td>69±26</td>
</tr>
<tr>
<td>3</td>
<td>105±34</td>
</tr>
<tr>
<td>4</td>
<td>69±28</td>
</tr>
<tr>
<td>5</td>
<td>33±15</td>
</tr>
<tr>
<td>6</td>
<td>48±33</td>
</tr>
<tr>
<td>7</td>
<td>73±24</td>
</tr>
<tr>
<td>8</td>
<td>70±30</td>
</tr>
<tr>
<td>9</td>
<td>175±59</td>
</tr>
<tr>
<td>10</td>
<td>84±43</td>
</tr>
</tbody>
</table>

Agreement between specialists was disappointing, with kappa 0.12±0.08. There was only slight improvement for the same operator assessing identical images of the same patient twice (kappa 0.23±0.07). The average absolute difference in optimum AV delay between 1st and 2nd viewing for same specialist assessing the same traces was 83±33 ms. This is wide variation and cannot be disregarded as clinically non-significant.

Our kappa values show that it is not simply that different operators are using different criteria to assess the optimal trace, because if this were the case, the intra-operator kappa (for the same specialist assessing the same trace twice) would be good and only the inter-operator measure poor. Our results show that the iterative method when implemented by experts looking at the same trace twice cannot yield consistent decisions: there is therefore almost zero opportunity for consistent decisions with real-life data which include natural variability.

4.4. Clinical and research implications

The most important clinical implication of this finding is that while the iterative method is easy to describe and recommend, in real life it appears difficult to execute. The problem is not discordance of opinions from different experts, but rather fundamental inability of the instructions to be carried out reliably by anyone.

So what explanations are there for improvements reported with AV optimisation of CRT? First, the placebo effect of any medical intervention means that improvement of symptoms cannot be reliably attributed to the effect of intervention [9,10]. Second, studies of serial assessment before and after optimisation that occurs soon after implantation might capture part of the improvement that comes from CRT itself [11]. Third, studies that involve adjusting settings serially, measuring a physiological response, and then adopt the setting with the highest physiological response, are subject to the illusion of benefit arising from measurement scatter [7]. Fourth, it is well documented in the psychological literature that we systematically overestimate the effect our actions: [12] this “illusion of control” encourages us to believe that we are indeed selecting the best trace. Finally, another well-recognised psychological principle comes into play: “cognitive dissonance”. Every time a specialist carries out an iterative optimisation, they increase their investment in the concept, and become increasingly unable to seriously consider the possibility that it was largely a process of random selection within a constrained range [7].

SMART-AV, a large, clearly-powered, carefully conducted, externally-monitored, prospectively-recruited randomised controlled trial, showed no significant improvement in LV dimensions, quality of life measures or distance walked in 6 min from iterative optimisation of AV delay [6].

SMART-AV does not teach us that AV delay setting is immaterial. Rather, it teaches us that selecting an AV delay setting effectively at random is not significantly better than a fixed AV delay across all patients.

It would be very unwise to test a proposed optimisation scheme against the iterative method as though it were a gold standard: the inability of the iterative method to give a clear optimum would leave such a study exquisitely vulnerable to confirmation bias. On the one hand, iterative optimisation does not agree with itself, and therefore a single execution of iterative optimisation would never be able to agree well with any other method, however good it is. On the other hand, carrying out several executions of iterative optimisation would generate several possible AV optima for each patient: if it is considered acceptable for any of these to be used as the gold standard, then any new optimisation algorithm, however nonsensical, could truthfully be said to agree with the iterative method [13].

4.5. Study limitations

Our study enrolled 36 specialists in echocardiographic optimisation and showed that there was poor agreement between specialists as to what represented an optimal AV delay. While the number of specialists may not seem large (compared to the number of patients expected to be seen in an endpoint trial), this was not a study of patient endpoints but of agreement between specialists. Because each specialist examined many datasets which were also evaluated by many other specialists, a very large number of comparisons is possible between observers: as a result this was able to be a sensitive analysis. We did not ask how frequently the specialists carried out iterative optimisation, and therefore it is possible that some may have been conducting them only infrequently or have been out of practice. However, they were reminded of the principle before they carried out the selections of optimal traces, and were permitted any degree of further reminder if they asked.

We did not study other methods of optimisation and therefore cannot comment on the relative merits of those versus the iterative method. However the present data suggest the iterative method is not sufficiently consistent in identification of the optimum AV delay to make it a credible representative [7] of the concept of optimisation. Other methods should be tested independently for simple properties such as a unique optimum (singularity), reproducibility, and plausibility of the distribution of obtained optima.

This is not a study of test–retest reproducibility, i.e. patients were not required to attend twice with independent datasets. This is because it is difficult to justify on ethical grounds asking patients to enrol for dual visits in a study of test–retest reproducibility of a test whose intraobserver consistency of analysis of identical printouts is already so poor. Difficult thought it may be to stomach, test–retest reproducibility must be worse: i.e. its kappa must be even closer to 0.0 than the kappas described in this paper.

This study may appear to be destructive by highlighting a problem without suggesting a solution. However fields that require a complete solution to be developed and implemented before the corresponding problem can be aired, characteristically make only slow, halting and unreliable progress. This study discloses a problem clearly and provides a simple, cheap, early test of any proposed solution before large-scale clinical implementation is mandated.

5. Conclusions

The decision making process of the iterative method of AV optimisation of CRT pacemakers may not be capable of being executed reliably, even by specialists experienced in the field. This is not due to disagreement between specialists, since the decision making process carried out by the same specialist viewing (unbeknownst to them) the same Doppler images a few minutes later, is just as inconsistent as between different specialists. Nor is it likely that inadequate experience or training is to blame, since all specialists showed similar variation in their choices. Rather, analysis of the 1000 assessments in this study, the most comprehensive blinded study on this subject,
indicates that reliably executing the conventional iterative protocol for AV optimisation of CRT in real life patients is almost impossible.

The iterative method should be retracted from the guidelines, and the guideline development process reformed to include a quick check. For any proposed algorithm, data of a handful of patients could be inspected by all guideline committee members independently to see whether concordant conclusions are drawn. This rudimentary precaution would have prevented the current uncomfortable need for retraction.

**Role of funding source**

DPF is a BHF Senior Research Fellow (FS/10/038).

**Conflicts of interest**

Imperial College, the employer of several authors, has filed patents on technologies that give reproducible optimisation of AV delay.

**References**


