Review

Cardiac resynchronisation therapy optimisation strategies: Systematic classification, detailed analysis, minimum standards and a roadmap for development and testing


In this article an international group of CRT specialists presents a comprehensive classification system for present and future schemes for optimising CRT. This system is neutral to the measurement technology used, but focuses on little-discussed quantitative physiological requirements. We then present a rational roadmap for reliable cost-effective development and evaluation of schemes. A widely recommended approach for AV optimisation is to visually select the ideal pattern of transmitral Doppler flow. Alternatively, one could measure a variable (such as Doppler velocity time integral) and "pick the highest". More complex would be to make measurements across a range of settings and "fit a curve". In this report we provide clinicians with a critical approach to address any recommendations presented to them, as they may be many, indistinct and conflicting. We present a neutral scientific analysis of each scheme, and equip the reader with simple tools for critical evaluation. Optimisation protocols should deliver: (a) singularity, with only one region of optimality rather than several; (b) blinded test–retest reproducibility; (c) plausibility; (d) concordance between independent methods; and (e) transparency, with all steps open to scrutiny. This simple information is still not available for many optimisation schemes. Clinicians developing the habit of asking about each property in turn will find it easier to winnow down the broad range of protocols currently promoted. Expectation of a sophisticated enquiry from the clinical community will encourage optimisation protocol-designers to focus on testing early (and cheaply) the basic properties that are vital for any chance of long term efficacy.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Clinicians rightly look for large endpoint trials to guide therapeutic choices. While for dichotomous choices with large effects (such as implanting a device versus not) this has been effective, for therapeutic decisions with more numerous choices and likely smaller effects this approach in isolation may be inefficient. The process of optimising atrioventricular (AV) and interventricular (VV) delay of biventricular pacing (cardiac resynchronisation therapy, CRT) devices is an example. Well-
conducted bias-resistant long-term trials are expensive and therefore few [1]. If the optimisation methods evaluated in them have not gone through a series of screening steps, the substantial investment may become allocated to strategies that are mathematically or physiologically implausible.

In this report we lay out a rational pathway for the development and testing of optimisation protocols. This is useful both to researchers and to clinicians not considering themselves researchers. We provide key questions to ask of strategies being developed, to encourage early recognition of some strategies that have no chance of ultimate effectiveness.

It does not instruct clinicians on what approach should be used to optimise CRT, because a reliable answer is not yet available. Instead it provides clinicians with questions which may enable them to reject many methods currently proposed to them, by finding the answers to be unsatisfactory or unavailable, or even in some cases just by careful consideration of the quantities involved. The same logical sequence of questions should also be followed by researchers developing a protocol, to prevent waste of research resources.

Small studies have consistently shown acute haemodynamic benefit of atroventricular (AV) and interventricular (VV) delay optimisation [2,3]. However, evidence on the long-term benefit of cardiac resynchronization therapy (CRT) optimisation from large clinical trials has focussed on a few methods and none has been convincingly positive. Guidelines therefore do not provide direction on how to programme CRT [4,5].

Many different variables can be measured to guide the programming for biventricular pacing, ranging from echocardiographic guidance based on diastolic or systolic haemodynamics, to electrogram guidance, as well as blood pressure and its derivatives [6]. In some cases, even after deciding which variable to monitor during optimisation, different protocols may be used to select the best pacemaker setting for that variable. This review systematically classifies the broad strategies for optimisation, provides detailed descriptions of the individual methods, and for each provides a practical perspective (Table 1).

We then describe a series of rational steps that should be performed by any investigator designing and evaluating an optimisation protocol.

Table 1
Common strengths and weaknesses of different approaches to optimisation.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Common strengths</th>
<th>Common weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Spot the pattern</td>
<td>- Easy to describe</td>
<td>- Potentially susceptible to multiple sources of variability (inter and intraobserver, in addition to biological)</td>
</tr>
<tr>
<td>2 and 3: Pick the highest or lowest</td>
<td>- Usually requires maximisation of a relevant physiological parameter</td>
<td>- Very large numbers of beats need to be collected to adequately reduce the effects of variability</td>
</tr>
<tr>
<td>4: Predict the optimum</td>
<td>- Time efficient method for the operator</td>
<td>- No direct measurement of physiologic effect</td>
</tr>
<tr>
<td>5: Fit a curve</td>
<td>- Usually requires maximisation of a relevant physiological parameter</td>
<td>- Substantially conflicting algorithms suggest the majority of these must be wrong</td>
</tr>
<tr>
<td>6: Find the inflection</td>
<td>- Potentially highly reproducible</td>
<td>- Validation and clinical studies have been susceptible to pitfalls in evaluating optimisation technology</td>
</tr>
</tbody>
</table>

Clinicians, when faced with a proposed optimisation protocol, might apply these as a sequential checklist of properties, which would help avoid fruitless endeavour.

2. Approach 1: spot the pattern

2.1. Example: mitral inflow pattern [7]

2.1.1. Also applicable to: TDI VV optimisation [8]

This may be the earliest form of physiological AV optimisation, preceding even the advent of biventricular pacing. Across the range of different AV delays, the clinician chooses the setting which gives the qualitatively most desirable pattern using the measurement tool. The paradigm case is Doppler interrogation of mitral valve inflow [7]. The preferred Doppler pattern is separation of the E wave and A wave on Doppler without truncation of the A wave. This test is performed in the supine, resting patient.

2.2. Protocol

Pulsed-wave Doppler is used to measure transmural flow during diastole. The pattern is recorded at each tested AV delay. There is no guidance on the number of beats to be recorded at each setting. Publications describing the technique typically show just one beat. At each AV delay, the pattern is noted (Fig. 1). Adequate separation of the E and A wave should allow selection of an AV delay which maximises LV filling in diastole. With a long AV delay, there is fusion of the E and A wave, with a short AV delay there is truncation of the A wave. Different protocols have been described to select the optimal AV delay. One algorithm described by Ritter et al. [7] estimates this AV delay by assessing time from the onset of the QRS to the end of the A wave at a long and short

Table 2
International consensus recommendation for evaluation of optimisation technology.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Implications &amp; pitfalls for clinical trial &amp; clinical study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Singularity</td>
<td>The optimisation scheme should provide, one small region for the optimum. Pre-requisite for a clinical trial</td>
</tr>
<tr>
<td>Step 2: Reproducibility</td>
<td>The scheme should pick almost the same value when the entire optimisation acquisition and analysis process is repeated, with blinding. Pre-requisite for a clinical trial</td>
</tr>
<tr>
<td>Step 3: Plausibility</td>
<td>Values selected by the scheme should be physiologically realistic. Advisable if a clinical trial is to have some chance of success.</td>
</tr>
<tr>
<td>Step 4: Clustering</td>
<td>Does the optimisation scheme under assessment choose values similar to another scheme with adequate &quot;SRP&quot;? Test for agreement with the pacemaker setting chosen by another SRP passed scheme. (Do not waste time correlating the physiological measures at the two settings.)</td>
</tr>
<tr>
<td>Step 5: Cluster selection</td>
<td>A scheme with adequate &quot;SRP&quot; may consistently agree with certain schemes but not others. If there is more than one cluster, each is likely to be preferentially optimising one aspect of physiology at the expense of another: pick the cluster maximising the most desirable physiological features (pressure and flow).</td>
</tr>
<tr>
<td>Step 6: Scheme selection</td>
<td>From the chosen cluster of &quot;SRP&quot; schemes, which scheme is most practical? An outcomes trial could be realistically designed at this stage. The control arm could be nominal settings or any irreproducible scheme (since both give some chance of accidentally giving the optimum).</td>
</tr>
</tbody>
</table>
AV delay, and using these figures predicts the optimal AV delay. The algorithm was originally developed for optimising AV intervals during RV pacing with AV block, but has been widely adopted for CRT optimisation. Another approach, the “iterative method”, starts with a long AV delay, and shortens it in 20 ms intervals until there is A wave truncation, and subsequently extends it in 10 ms increments until a desirable pattern is obtained [9]. Another variant of this approach is where a long AV delay is programmed, and the time from the end of the A wave to the onset of systolic mitral regurgitation is deducted to calculate the optimal AV delay [10].

Fig. 2 shows this method in a patient, at three different AV delays: 80 ms, 120 ms, and 160 ms. It shows more than one beat at each AV delay, so that beat-to-beat variability is obvious. While cartoons (Fig. 1) – or example cases commonly used in teaching – may have obvious, dramatic differences between settings, accentuated by selective display of convincingly different profiles (and not showing between-beat variability), in unselected case series the differences between settings are often more subtle [11]. Despite this, the protocol described above is typically recommended for selecting the optimal programmed settings.

A clinician with only the information above and an echocardiography machine can easily test the iterative scheme by a series of quick, low-cost explorations:

Exploration 1 Have two independent staff members conduct separate optimisations on the same handful of patients, blinded to each others’ findings. Is the optimal AV delay very similar (to within 10 or 20 ms) between these two sessions? Time can be saved by having the sessions in immediate succession, but independent data must be collected: re-reading the same scans is not relevant at this stage. If this test–retest reproducibility is strong, the method is at least internally valid. If test–retest reproducibility is weak, then the further explorations below can identify why.

Exploration 2 Eliminate acquisition variability by using a single set of acquired images. Do two mutually blinded reporters reviewing the same images consistently agree on the optimum? If they do, then biological (or equipment) variability between sessions is the explanation for lack of reproducibility. If they do not agree, then either the protocol instructions are being understood differently by different operators, or the protocol is not really an algorithm. A third exploration is required to separate these possibilities.

Two large studies (InSync-III completed in 2005 [12], SMART-AV completed in 2010 [13]) appear to have collected formal test–retest reproducibility of the optimum using a “pick the pattern” strategy, but the results are unpublished. Explorations 2 and 3 for transmitral Doppler [14] suggest that between-observer and even within-observer, same-image agreement is already poor (kappa = 0.23) and therefore test–retest reproducibility (which must additionally include biological variation) must be worse.

3. Approach 2: pick the highest

3.1. Example: LVOT VTI

3.1.1. Also applicable to: impedance cardiography [15], invasive dP/dt\(_{\text{max}}\) [2], mitral inflow E-A VTI [16], mitral inflow E-A duration [16], cardiac output [17,18], stroke volume [19]

Moving beyond qualitative pattern recognition, another approach is to quantify some variable describing the effectiveness of cardiac function, and then pick the pacemaker setting that maximises this variable. This could be applied to a range of physiological measures that can be quantified invasively during the time of device implant, or non-invasively using echocardiography, impedance cardiography, or non-invasive blood pressure measurements. Such measures include peak blood flow, stroke volume, stroke distance (velocity–time integral), arterial or ventricular pressure, ventricular dP/dt\(_{\text{max}}\), and bio-impedance. This category encompasses many different variables where the same overall approach is used to pick the optimum, but each of these have very different signal to noise characteristics, and protocols vary in using a single measurement at each AV delay, repeated measurements, or comparison to a reference state. “Pick-the-highest” schemes can be used for either AV or VV optimisation. The example protocol given here is for AV optimisation using stroke volume (LVOT VTI) measured by Doppler echocardiography.

3.2. Protocol

LVOT VTI can be used for AV optimisation, and is often recommended for VV optimisation [20]. At each pacemaker setting, a sample of

![Fig. 1. Iterative optimisation: protocol. Illustrative example of pattern changes to identify during iterative optimisation.](image1)

![Fig. 2. Iterative optimisation: clinician’s perspective. Real world data: pulsed wave Doppler traces of mitral valve inflow at three different AV delays.](image2)
pulsed wave Doppler images is acquired from the LVOT, while keeping the probe position constant between settings. It is often recommended to average the VTIs of 3 beats, although articles describing the technique commonly show the process being carried out on a single beat per setting [20]. Increments of 20 ms in the AV delay are commonly recommended. The AV delay setting which yields the highest VTI is selected as the optimum. The same approach is used for VV optimisation.

3.3. Clinician’s perspective

The clinical data from one patient in Fig. 3 illustrates several features that are rarely highlighted. First, differences between VTIs at the same setting are not trivial. Second, if only one beat had been measured at each setting, the selection of the optimum would be largely a matter of chance. Third, on any one single-beat dataset, the pattern of VTI against AV delay has not a single peak, but several alternating local maxima and minima [21].

Variability is not peculiar to Doppler: it is present in all physiological variables. In some cases it is almost exactly concordant in simultaneous measurements from sensors using different physical principles from different sites in the body [22], which suggests that it is not caused by equipment or operator error, but instead genuine biological variation between beats. It is preferable to test AV delays in random order rather than sweeping, for example, from short to long. Otherwise, a trend in interpretation variability, and do not critically depend on stable positioning of sensors, but do not eliminate natural biological variability from respiration and other sources, which should be quantified in the protocol-planning phase, in order to determine how many replicates are needed.

To determine the number of repeat measurements that are required, the likely true physiological difference between the two different tested settings is required (for example 2–5%), as well as the variability of measurements being taken (noise, or scatter, for example 5–10%) [23]. To be 90% certain that the correct setting has been chosen, the number of beats required can be calculated [23], which can easily be over one hundred times larger than the three beats usually advised.

4. Approach 3: pick the lowest

4.1. Example: minimising dyssynchrony using echocardiography [27]

4.1.1. Also applicable to: myocardial performance index [28]

Minimising an undesirable characteristic is an alternative to maximising a desirable one. Mathematically the processes are equivalent. This can be applied to markers of dyssynchrony. For those who
believe that the principal mode of action is to correct inter and intraventricular dyssynchrony, this is a logical method for optimisation. However, unlike the other methods which are described here, there are relatively few examples of its use in the scientific literature. This approach has the same formula as Approach 3 for calculating the number of replicates needed which, because of the greater variability [27], can become very large.

Minimising the ratio of the isovolumic times to the ejection times (the myocardial performance index) has also been described [28].

5. Approach 4: predict-the-optimum

5.1. Examples: the electrogram methods: QuickOpt™, SmartDelay™, Adaptive CRT™

Biventricular pacing is implemented by electrophysiologists using devices with a prominent capacity for recording electrograms, and therefore it is not surprising that a variety of algorithms have been proposed to use information related to electrical activation, either from these intracardiac electrograms, or the surface ECG to predict which AV and/or VV delay will deliver the greatest physiological effect [29]. These proposed algorithms differ substantially in their proposed AV or VV optima, which implies that for most of them the claim that they recommend the optimum AV or VV delay must be incorrect (see below). Algorithms based on the surface ECG target the settings which deliver the narrowest QRS complex [29], device based algorithms employ different elaborate formulae to predict the best setting for the AV or VV delay.

5.1.1. Protocol: QuickOpt™

The QuickOpt™ method recommends values for both AV and VV delay [30–33]. For sensed AV delay, the device analyses the duration of the atrial intracardiac electrogram and adds either 60 ms (if electrogram duration < 100 ms) or 30 ms (if electrogram duration > 100 ms). This is a zigzag relationship (Fig. 4), meaning for example that, counter intuitively, all patients with atrial electrogram durations between 100 and 128 ms receive shorter recommended AV delays than those with the shorter electrogram duration of 99 ms. Another counterintuitive feature is that every patient in the region of 70 to 129 ms receives an AV delay recommendation identical to that of some patients that are exactly 30 ms away but different from that of all patients with electrogram durations which are closer.

Why a zigzag relationship is thought to be biologically plausible has never been explained, nor has the data which originated the particular values 100, 60, and 30 ms been made public. The QuickOpt™ VV delay is calculated using the following method [33]:

QuickOpt™ VV is defined as the average of two quantities:

- time delay from RV sensing to LV sensing
- how much longer an LV paced activation takes to reach the RV (conduction time left to right) than an RV paced activation takes to reach the LV (conduction time right to left) (a manifestation of directionally sensitive conduction velocity)

For example if during sensing the RV senses 90 ms before the LV, then the first quantity is +90 ms. Then if LV pacing takes 110 ms to activate the RV, and RV pacing takes 108 ms to activate the LV, then the second quantity is 2 ms. In this example the QuickOpt™ VV formula would give 90 ms + 2 ms = 88 ms, multiplied by 0.5 = 44 ms.

5.1.2. Protocol: Expert Ease for Heart Failure + (EEHF +)™ & SmartDelay™

EEHF +™ (Boston Scientific, Minnesota, USA) calculates the AV optimum from the intrinsic sensed and paced AV intervals, QRS duration, and a series of 6 constants, two of which are chosen based on the position of the LV lead [34]:

\[
\text{Sensed AV optimum} = \text{constant} \times \text{QRS} + \text{another constant} \\
\times \text{sensed AV interval} \\
+ \text{another constant depending on lead position}
\]

\[
\text{Paced AV optimum} = \text{another constant} \times \text{QRS} + \text{another constant} \\
\times \text{sensed AV interval} \\
+ \text{a different constant depending on lead position}
\]

SmartDelay™ (Boston Scientific, Minnesota, USA) appears to be a modification of the EEHF + formula, both of which are said to be derived from the PATH-CHF data in a way that appears not to be publicly described [34,35]. It was used in the SMART-AV trial [13]. The SmartDelay™ algorithm makes recommendations for both sensed and paced AV delays. If the LV lead position is not stored, an automated guess of the position is made using the RV sense to LV sense interval. If the LV sensed electrogram is >40 ms after the RV, the assumption is that the position is made using the RV sense to LV sense interval. If the LV sensed electrogram is >40 ms after the RV, the assumption is that the LV lead is considered to be in a conventional free wall position, otherwise a more anterior position. Depending on the A-LV time and the A-RV time, and the lead position, the SmartDelay™ may also choose to pace only the LV or both LV and RV [35]. An electrogram-based optimisation scheme from this manufacturer was used in the COMPANION Trial [36].

5.1.3. Protocol: Adaptive CRT™

Adaptive CRT™ (Medtronic, Minnesota, USA) is a third electrogram based method [37].

![Fig. 4. Graphical representation of QuickOpt™ AV delay. Optimum AV for the underlying atrial EGM duration is plotted.](image-url)
Initially a set of baseline intervals is measured by the device.

A-RV: atrial sensing (As) or pacing (Ap) to RV sensing
A-Pend: P-wave conduction interval determined as the time from atrial sensing (As) or pacing (Ap) to the end of the P wave in the far-field electrogram
RVs-QRSend: the QRS conduction interval, determined as the time from RV sensing to the end of the QRS complex in the far-field electrogram

A-Pend and A-RVs are quantified in the atrial sensed state and separately in the atrial paced state. The algorithm dichotomises patients into normal AV conduction versus abnormal AV conduction (A-RV > 200 ms, or A-P > 250 ms) [37,38]. The recommended AV delays are then defined in the following way.

If AV conduction is normal and heart rate is below 100 beats/min, LV-only pacing is delivered with AV delay calculated from A-RVs as below.

\[
\text{AV optimum} = 70\% \text{ of } A \rightarrow \text{ RVs if } A \rightarrow \text{ RVs} > 133.3 \text{ ms} \\
or \ (A \rightarrow \text{ RVs} - 40 \text{ ms}) \text{ if } A \rightarrow \text{ RVs} < 133.3 \text{ ms} \\
[39]
\]

In the other cases, i.e. either AV conduction is abnormal or heart rate exceeds 100 beats/min, the algorithm jumps to two different formulae.

For sensed
- Adaptive CRT AV = the smaller of: As → Pend + 40 ms and A → RVs − 50 ms.

For paced
- Adaptive CRT AV = the smaller of: Ap → Pend + 30 ms and A → RVs − 50 ms [40].

The Adaptive CRT VV delay recommendation is calculated from the intrinsic A-RVs and RVs-QRSend intervals by a process which appears to be still confidential, although it is disclosed that shorter intervals lead to the time of RV pacing to be progressively more delayed [37]. Although the VV algorithm itself has not been publicly disclosed, it is believed to delay the right or left ventricular lead by a maximum of 20 ms for all patients in whom the native QRS width is greater ≥80 ms. Again, the nature, physiological justification, and clinical derivation and supporting data, appear to be secrets.

5.1.4. Clinician’s perspective
This approach shares the advantage with programming nominal settings, of being instant, potentially highly reproducible, and effortless, but has the extra advantage of allowing the clinicians to claim to “have carried out an optimisation”. However, these electrogram-based algorithms, and various surface-ECG-based algorithms, fall short of what can reasonably be described as an optimisation.

First, the clinician does not establish directly in the individual patient that the AV or VV setting programmed is delivering the greatest physiological effect: instead the clinician must trust that the algorithmic predicts the correct setting. Second, the three methods give contradictory recommendations as can be seen from their formulae. Third, some of the formulae stretch credulity to its limit. For example, the zigzag shape of Fig. 4 implies a rare phenomenon: a biological response curve with not only a sharp discontinuity, but a double switchback.

Clinicians choosing to trust this prediction approach, without themselves making measurements of cardiac function, are implicitly hoping that (a) electrogram data indeed contain all the required information to identify a physiological optimum, (b) of the 4 mutually contradictory electrogram formulae, only 3 are incorrect while 1 is correct, or (c) the clinician has happened to choose a manufacturer who happens to have the only correct formula.

The secrecy around their origin or even their true nature gives little reason for such trust. Clinicians hoping to gain confidence from the studies used to validate the formulae may be disappointed. They fall into four families:

5.2. Agreeing with methods that do not agree with themselves
Some of the electrogram methods were reported to give the same optima as Approach 1 (spot the pattern) using transmitral Doppler, or Approach 2 (pick the highest) with aortic VTI or left ventricular dP/dt. The weak point in this chain of reasoning is that bias-resistant, independent assessments of the test–retest reproducibility of methods in Approach 1 and Approach 2 are rather scarce, which makes them uncertain gold standards [7,20,33]. A method under test cannot reliably agree with a reference standard that does not agree with itself [41].

5.3. Studies whose results show the opposite of the reported conclusions
Some electrogram methods have been shown to give VTI values that correlate with the highest VTI achievable across patients, but this only shows that VTI varies between patients much more than it varies between AV delays [42]. Indeed the strength of the correlation would become 1 if optimisation had no effect at all (in a noise-free measurement). Strong correlations can indicate simply the smallness of the effect of changing the pacemaker setting (Figs. 8, 9, online supplement), and therefore that the analysis is irrelevant.

5.4. Agreement with physiological optima poor
When the electrogram optima are directly compared to the physiological optima, correlation has been found to be poor [42].

5.5. Clinical endpoint impact has been neutral
The Adaptive CRT™ methods showed non-inferiority in outcome measures to echo based methods for optimisation (spot the pattern for AV, pick the highest for VV) [43]. Non-inferiority in outcome measures does not guarantee that either is an effective optimisation technique. It may signify that both might be similarly inadequate at optimising, or both similarly adequate.

A simple test of the plausibility of the predict-the-optimum methods is that they should agree with each other, but this is difficult to confirm because they are implemented on different pacemakers. An alternative is to compare their formulae, but these – when disclosed – appear to be substantially different. Independent studies with blinded, bias-resistant design (see Section 8.1.2), showing in individual patients the spectrum of differences in measures of cardiac function between AV programmed to nominal and to predicted optimum, are needed to underpin these methods.
6. Approach 5: fit a curve

6.1. Example: finger photoplethysmography [3]

6.1.1. Also applicable to (potentially): all the pick the highest approaches

The concept of an “optimum” AV delay implies that there is a region on the spectrum of AV delay where cardiac function is good, and when AV delay is changed in either direction function becomes progressively worse. Biological relationships of this nature are typically curved, with small departures from the peak having only small effects on the physiological variable, but progressively larger departures having effects that grow proportionally to approximately the square of the distance. Such curved relationships to a first approximation can be described by a parabola [44], which can be fitted to observed data using any standard spreadsheet software. This “fit a curve” approach [3,45] can be used for any physiological measurement that the clinician wishes to maximise, such as blood pressure, aortic VTI, or impedance cardiography. Invasive measurements were used with this approach in the early trials of CRT [2,46,47]. It can be used as a direct replacement for the pick-the-highest approach. Two settings such as AV and VV delay could even be optimised simultaneously by testing a grid of combinations and plotting the surface of haemodynamic response, so that the peak of the resulting dome indicates the optimum combination of AV and VV delays [3,24,48].

6.2. Protocol

Fig. 5 shows an example of an AV optimisation using non-invasive blood pressure monitoring [45].

The protocol begins like pick-the-highest, but then involves fitting a curve, which may be a simple parabola or require a more complex shape [24,26,49] if very long AV delays are covered that take physiology into a plateau region. Instead of picking the setting which gives the highest measurement, the peak of the curve is calculated, which may be in between two tested settings. The confidence interval of the optimum can also be estimated [24,26,49,50].

Raw datasets that show multiple peaks and troughs (Fig. 10) are treated by the curve-fitting approach to be uninformative noise. This is in contrast to a naive interpretation that every difference between measurements at two settings represents biologically important differences between the settings. A second undesirable result is a parabola oriented upside down since, in such a dataset, noise has overwhelmed signal. While such an individual dataset can easily be rejected, if this happens often it should be remembered that in an equal number of cases noise will have overwhelmed signal, but by chance the resulting parabola happened to be oriented in the expected direction (Fig. 10) [58].

Uninformative data sets can be exposed either by repeating the optimisation and showing no relationship between successive results, or by calculating the confidence interval of the optimum and finding it to be unacceptably broad [24,26,49]. Optimisations of AV and/or VV delay can be presented with a 95% confidence interval of the optimum, to give the reader an idea of the degree of precision achieved. Curve-fitting allows this to be established from a single optimisation [50], from the variability between replicate physiological measurements (expressed by their standard deviation or “scatter”) and the curvature of the response [23]. The curvature is expressed in physiological units of response (e.g. mm Hg, mm Hg/s, ml, or %EF) per ms², if the AV or VV delay is measured in ms. Curvature is the quadratic coefficient (the coefficient of \( x^2 \)) in the curve that fits response to AV or VV delay as shown in Fig. 5. If the dataset is width ms wide, and it is desired to know the optimum to with a standard error of precision ms, and the dataset can be positioned to straddle the optimum approximately centrally, then the total number of individual measurements required (number of settings × number of replicates) can be planned by the following formula [50]:

\[
\text{Total number of measurements required} \approx 3 \left( \frac{\text{Scatter}}{\text{Width} \times \text{Precision} \times \text{Curvature}} \right)^2
\]

This can be demonstrated with reported values for pressure [50]. Taking each sample as a ten beat average, where scatter is 3.9 mm Hg, a range (width) of 160 ms of AV delays is tested (i.e. 80–240 ms), curvature 1194 mm Hg/s², with a 95% confidence interval of ±10 ms, i.e. a standard error of 5 ms to identify a programmable AV delay. Using these values, a minimum of 50 measurements is required. If the scatter is doubled in this example, the number of measurements required will increase four-fold to 200.

6.3. Clinician’s perspective

This approach has much in common with pick-the-highest. It differs only in fitting a curve to identify the optimum, which if conducted on paper alone makes it slightly more complicated. If the measurements are being documented electronically then curve fitting requires no additional effort.

Fit-a-curve and pick-the-highest have never been tested head to head for their ability to identify the optimum reliably (i.e. with good test–retest reproducibility). It is possible that they may behave differently in response to changes in protocol. For example, testing additional settings under the pick-the-highest approach might increase the chance of picking the wrong setting (because there are more wrong settings) [23], but under the fit-a-curve approach might decrease the chance of picking the wrong settings (because additional data would improve the precision of determination of the optimum) [50].

There are some theoretical advantages of fit-a-curve. First, it can interpolate optima in between tested settings. Second, it automatically identifies some datasets that have too little signal for the amount of noise, so that an extended dataset can be acquired [50]. Third, the habit of formally measuring the curvature (and biological noise) may assist in protocol planning [50]. Finally, the estimated location of the optimum is based on the entire ensemble of data so that measurements at settings far from the optimum (where signal is larger) can contribute to its localisation.

Although fit-a-curve is sometimes used in research optimisations [3,24,26], it is not standard clinical practice [50].
7. Approach 6: find the inflection

7.1. Example: peak endocardial acceleration [7,51]

More recently a family of approaches has been introduced which involves measuring a variable whose response to changes in settings is sigmoidal, with a plateau of low values at one set of extreme settings, and a plateau of high values at the other extreme, and progressive changes in the intermediate zone. Often the point of inflection is defined as the optimum, as is the case for the SonR™ method (Sorin Biomedical, Milan, IT), which uses a heart sound sensor in the lead tip [51] (Fig. 6). It is not entirely clear why this middle value of heart sound should be considered desirable for cardiac resynchronisation.

Difficulty with these inflection based schemes is that any variable which has a plateau of low values at one extreme of AV delay, and a plateau of high values at the other extreme, will have a point of inflection in between, but this gives no reassurance that this setting has any desirable physiological characteristics. For different variables in the same patient, this inflection point might easily occur at different AV delays which undermine the belief that the inflection point of any particular variable is “the” optimum.

A thought experiment illustrates the difficulty in putting one’s trust in an inflection-based scheme. Fig. 7 shows a family of inflection schemes each of which measures one variable that has one plateau level at one extreme, a different plateau level at the other extreme, and a progressive change in between. In each case there will be a mathematically discoverable point of inflection, but the different points of inflection have no reason to agree. That the point of inflection on a single variable is reproducible, therefore, does not give reassurance that cardiac function is maximal at that setting.

In some circumstances, it may be rational to seek the halfway point between two plateaus of a measured variable. For example if it is desired to have the ventricle half captured by pacing, and half activated natively, then it may be rational to seek the AV delay setting that gives a vectorcardiogram halfway between its fully paced and fully native states. Such concepts have been tested against separate haemodynamic measurements [52,53].

8. Part B

8.1. Recommendation for an efficient approach for evaluating optimisation protocols

8.1.1. Need for a new approach

Clinical trials remain the gold standard to test the efficacy of a therapeutic intervention. However, it is unrealistic to skip directly to endpoint clinical trials for all potential optimisation schemes because the cost of these is very large and therefore the number of methods that can be tested adequately is very small.

Negative trials of unreliable methods may create an impression that it does not matter what AV or VV delay settings are programmed. However, there are many possible interpretations:

(a) Choice of AV and VV settings other than nominal really has no effect. If all settings are just as effective as nominal, then all settings must be just as effective as each other. If CRT really does work equally well regardless of programming, then the belief that it works by a sophisticated synchronisation effect is false.

(b) AV and VV delay do matter, but the effect of the choice of these is much smaller than the effect of switching on CRT with any setting because the baseline unsynchronised state is so far below optimal that any AV/VV combination in the reasonable zone is better than no CRT, but not much different from each other. A more sensitive marker of function, than clinical outcomes, might be able to detect it.

(c) AV and VV delay might matter for some patients, but have little or no impact on most patients. While this hypothesis is easy to propose, it is difficult to test unless there are reproducible measures of response within individual patients, which are only just emerging now [54,55]. There is some evidence from the SMART AV trial that optimisation using that one particular “predict the optimum” approach works more effectively in individuals who have a long Q-LV time [56].

(d) AV and VV delay may matter, and a particular method might in principle give an unbiased assessment of the optimum combination but with a great deal of random noise, i.e. irreproducibility due to a poor signal to noise ratio. This can be introduced by biological sources of beat to beat variability such as respiratory variation, which is why some attempts are made to reduce this by ensuring measurements are made in the same phase of respiration. A large number of estimates of the optimum for one patient would yield results whose average accurately defined the optimum but whose individual values might be widely scattered and therefore erroneous. Protocols that use too few replicates might be like this, often inadvertently recommending a random setting. Additionally, optimal device settings may change with physiologic state, posture, heart rate and type of activity; but answering such questions can only begin once reproducible optimisation is available.

(e) The particular method might include an element of bias which is specific to each patient (as well as noise). For example, in Patient 1 the average of very many obtained optima might be 40 ms shorter than the true optimum while in Patient 2 the obtained optimum might average 50 ms longer than the true optimum. In this case, then no amount of averaging will resolve or even expose this: comparison with other candidate optimisation schemes is essential.

(f) The candidate optimisation method might contain no information at all. In a thought experiment, an example might be random selection amongst the range of plausible values. In this case, failure of an endpoint trial is guaranteed but casts no light on the concept of optimisation other than that this candidate method is not it.
8.1.2. Pitfalls to avoid when evaluating optimisation methods

New protocols for optimisation typically undergo clinical testing. There are several common errors that research planners can easily fall into, which can create artificially large appearances of the efficacy of optimisation. Clinical readers should watch out for, and disregard, reports based on these types of fallacious reasoning.

8.1.2.1. Mistaking noise for benefit. The biggest trap is to misinterpret random variability in measurements as evidence of having delivered benefit [55]. The “pick the highest” approach is particularly vulnerable. Natural biological variation will ensure there is always a highest value (even if changes in AV delay have no underlying effect). This will always be higher than (or equal to) the value at nominal settings. Across a group of patients, this will produce a highly statistically significant, but spurious, p value. The p-value is only identifying, correctly, that random variability in measurements as evidence of having delivered benefit. (Fig. 9, online supplement) explain this problem.

Avoiding this bias is easy and quick. Once the measurements at the various settings have been made, the setting with the highest value is defined as the proposed optimum. Re-measurements are then made for just the reference and the proposed optimum setting, by a blinded observer. The increment now seen between reference and pre-defined optimum in these re-measured values represents an unbiased quantification of the effect of optimisation.

8.1.2.2. Mistaking large between-patient difference for information about optimisation reliability. A second common trap is to mistake a high correlation between VTI_{AVdelay1} and VTI_{AVdelay2} across patients as evidence of successful optimisation. Such correlations are always high when the between-patient difference in VTI is much larger than the between-setting difference, even if the proposed optimisation scheme is simply randomly selecting a setting. (Fig. 9, online supplement) show this error in detail. Researchers should correlate not VTIs, but AV delays.

8.1.3. Roadmap for way forward in developing and evaluating optimisation protocols

We should avoid repeating the considerable expense of clinical trial such as SMART-AV [13] and FREEDOM [32] that despite meticulous conduct showed no objective endpoint benefit. Commercial pressure to move directly to endpoint trials, without opening the physiology of the confidential methods to normal scientific discourse, may be partly to blame.

In SMART-AV, the blinded test–retest reproducibility of iterative Doppler optimisation appears not to have been formally explored, the justification for the SmartDelay™ formula was not open to scientific enquiry, and whether it even acutely optimised cardiac function appears never to have been independently tested.

We should only embark on large clinical trials when we have actively chosen optimisation schemes that consistently withstand open scientific critique in reliably-measured physiological studies under blinded scientific conditions.

It may be tempting [4] to recommend limiting optimisation to patients who appear to have not responded to CRT with nominal programming, but this has two undesirable consequences. First, patients whose physiology is such that there is no possibility of improvement by CRT, might be obliged to undergo unnecessary additional clinical manipulations. Second, patients who are fortunate enough to have a partial physiological response, or an optimistic outlook that biases them to report symptomatic improvement, would be denied the opportunity for further physiological benefit. It would be better to develop a reliable understanding of the physiology of optimisation based on reproducible physiological methods, and only then make a judgement on how the provision of optimisation should be restricted.
Evaluating optimisation schemes should begin with quick, inexpensive experiments that can be used to improve basic protocol properties such as reproducibility, and to abandon avenues of optimisation that never fulfilled rudimentary requirements of a decision-making algorithm. Progressively more elaborate experiments can then be carried out, on progressively fewer optimisation algorithms. A small number of algorithms with strong internal validity will survive. If they all give concordant optima, then the clinician is free to choose whichever is most convenient. If they give discordant optima (beyond their own test–retest variability) then they will likely form a small number of clusters, with each cluster optimising one aspect of physiology at the expense of another. In that case the clinician can then choose which aspect of physiology (such as pressure, flow, intensity of heart sounds, etc.) should take priority. If there is dispute, an endpoint trial could be carried out. This selection process can be described by a series of steps that each optimisation scheme can be taken through, beginning with the simplest tests that can be done in a few minutes in one or two patients that optimisation is poorer (otherwise an optimisation scheme that reports for every patient that the optimum lies in the range AV 40 ms to 300 ms, would be considered perfect). It should not be expected for non-adjacent settings to be optimal, with intervening non-optimal settings. The optimal setting or range of settings should therefore be singular.

8.1.5. Step 2: reproducible?
Singular schemes can then be tested for reproducibility. (Schemes which give non-singular optima or a very broad optimum range in individual patients can be discarded before this stage.) If the same patient were to undergo a separate optimisation in the same physical state, by staff blinded to the original findings, the second optimum AV or VV delay should be nearly the same as the first (Fig. 10: panel b). The investigators should not give in to the normal clinical temptation to peek at the previous value [57], because doing this destroys the scientific value of the experiment. If the optimum value changes between datasets acquired in the same clinical state, then either the optimum is truly changing rapidly (in which case there is no point carrying out optimisation) or the optimisation scheme is unreliable.

The simplest description of the reproducibility of the optimum is the standard deviation of differences between successive AV or VV delay optima. A more advanced description (to prevent an optimisation scheme appearing to be perfect by simply reporting an identical value across all patients) is the intra-class correlation coefficient of the optima across patients.

8.1.6. Step 3: is the value plausible?
Singularity and reproducibility alone do not give reassurance that optima are physiological. For example, defining the optimum AV delay in milliseconds as the patient’s height in centimetres is singular and reproducible, but biologically implausible. Likewise a scheme that

<table>
<thead>
<tr>
<th>VTI measured at various AV delays</th>
<th>Apparent VTI optimum</th>
<th>Apparent VTI increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>Pat 1</td>
<td>229</td>
<td>223</td>
</tr>
<tr>
<td>Pat 2</td>
<td>747</td>
<td>695</td>
</tr>
<tr>
<td>Pat 3</td>
<td>479</td>
<td>541</td>
</tr>
<tr>
<td>Pat 4</td>
<td>647</td>
<td>635</td>
</tr>
<tr>
<td>Pat 5</td>
<td>801</td>
<td>735</td>
</tr>
<tr>
<td>Pat 6</td>
<td>153</td>
<td>120</td>
</tr>
<tr>
<td>Pat 7</td>
<td>180</td>
<td>174</td>
</tr>
<tr>
<td>Pat 8</td>
<td>43</td>
<td>433</td>
</tr>
<tr>
<td>Pat 9</td>
<td>372</td>
<td>342</td>
</tr>
<tr>
<td>Pat 10</td>
<td>111</td>
<td>132</td>
</tr>
<tr>
<td>Pat 11</td>
<td>844</td>
<td>894</td>
</tr>
<tr>
<td>Pat 12</td>
<td>779</td>
<td>811</td>
</tr>
<tr>
<td>Pat 13</td>
<td>565</td>
<td>585</td>
</tr>
<tr>
<td>Pat 14</td>
<td>200</td>
<td>183</td>
</tr>
<tr>
<td>Pat 15</td>
<td>148</td>
<td>128</td>
</tr>
</tbody>
</table>

Mean increment in VTI +35
SD of increment in VTI 26
p = 0.00014

Fig. 8. Simulation to demonstrate how noise can be mistaken for benefit during optimisation. This tool is available as an editable spreadsheet in the data supplement. It may be used not only for VTI but for any variable which is under consideration for an optimisation protocol. The user enters a standard deviation to indicate the variability of the parameter within an individual, followed by the maximum and minimum plausible difference in values between two individuals. Noise is easily mistaken for benefit by simply picking the highest value without consideration of underlying variability, and a false impression of statistical significance is given when increments are averaged across a group of individuals.
always defined the optimum VV delay in females as RV-first by 60 ms, and males as LV-first by 80 ms would also be singular and reproducible, but not biologically plausible (Fig. 10, panel c).

In the examples above, implausibility is immediately evident, but in other cases it may be contentious. For example, a scheme may recommend that on the transmitral Doppler, the area under the curve of the A wave (including any overlapping E wave) should be measured, and that the optimum is not the setting that minimised it (e.g. AV 0) or that maximised it (AV so long as to cause intrinsic conduction), but the setting which is half-way in between, or at the point of inflection.

Many variants of such schemes could be proposed using the numerous available variables, producing a spectrum of contradictory optima (Fig. 7).

Any scheme which consistently chooses an extremely long or short AV delay is implausible, but equally a scheme which arbitrarily chooses a value that is a fixed portion between the two extremes could also be doubted.

Plotting the distribution of optima obtained may give an additional clue to plausibility, since (for example) simultaneous atrial and ventricular activation, can be detected as implausible [58].
If a singular, reproducible scheme passes the step of plausibility, it can be admitted to the “elite” of schemes, which can then be tested for clustering. Schemes that are non-singular, irreproducible or implausible, need not go forward for testing for clustering because this would be a waste of resources (including patient time).

8.1.7. Step 4: clustering of schemes

Singular, reproducible, and plausible optimisation schemes are still not yet necessarily ready for large-scale trialling. There are simple, cheap tests that can still winnow out unsuitable methods.

Since they all claim to have identified “the” optimum, they should all report the same setting (Fig. 11). Each scheme will have an error bar in its determination of the optimum, which is known from the test–retest reproducibility studies (Step 2) and therefore the only discrepancy beyond this need be considered significant disagreement.

The schemes will fall into one or more clusters. For example, schemes that aim to maximise systemic arterial blood pressure will tend to identify the same optimum regardless of how pressure is measured. Meanwhile, if there were several schemes that (for example) maximised peak velocity of the tricuspid valve, they would tend to identify the same optimum, and this optimum may be rather different from that obtained by the arterial-pressure-maximising schemes. Optima based on intracardiac measurements need not necessarily agree with each other, since there are many potential variables, and maximisation

Fig. 10. Optimisation schemes should be (a) singular, (b) reproducible, and (c) plausible.

Fig. 11. Identifying “clusters” of concordant schemes: In this sketch, one patient undergoes optimisation by 4 schemes, and the results are two clusters. One cluster is schemes 1 and 2: they agree with each other in this patient. A separate cluster is schemes 3 and 4: they agree with each other (but not with schemes 1 and 2). This is reproduced in three further patients.
of one may be at the expense of another. Outside of the heart, however, there are fewer opportunities for biventricular pacing to manipulate different variables discordantly, since the heart ejects its stroke volume and the observed pressure is the consequence of this, and all extracardiac variable arise from pressure or flow or both. Measures such as systemic pressure, cardiac output, and cardiac power output might therefore have mutually consistent optima [59,60].

8.1.8. Step 5: choosing a cluster

It is not known whether there will be only one cluster of optimisation schemes, or several separate clusters. If there are separate clusters of schemes, clinicians should select which cluster represents the variables they believe should be maximised at the expense of the variables in the other clusters. It may be that the choice is obvious. If not, an endpoint trial would at that point be justified, and could confidently be carried out between two very reliable schemes in separate clusters (i.e. schemes that disagree with each other but are each individually singular, reproducible, and plausible).

8.1.9. Step 6: choosing an optimisation scheme

After these 5 stages, there will be a single cluster of optimisation schemes, each individually being singular, reproducible, and plausible. Each scheme would in effect be already validated against all the others in that cluster. Since all schemes in the cluster reported the same optimum, any could be chosen for clinical use, perhaps based on cost or convenience.

No endpoint trials would be carried out until this stage, unless at stage 5 a trial was needed to decide which cluster to reject. An endpoint trial could now be rationally planned. If several important physiological variables were consistently maximised simultaneously by an optimisation scheme, then choosing a scheme from a contradictory cluster would be deliberately choosing to depress several physiological variables below their maximum, which would need careful justification. Identifying these clusters might therefore permit elimination of many schemes before having to conduct long term studies.

Application of these steps would arrive automatically at one or more mutually consistent suitable optimisation methods before any single major expenditure on endpoint trials. Steps 1 to 5 are inexpensive, and could be carried out in multiple independent environments, with results only being considered verified when independent studies concur.

9. Conclusion

Clinicians planning to actually carry out optimisation must select amongst the five optimisation strategies recommended (spot-the-pattern, pick the highest or lowest, predict the optimum, fit a curve, find-the-infection) and the range of measured variables recommended. Published protocols and even guideline recommendations cannot necessarily be trusted because study authors may inadvertently exaggerate the utility of their locally-favoured methods, and guideline preparation does not involve checking whether protocols are reliable, reproducible, or even possible [4,5]. In the absence of trustworthy protocols, clinicians may feel forced to fall back on arbitrary processes for conducting optimisation, but it is difficult to recommend this since such adjustment might just as easily make physiology worse as better. Readers of this document now have a rational process for selecting between optimisation schemes, or several separate clusters. If there are separate clusters, clinicians should select which cluster represents the optimum of clinically satisfactory precision. If ultimately an approach to maximise a physiological marker is taken, fitting a curve does allow comprehensive use of the acquired data, but even still sufficient time must be allocated to achieve a narrow confidence interval, otherwise the optimisation protocol may make up to half of patients worse.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijcard.2013.10.069.

Funding sources, disclosures, and conflicts of interest

DF, SMAS, and ZIW are supported by research grants from the British Heart Foundation (FS/10/038/28268, FS/11/92129122, BRAVO SP/10/ 002/28189, FS/13/44/30291). DF is consultant to Medtronic and Sorin. ZIW acts as a consultant to St Jude Medical. H.M.S is a founder of and unpaid consultant to Strategic Pacing Systems, LLC, and Clinical Scope Principal Investigator for the Columbia-Couter Translational Research Partnership: An Intracardiac Localization System to Cannulate the Coronary Sinus for Left Ventricular Lead Insertion. MM has received speaker fees from Medtronic, Boston Scientific, and St Jude Medical. BvG acts as a clinical advisor and is involved in training and education for pacing and CRT at Medtronic, a consultant to St Jude Medical RADI pressure wire systems, and provides training and education for pacing and CRT at St Jude Medical. FP has received research grants from Medtronic, EBR Systems, Proteus Biomedical, Biological Delivery Systems, MSD. KE has received fellowship support from Medtronic, Boston Scientific, and Biosense Webster, acts as a consultant to Boston Scientific, Medtronic, St Jude Medical, Biosense Webster, Janssens, Pfizer, Daiichi Sankyo, and has received honoraria from Medtronic, Boston Scientific, and Biotronik. CL was principal investigator of REVERSE and is principal investigator of the Miracle–EF study sponsored by Medtronic. She reports research grants, speaker honoraria, and consulting fees from Medtronic, and speaker honoraria and consulting fees from St. Jude Medical. She receives consulting fees from CardioMems. JvG received a ‘Dr. E. Dekker Grant for Research Fellow in Pediatric Cardiology’ from the Dutch Heart Foundation (Grant NHS-2010T07B). CS has received research sponsorship from Biotronik, St. Jude Medical, and Medtronic.

Acknowledgements

The authors from Imperial College acknowledge support from the National Institute for Health Research Biomedical Research Centre at Imperial College London and Imperial College Healthcare NHS Trust.

References


