Heart Fail Rev (2011) 16:277–290
DOI 10.1007/s10741-010-9203-5

When is an optimization not an optimization? Evaluation of clinical implications of information content (signal-to-noise ratio) in optimization of cardiac resynchronization therapy, and how to measure and maximize it

Punam A. Pabari · Keith Willson · Berthold Stegemann · Irene E. van Geldorp · Andreas Kyriacou · Michela Moraldo · Jamil Mayet · Alun D. Hughes · Darrel P. Francis

Published online: 26 November 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Impact of variability in the measured parameter is rarely considered in designing clinical protocols for optimization of atrioventricular (AV) or interventricular (VV) delay of cardiac resynchronization therapy (CRT). In this article, we approach this question quantitatively using mathematical simulation in which the true optimum is known and examine practical implications using some real measurements. We calculated the performance of *any* optimization process that selects the pacing setting which maximizes an underlying signal, such as flow or pressure, in the presence of overlying random variability (noise). If signal and noise are of equal size, for a 5-choice optimization (60, 100, 140, 180, 220 ms), replicate AV delay optima are rarely identical but rather scattered with a standard deviation of 45 ms. This scatter was overwhelmingly determined ($\rho = -0.975$, $P < 0.001$) by Information Content, $\frac{\text{Signal}}{\text{Signal+Noise}}$, an expression of signal-to-noise ratio. Averaging multiple replicates improves information content. In real clinical data, at resting, heart rate information content is often only 0.2–0.3; elevated pacing rates can raise information content above 0.5. Low information content (e.g. <0.5) causes gross overestimation of optimization-induced increment in VTI, high false-positive appearance of change in optimum between visits and very wide confidence intervals of individual patient optimum. AV and VV optimization by selecting the setting showing maximum cardiac function can only be accurate if information content is high. Simple steps to reduce noise such as averaging multiple replicates, or to increase signal such as increasing heart rate, can improve information content, and therefore viability, of any optimization process.

Keywords Cardiac resynchronization therapy · Biventricular pacemaker · Optimization · Echocardiography · Velocity–time integral · Blood pressure

P. A. Pabari (✉) · K. Willson · A. Kyriacou · M. Moraldo · J. Mayet · A. D. Hughes · D. P. Francis
International Centre for Circulatory Health, St Mary's Hospital and Imperial College, 59-61 North Wharf Road, W2 1LA London, UK
e-mail: punampabari@doctors.org.uk

B. Stegemann
Medtronic Bakken Research Center, Medtronic Inc, Maastricht, The Netherlands

I. E. van Geldorp
Department of Physiology, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University, Maastricht, The Netherlands

Background

After implantation of a resynchronization device (biventricular pacemaker or defibrillator) not all patients undergo optimization even though guidelines recommend that AV and VV delay should be optimized, and even though clinical trials have only demonstrated survival benefit of individually optimized CRT. Are clinicians right to cut corners from the trial-validated, guideline-mandated process? To answer this, the basic science of optimization needs to be examined.

For optimization of atrioventricular (AV) delay, commonly a range of AV settings is tested, whilst monitoring a marker of cardiac function such as echocardiographic velocity–time integral [1, 2] (VTI, a surrogate of stroke volume [3]) or left ventricular dP/dt [4, 5]. The pacemaker setting that gives the best cardiac function is then defined as the optimum. A similar process can also be carried out

for the delay between activation of left and right ventricular leads (VV delay).

However, every measurement has uncertainties, which might conceal the true optimum. This uncertainty in our measurement of VTI (or of any other marker for monitoring cardiac function [6, 7]) arises from numerous factors including natural biological variability [8]. Therefore, repeating the "optimization protocol" often provides different optima, as shown in Fig. 1.

There are several clinically important questions. First, if the optimum is not necessarily the 'true' underlying optimum, can we at least express its precision, for example, as a 95% confidence interval?

Second, can we trust the measured increase in VTI as a good estimate of the 'true' average underlying increase in VTI?

Third, if optimizations 6 months later show that many patients' optima have changed, would this imply that patients require more frequent re-optimization? [7, 9]

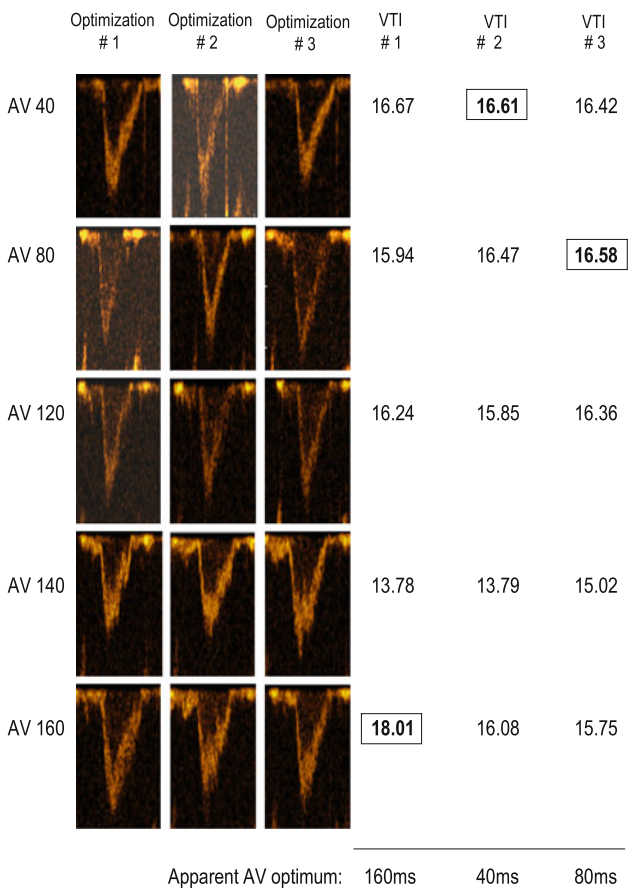

Fig. 1 An example of clinical data from typical patient undergoing three separate Doppler optimization processes (#1, #2 and #3) a few minutes apart, using one heartbeat of velocity–time integral as the measurement to be maximized. In this patient, the small differences in velocity–time integral between the three optimization processes are enough to cause different AV delay settings to be identified as apparently optimal on the three occasions

Finally, how can the precision of the optimization protocol be maximized?

It would be difficult and contentious to attempt to answer these questions by doing clinical studies. This is partly because in clinical practice it is normally assumed that the apparent optimum is indeed the true optimum (or at least the nearest of the tested settings to the true optimum). Persons other than the operator conducting the optimization process itself rarely entertain the possibility that spontaneous variability of the monitored measurement during the optimization procedure arising from beat-to-beat variability and inherent measurement uncertainty has caused the optimum to be misidentified. Confidence intervals are not reported for individual clinical patients' optima [1, 2, 4, 10].

In this study, therefore, we created mathematical simulation having properties exactly like real-life studies, but in which we could truly know the underlying optimum, despite the presence of overlying noise. To understand the realistic balance between underlying optima and overlying noise, we looked at published studies of optimization.

Information content

A convenient way of quantifying in real-life optimizations the relative contributions of underlying true signal information versus overlying random noise (illustrated in Fig. 2) is using "information content". Signal, in this context, is the genuine underlying between-setting difference in VTI, which for computational convenience can be expressed as a variance (average of the squared deviate between the underlying value of each setting and the mean of all settings). Noise, correspondingly, is the unwanted variability that occurs when measures are repeated at the same setting. This too can be expressed as a variance (average of the squared deviate between individual replicate measurements at a setting and the underlying value of that setting). The advantage of using variances is that their sum is the total observed variance. The variance observed over a series of settings can be decomposed into the variance arising from the genuine between-setting differences (signal magnitude) and the remainder which is noise variance. The proportion of the total variance which is signal can be called "information content".

$$\text{Information Content} = \frac{\text{Signal Variance}}{\text{Signal Variance} + \text{Noise Variance}} \quad (1)$$

The reasons to use information content rather than simply signal-to-noise ratio are three fold. First, the information content conveniently varies between 0 and 1, rather than extending to infinity. Second, it is symmetrical:

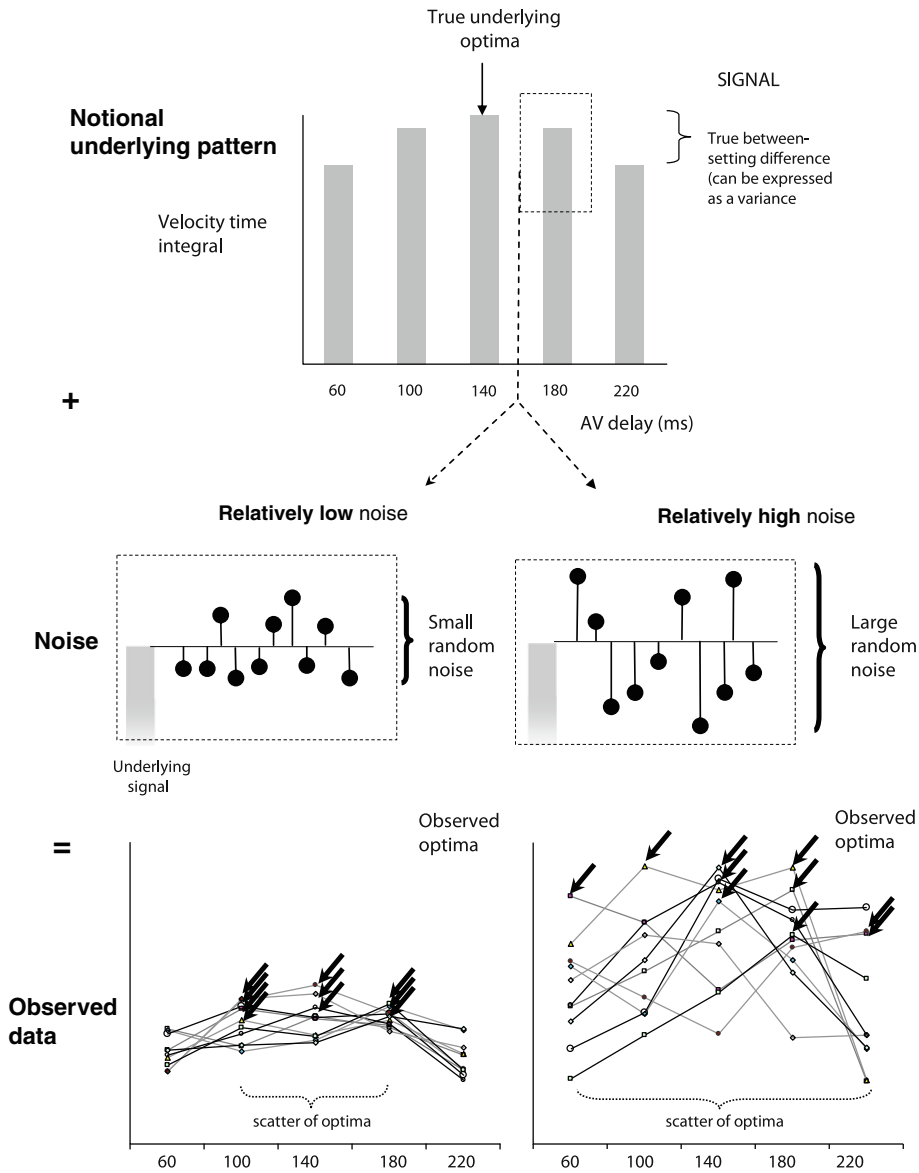

Fig. 2 Observed measurements are composed of underlying true difference between settings ("signal", *top* panel) and beat-to-beat variability ("noise", *middle* panel) which may be small (*left*) or large (*right*) relative to the signal. The relative sizes of underlying signal and overlying noise determine whether the observed measurements (*bottom* panel) reflect the underlying signal faithfully (*left*) or not (*right*). When the noise variability is relatively large (*right*), the observed optimum (*arrow*) is often not the true optimum (140 ms in all cases in this figure)

noise content is 1 minus information content, which makes it clear that there are two contributors to observed differences between settings. Third, it is numerically identical to the intraclass correlation coefficient, a simple index of reproducibility used in biological research.

Published data

Information content can be calculated in any study for which both the overall variability and the noise variability are available. We present in Table 1 information content for three detailed physiological studies conducted in research environment where special attention was given to accuracy [11–13]. For each row of this table, we calculated for each patient the signal size (expressed as a variance) and the noise size (expressed as a variance), and displayed the average values across all patients. Each study had measurements at more than one heart rate, or via more than one monitoring technique, and so had more than one row. Where raw data of multiple replicates were available to us [12], noise variance was quantified directly. Where data of only a single replicate were available [13], noise variance was defined as the dispersion (expressed as a variance) of raw data away from a best-fit regression parabola between the observed measurements and the AV delay. Where noise variance was published graphically [11], it was read off the graph. Signal variance was defined as the total observed variance of that patient minus noise variance. Because the protocols differed between studies, this table should not be used to compare optimization technologies, but rather just to obtain an idea of the realistic range of information content achievable. It

Table 1 Information content calculated from published studies of optimization

Average heart rate (bpm)	Pacing parameter being optimized	Study	Definition of 1 replicate	Number of replicates (*R*)	Parameter used for optimization	Number of patients with sufficient raw optimization data	Signal size, expressed as a variance	The noise size, expressed as a variance	Information content of parameter, when used as an *n*-replicate average (mean ± SEM)
Resting	AV	Auricchio et al. [11]	Average of first 5 beats at an AV delay setting, minus average of previous 6 beats at reference setting (intrinsic)	5	% change in aortic pulse pressure	4 (pts 18, 5, 17 and 13)	10.74%^{2}	45.76%^{2}	0.20 ± 0.15
Resting	AV			5	% change in LV dP/dt	4 (pts 18, 5, 17 and 13)	18.21%^{2}	19.44%^{2}	0.32 ± 0.05
90	AV	van Geldorp et al. [13]	Average of first 8 beats at an AV delay setting, minus average of previous 10 beats at reference setting (intrinsic conduction)	1	Relative change in stroke volume by Nexfin	20	21.78%^{2}	8.70%^{2}	0.63 ± 0.05
90	AV			1	Relative change in stroke volume by echo	20	40.31%^{2}	23.84%^{2}	0.71 ± 0.05
90	AV	Whinnett et al. [12]	Average of first 10 beats at an AV delay setting, minus average of previous 10 beats at reference setting (AV 120 ms)	8	Relative change in systolic blood pressure	10	29.06 mmHg^2	3.81 mmHg^2	0.51 ± 0.09
110	AV			8	Relative change in systolic blood pressure	10	35.58 mmHg^2	2.44 mmHg^2	0.86 ± 0.02
130	AV			8	Relative change in systolic blood pressure	10	55.07 mmHg^2	4.60 mmHg^2	0.90 ± 0.01

should be remembered that these were conducted in ideal research environments when there was effectively no time pressure. Routine clinical practice, because of time pressure, typically falls short of such ideal protocols that might require as many as 1,500 beats to be acquired and analysed [14].

In this study, we present a simple way to establish the impact of spontaneous beat-to-beat variability, by simulating an optimization in which there is a known underlying optimum setting at which cardiac function is best, and alternative settings at which cardiac function decays away. In the simulation, we can then superimpose random variability simulating clinical beat-to-beat variability (the "noise"). This simulation gives information whose applicability is completely general across any method of optimization that is based on selecting the settings which gives the most favourable value of a cardiovascular measure.

We aimed to determine

- how reliable optimization is
- how one can quantify the confidence interval of any observed optimum
- whether one should trust an apparent increment in cardiac function
- whether the observation that optima change over time is a good reason to increase the frequency of repeat optimization, and finally
- whether there are any straightforward steps we can take to improve the quality of the optimization process.

Methods

Observed measurement = underlying signal + superimposed noise

We constructed a simulation to identify the impact of noise variance, which is the random variability occurring between one beat and another. This noise is superimposed on the signal, which is the "true" underlying effect of the pacemaker setting changes in real patient data. In clinical practice, signal and noise cannot be separated in individual raw data points because each such observed measurement contains both contributions mixed together (however, if replicate measurements are made, their inter-replicate variance can be subtracted from the total variance of the observed raw data to reveal the signal variance).

Simulation

In keeping with real patient data [15], the underlying signal in our model was constructed as an inverted parabola with its peak—the underlying optimum—at 140 ms. The vertical

size of the parabola was scaled to have the desired signal magnitude. The magnitude was defined as the average of the squared deviation from the mean: this definition is computationally identical to that of variance. Separately, we programmed noise as normally distributed random values with mean zero and variance as desired. The signal and noise were added together to create the simulated observations. This process was repeated separately for each simulated patient. For each analysis in this study, 1,000 patients were simulated.

We tested signal and noise sizes over a wide range, but for clarity in this paper we have presented a limited number of values, ensuring that the full spectrum of relative sizes of signal magnitude and noise variance is encompassed.

Identification of optimum

We defined the optimal setting as the one which gave the highest measurement of cardiac function [3, 7, 16]. Because of the presence of noise, the measured value of this optimum may not be the same as the underlying optimum. The measured hemodynamic parameter is not specified, but it could represent VTI [3], blood pressure or dP/dt. The measurement is expressed without physical units, for simplicity and generality. Because signal and noise always will have the same unit, the choice of the unit has no impact on reliability of optimization.

Confidence intervals of the optimum

We simulated repeat optimizations within the same individual and collected the resulting optima in order to see how widely these optima were scattered. We defined the 95% confidence interval of a single optimization as 1.96 × the standard deviation of this collection of observed optima. This is the confidence interval that would be appropriate to report for each patient's individual optimization, although by this method it is of course necessary to carry out several optimizations per patient in order to calculate the confidence interval.

Results

Impact of information content on consistency of detecting optimum, using a single beat at each setting

With signal and noise both configured to be the same size, the underlying curved shape of the signal was not always evident in the observed measurements (signal + noise). Nevertheless inevitably, in each run, one of the settings yielded the highest measurement and was duly selected as the observed "optimum". Since this was not always the true *underlying* optimum, the *observed* optima showed some scatter (as shown schematically in Fig. 2).

For each combination of signal and noise size, we quantified the observed scatter of optimization as the standard deviation of difference between the optima obtained on two successive optimizations of the same patient. We calculated the information content from the known sizes of signal and noise.

When signal and noise were equal, there was an optimization scatter (standard deviation) of 45 ms. Making the signal magnitude small made the scatter of the observed optimum wider. Making the signal larger made the scatter of the observed optimum narrower (Fig. 3, Spearman rank correlation coefficient $\rho = 0.973$, $P = 0.021$). When the noise was made smaller, the scatter of the observed optimum narrowed. When the noise was made larger, the scatter of the observed optimum widened (Fig. 3, $\rho = 0.991$, $P = 0.0017$).

The information content was the overwhelming determinant of the scatter of optima ($\rho = 0.979$, $P < 0.001$, Fig. 3). In the worst case scenario, i.e. information content near zero, the scatter of optimization was ~80 ms, the implied range, 60–220 ms, covers the full range of settings over which the simulations are performed.

We can compare this to the expected behaviour of an entirely worthless optimization method, which would be to use no physiological information but simply to select one of the settings (60, 100, 140, 180, 220 ms) at random and announce it to be the optimum. From first principles, the mean "optimum" expected from such an approach is 140 ms, and the expected variance (average square of deviate from that mean) is simply $(80^2 + 40^2 + 0^2 + 40^2 + 80^2)/5 = 3{,}200\ \text{ms}^2$, giving an expected optimization scatter (SD of difference, SDD) of $\sqrt{2} \times \sqrt{3{,}200} = 80$ ms. This forms an effective limit on how poorly reproducible any optimization amongst these settings can be: SDD can never be more than 80 ms, for this range of tested settings.

Figure 3 shows that the information content needs to be rather high before the scatter of optimization even comes close to values that clinicians may consider acceptable. Even to get the SDD of successive optima down to 25 ms, for example, we need information content of 0.91, i.e. signal-to-noise ratio of 10:1.

Size of confidence interval of the observed optimum

We calculated the size of the confidence interval of the observed optimum for a range of possible signal and noise size combinations (and therefore information content) as shown in Table 2.

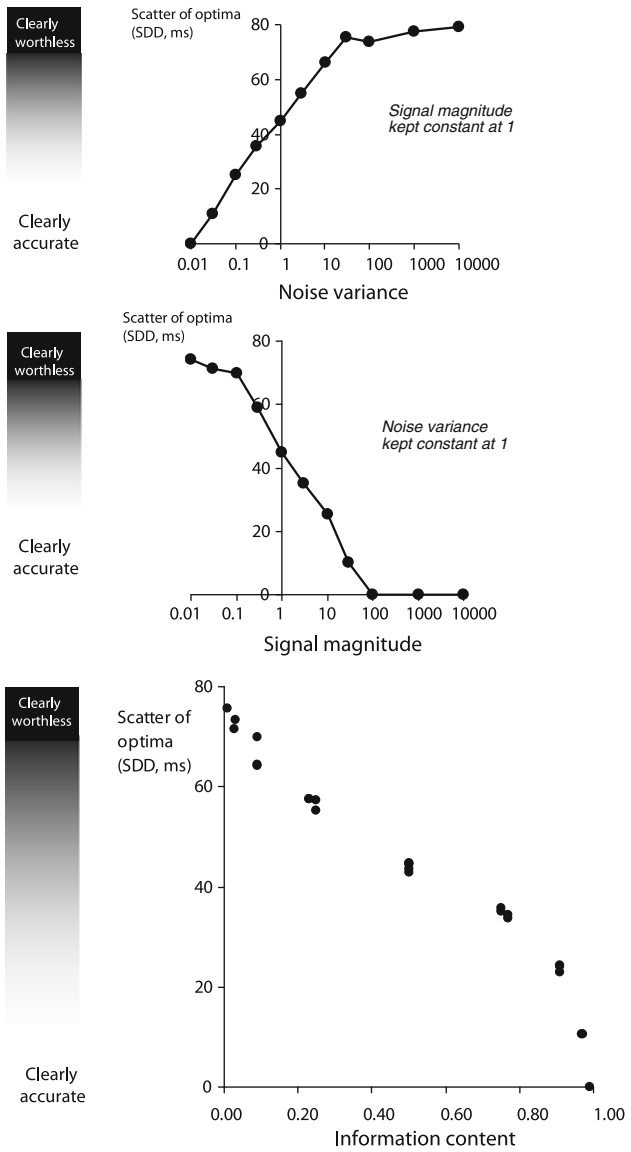

Fig. 3 Scatter between successive optima increases when noise variance is increased (*Top* panel) and decreases when signal magnitude is increased (*Middle* panel). Information content, encompassing the relative sizes of signal and noise, has a powerful effect on the scatter between successive optima. The *bottom* panel shows the effect of information content (the proportion of variance that arises from signal) on the scatter between successive optima

Impact of averaging multiple replicates on reproducibility

We tested the impact of changing a clinic's optimization policy to making, not just a single measurement at each pacemaker setting, but several raw replicates (3, 10, 30 or 100), with average of all those replicate raw measurements in that patient being plotted and used to select that patient's optimum setting. This process improved the fidelity with which the observed measurements reflected the underlying physiological value.

Effectively, the absolute impact of noise was reduced. For example, using averages of 3 replicate raw measurements reduced effective noise variance to one-third (Table 2). With this elevation of the signal-to-noise ratio, the shape of the underlying signal was more faithfully depicted in the observed measurements (Fig. 4, left panels), and the true optimum more likely to be detected (Fig. 4, right panels).

Apparent versus true size of improvement on optimization

We measured the apparent size of the increase in the measured variable upon optimization. To make the results easy to interpret, we simulated patients to arrive in the optimization clinic with a reference setting of 100 ms and undergo an optimization procedure. In each case, the underlying true optimum is 140 ms, but because of noise variability, the setting selected as optimum may be this or another setting.

We calculated several aspects. First, the proportion of patients in whom the observed optimum was a correct reflection of the underlying true optimal AV delay.

Second, we calculated by how much the *observed* optimum appeared to be better than the reference state. In reality, we also knew how much the *underlying* optimum was better than the underlying reference state, and we reported this value too, for comparison. This enabled us to report the extent to which the apparent increase over- or under-estimated the underlying benefit. It was always an over-estimation, as shown in the column "Extent of Illusion" in Table 3. The size of this illusion was strongly determined by the information content, with lower information contents leading to larger illusory improvements ($\rho = -0.975$, $P < 0.001$).

Third, we calculated the *observed* difference between the "best" setting and "worst" setting. Because we knew the *underlying* difference between the true best and worst, we were able to report this too, for comparison. Again we were thereby able to calculate the illusory element (Fig. 5).

Apparent change in optimum over time

We simulated repeating the optimization process after the passage of time, keeping the underlying optimum the same between sessions. We calculated whether the observed optima seemed to change between sessions and by how much.

For each signal and noise combination, we observed the resulting distribution of differences between the optima found at the 1st and 2nd optimization visits. Figure 4 shows these distributions which have information content of 0.91, 0.50 and 0.09, respectively. Since the true

Table 2 Effect of signal and noise on the information content and on widths of 95% confidence limits

Properties of isolated measurements			Effective properties of averaged replicate measurements				Size of 95% confidence interval of an observed optimum
Signal magnitude	Noise variance	Information content	Number of replicate measurements	Effective noise variance	Effective information content	Scatter of optima, SDD (ms)	
1	1	0.50	1	1.00	0.50	45	±62
1	1	0.50	3	0.33	0.75	37	±51
1	1	0.50	10	0.10	0.91	24	±33
1	1	0.50	30	0.03	0.97	12	±16
1	1	0.50	100	0.01	0.99	0	±0
1	10	0.09	1	1.00	0.50	66	±91
1	10	0.09	3	0.33	0.75	58	±80
1	10	0.09	10	0.10	0.91	42	±58
1	10	0.09	30	0.03	0.97	35	±49
1	10	0.09	100	0.01	0.99	26	±36
1	100	0.01	1	1.00	0.50	74	±103
1	100	0.01	3	0.33	0.75	74	±102
1	100	0.01	10	0.10	0.91	68	±94
1	100	0.01	30	0.03	0.97	57	±79
1	100	0.01	100	0.01	0.99	43	±60

For simplicity, the confidence intervals are shown centred on the "true" value. Greater information content gives narrower confidence intervals. The effect of averaging multiple replicate measurements is to reduce the effective noise and therefore narrow the confidence interval

underlying optimum did not change between visits, all changes in observed optimum were false. The proportion of patients of patients giving this false apparent change in optimum is shown by dark shading on the Fig. 6.

Low information content was strongly linked to the likelihood of false-positive detection of change in optimum ($\rho = 0.975$, $P < 0.001$)

When signal and noise were of equal size (information content $\sim$0.50) about two-thirds of the patients have spurious apparent changes in optimum between visits. Even when the signal-to-noise ratio was 10:1, giving an information content of 0.91, still one-third of patients had spurious apparent changes in optimum. Only when signal-to-noise ratio reached 30:1 (information content $\sim$0.97), did the proportion of patients getting false-positive apparent change in optima fall to a clinically respectable 7% (top panel, Fig. 6).

Discussion

In this study, we have shown that uncritically selecting the pacemaker setting which gives the best value of a monitored variable might be little better than random selection amongst a set of AV settings. These findings are generally applicable to any optimization method that relies on testing a series of settings whilst monitoring some measure of cardiac function (such as echocardiographic velocity–time integral or pressure or any other cardiovascular marker) and then picking the setting that gives the highest measurement.

It is overwhelmingly important for signal-to-noise ratio (information content) to be high, otherwise a series of illusions automatically arise in any clinical data analysis.

Illusion 1: "We have selected the true underlying optima"

One tends to assume that the setting which gives the highest measurement is the best. However, our study shows that only a very small amount of variability is enough to seriously compromise this assumption because the true biological effect may also be very small. With signal and noise of equal size for example, in $\sim$50% of cases (Fig. 6) the optimum detected will not be the true optimum but an erroneous alternative.

The confidence interval of a clinical optimization is never reported and (surprisingly) rarely asked for. A wide confidence interval will have immediate comprehensibility to any clinician reviewing the result. The simplest way to calculate the confidence interval of optimization is to carry it out on several occasions (e.g. immediately, one after the other) and calculate the standard deviation. The 95% CI would be the mean $\pm$ 1.96 $\times$ standard deviation. To make this reasonably valid, we would need to perform at least three or four optimizations. Of course this would be

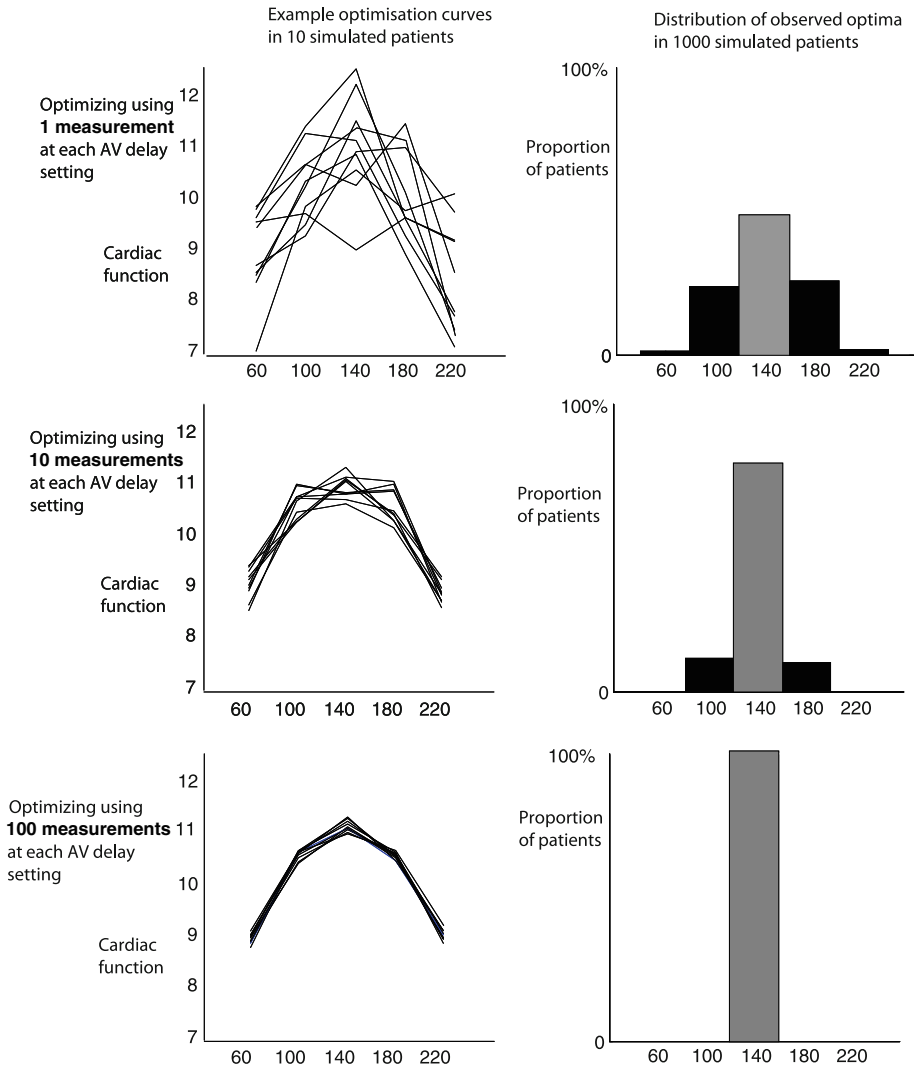

Fig. 4 Impact of switching from single measurements to average of multiple measurements for the optimization process. For simplicity of presentation, the simulated patients all have the same underlying optimum (140 ms), and the signal magnitude and the noise variance for a single measure is set to be 1. The *top* panel simulates optimization with a single measurement made per setting per patient. The *lower* two panels simulate multiple measurements made for each setting in each patient, each patient's optimum being determined using the averages of that patient's replicate measurements. We display detailed optimization curves in 10 example patients (*left*) and the overall distribution of the observed optimum in 1,000 patients (*right*).

Table 3 Impact of information content on the size of the apparent benefit of optimization

Signal magnitude	Noise variance	Information content	Benefit from optimization: optimum minus worst			Benefit from optimization: optimum minus reference		
			Apparent measured increase	True underlying illusion	Extent of illusion (%)	Apparent measured increase	True underlying increase	Extent of illusion (%)
1	1	0.50	3.3	2.1	+54	1.1	0.5	+105
1	10	0.09	7.8	2.1	+263	3.2	0.5	+497
1	100	0.01	23.7	2.1	+1,008	11.2	0.5	+2,002
10	1	0.91	7.5	6.8	+11	1.9	1.7	+10
10	10	0.50	10.4	6.8	+54	3.3	1.7	+98
10	100	0.09	24.8	6.8	+267	10.5	1.7	+522
100	1	0.99	21.9	21.4	+2	5.4	5.3	+0
100	10	0.91	23.6	21.4	+10	5.6	5.3	+6
100	100	0.50	32.9	21.4	+54	11.2	5.3	+110

For each combination of signal and noise variance, we show the apparent benefit of optimization (calculated from the measured data including noise) and the true benefit (calculated from the underlying benefit with no noise). The illusory element is also shown, defined as the degree to which the apparent measured increase overstates the true increase

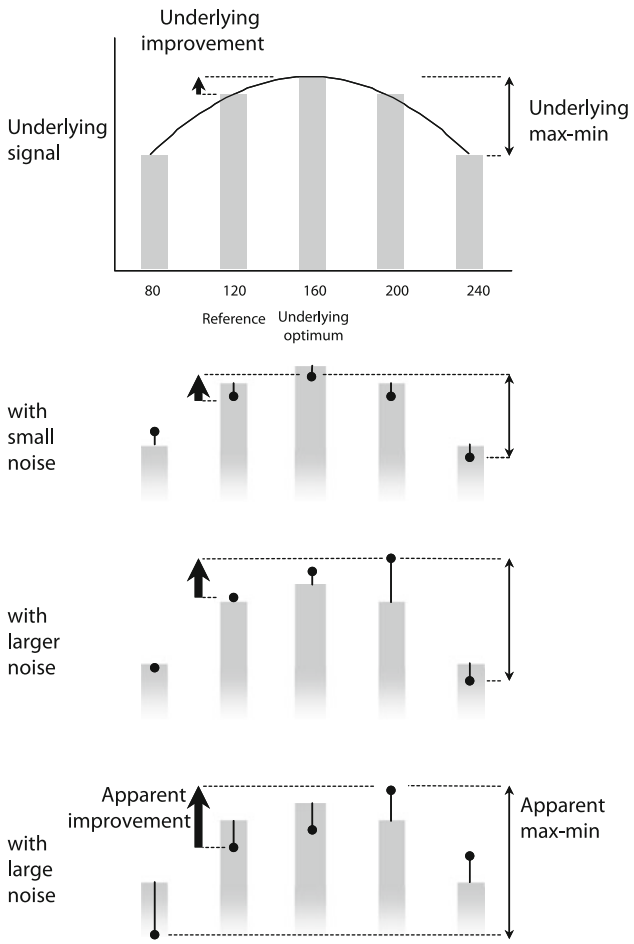

Fig. 5 Relation between variability and apparent benefit of optimization. The presence of noise on its own does not consistently inflate the apparent difference between two predetermined settings, since the noise effect is sometimes positive and sometimes negative. However, it can inflate the difference between a predetermined setting and the apparently best setting. When there is sufficient noise to cause a setting which is not the underlying optimum to appear to be the optimum (because it happens to have had a positive noise element), then we are effectively selecting the setting whose noise is most positive. This introduces a consistent, positive bias whose size increases as the noise becomes more dominant because the most positive noise element is larger, and because in noisier environments there is a larger group of settings amongst which the apparent optimum might plausibly be drawn. As a result, larger noise consistently inflates the difference between the apparent optimum and any reference setting. Likewise the difference between maximum and minimum is also artefactually inflated by increases in noise.

Information content	Proportion of patients whose optima **falsely** appeared to have changed	Proportion of patients whose optima are **correctly** identified as unchanged
0.99	0%	100%
0.97	7%	93%
0.91	31%	69%
0.75	56%	44%
0.50	65%	35%
0.25	75%	25%
0.23	73%	27%
0.09	78%	22%
0.03	80%	20%
0.01	79%	21%

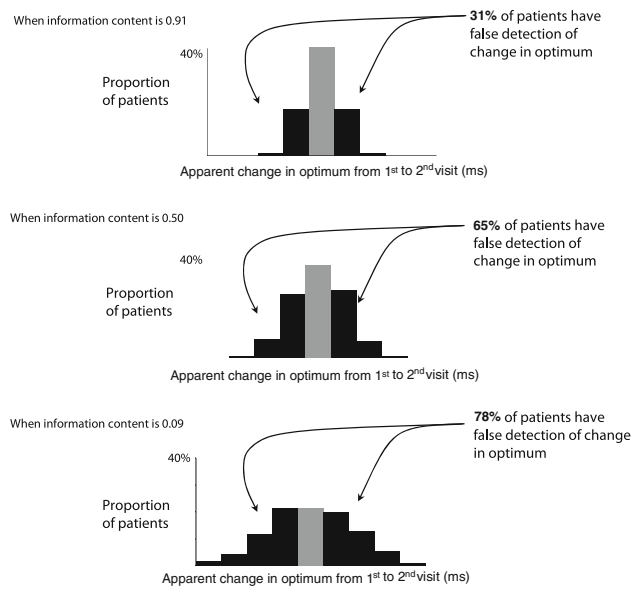

Fig. 6 Impact of information content on the probability of falsely detecting a change in optimum. A group of patients is simulated attending the optimization clinic twice, with their underlying optima truly unchanged between visits. We calculate the apparent change in optimum between visit 1 and visit 2. The table shows the percentage who have a spurious apparent change depending on the information content. Graphically we can see that when information content is high (*top* panel) only 33% of patients have a spurious apparent change in optimum. When information content is slower (*middle* and *bottom* panels), the proportion of patient having a spurious apparent change in optimum becomes much higher. It should be noted that simply randomly choosing between five settings, gives an 80% (4/5) rate of spurious detection of change in optimum, which is no worse than the *bottom* panel.

extremely time-consuming and is therefore not realistic for routine clinical practice with current monitoring techniques.

Alternatively one can determine information content of the clinic's optimization process in general. This could be calculated once and then applied to all similar patients without having to carry out multiple replicate optimizations in each new clinical patient. Fortunately, information content is easy to calculate: it is essentially the intraclass correlation coefficient. This can be calculated quickly for a representative group of patients by any laboratory. This is similar, in principle, to using concepts of statistical power analysis to routine clinical practice.

Illusion 2: "The optimization increased flow (or pressure) by *X* and was therefore worthwhile"

It is tempting to average the apparent increments in velocity–time integral (or whatever measure was used for optimization) achieved in an optimization service, and

believe that first, (a) the process is almost always increasing stoke volume, (b) the size of the average increase in stroke volume is '*X*' which sounds clinically worthwhile and (c) since the increment is statistically significant it is not likely to be a chance finding.

This study reveals all three of these tempting conclusions to be wrong. First, the setting selected as apparently optimum will always have a higher measured cardiac function than the reference setting (except where the reference setting happens to be selected as the optimum). Even if an optimization method was just roulette amongst n tested settings, then in $(n-1)/n$ cases (i.e. almost always) it would be selecting an optimum different from reference. Therefore, the statement that stroke volume is higher on the optimal setting is meaningless.

Second, ironically, the worse the optimization method, the larger the illusionary increase in stroke volume.

Third, unless carefully constructed [8], the statistical test is assessing whether changes in stroke volume are randomly distributed (some positive, some negative) with a mean of zero. But each patient's increment will always be either positive or zero (never negative), so the average increment will always be statistically significantly positive unless the sample size is very small. Indeed, the worse the optimization method, the more likely the apparent increment is to be statistically significantly positive.

Illusion 3: "The optimum has changed between *X* months and now"

A well-established and indispensable optimization clinic may start to consider how often these optimizations should be carried out [7, 9]. Is the contrast between patients' optima on subsequent visits a useful guide? Our analyses now show that if a technology has poor information content (low signal-to-noise ratio), reproducibility will be poor. For example, when signal and noise are approximately equal (information content = 0.5) at 6 months (or any other time), the optimum will falsely appear to have changed, purely through noise, 65% of the time (Table 4). Ironically, the worse the optimization process, the more the data will seem to encourage more frequent optimizations. The giveaway clue to this would be that however frequently we re-optimize, there would still be a similar proportion who would seem to need a change in setting.

Table 4 The number of replicates required when optimization is performed to reduce the scatter of AV optima obtained to a range of acceptable confidence intervals

	ICC of single replicate	For 95% confidence interval of an individual-patient optimum to be:				
		±5ms	±10ms	±15ms	±20ms	±30ms
		Required "effective ICC" of the multiple replicates				
		0.974	0.962	0.93	0.877	0.56
		Number of replicates required to achieve this				
***Projected* behaviour of VV optimisation algorithms which average 18 - 20 beats per single replicate**	0.02	1836	**1240**	651	349	62
	0.05	712	**481**	252	135	24
	0.1	337	**228**	120	64	11
***Actual* behaviour of AV optimisation algorithms which average 18 - 20 beats per single replicate (Table 1)**	0.2	150	**101**	53	29	5
	0.3	87	**59**	31	17	3
	0.4	56	**38**	20	11	2
	0.5	37	**25**	13	7	1
	0.6	25	**17**	9	5	1
	0.7	16	**11**	6	3	1

Illusion 4: "We should not waste time making multiple replicate measurements at each setting in clinical practice"

In a busy clinical department, it may seem an unnecessary multiplication of work to make more than one measurement at each setting. Instead, it may seem rational to concentrate on ensuring that each measurement is acquired and analysed properly by well-trained staff. Unfortunately, the reasons for beat-to-beat variability in measurement are many, and inadequate skill on the part of the sonographer or interpreter is typically not the dominant contribution. Rather, there is substantial beat-to-beat variability in transvalvular blood flow, ventricular volumes, arterial blood pressure and dP/dt. These variations may be due to respiration and numerous other less-easily monitored physiological processes that take place over periods of seconds and minutes. They will not disappear through wishful thinking alone. Instead, averaging multiple replicate acquisitions gives us a powerful method to reduce the effective noise. Effective noise (the variance of the averaged value from R replicate raw measurements) falls in direct proportion to $1/R$, providing a simple way to improve the information content. Another strategy is to elevate heart rate, since this increases the size of the signal [12].

Illusion 5: "We should optimize using whatever measurement method we are most familiar with"

Inter- and intra-observer variability may not be the dominant source of noise, rather there may well be genuine biological variation between beats. Even with excellent clinical acquisition and measurement technique, if the biological variability is large in comparison with the true signal between settings, information content will be low. We should quantify information content directly and not assume that the technique with which we are most familiar has a high information content.

Illusion 6: "Between separate beats, variability in my laboratory is only X%, therefore this measure is suitable for use in optimization"

That X%, being the ratio between variability and *mean* measurement, is not the relevant ratio for quality of optimization. Reliability of optimization depends on the ratio between beat-to-beat variability (noise) and *between-setting* variability (signal). The ratio is much less favourable than X%. For example, a VTI measurement might have a mean value of 10 cm and a standard deviation of 1 cm, giving a coefficient variance of 10%. However, the relevant signal is not 10 cm but the standard deviation *between settings* which may only be (for example) 1 cm. In this case, the information content would be $\frac{1}{1+1} = 0.5$. The naive figure of 10% variability, in isolation, is of no relevance.

A simple method of calculating information content of a cardiovascular measure used clinically for optimization

Because this study was carried out using computer simulation, it was possible to know the size of the true underlying signal, as well as the size of the noise, and thereby state the information content directly.

In vivo, one can calculate information content by measuring total variance and noise variance, since although the underlying signal magnitude cannot be directly observed, it is the difference between them. We need to carry out several optimizations in the same patients. Suppose one carries out R replicate sets of optimizations in one patient. First, calculate the variance of all the raw measurements (V_{raw}). Then one can calculate the mean measurement at each pacemaker setting and then the variance (V_{m}) of these means. V_{m} will tend to be smaller than V_{raw}, because the impact of noise is reduced by the averaging process. The lower the information content in the measurement, the larger its noise in comparison with its signal, and therefore the more markedly V_{m} will differ from V_{raw}. In brief, the information content is approximately the ratio $V_{\mathrm{m}}/V_{\mathrm{raw}}$, when K is large. More elaborately, accommodating for R not always being large,

$$\text{Information content} = \frac{R}{R-1}\frac{V_{\mathrm{m}}}{V_{\mathrm{raw}}} - \frac{1}{R-1} \qquad (2)$$

An example of how to calculate information content in a single patient, using only standard spreadsheet software, is shown in Fig. 7.

In practice, the examples of published data on information content in Table 1 show that even with time-consuming methodology, including a high number of replicates and many beats measured per replicate, information content can still be low.

How many replicates are really needed in clinical practice?

Clinicians cannot afford to waste time in clinical practice on performing unnecessary numerous measurements during optimization. Nor, though, can they waste time performing apparent optimizations that they should know will be worthless before the patient even lies down on the couch. To choose rationally the number of replicates to perform, it is vital to decide how precisely the patient's optimum needs to be identified.

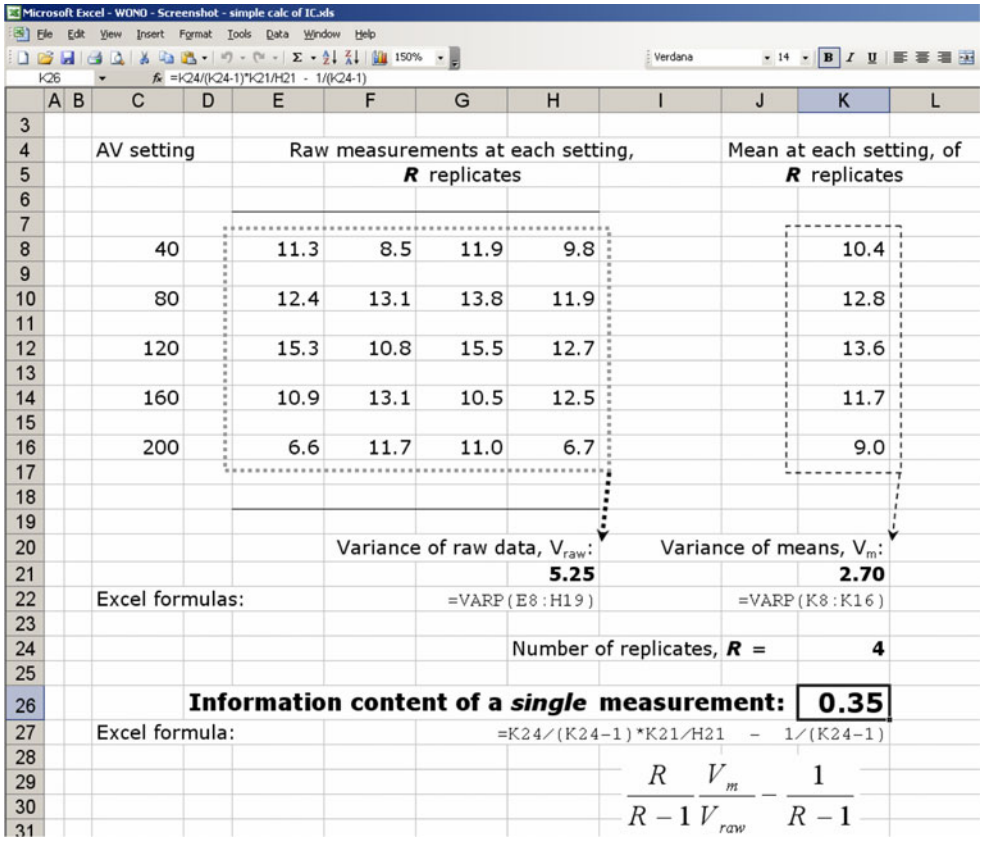

Fig. 7 Calculation of information content using raw clinical data from a single patient. In this example of real-life data from one patient, four replicates of measurements at five settings are entered into a table (columns *E* to *H*), and the mean at each setting calculated (column *K*). The variance of the raw data is calculated (Cell *H21*) using the formula shown immediately below it; the same is done for the variance of the means (Cell *K21*). Information content is calculated in cell *K26* using the formula shown immediately below it. The formulae shown are in appropriate form for standard spreadsheet software such as OpenOffice or Microsoft Excel. In particular, information content for several early patients (who would need to undergo replicate measurements) can be averaged to allow the laboratory to calculate typical confidence intervals to be reported alongside optimal settings in future patients.

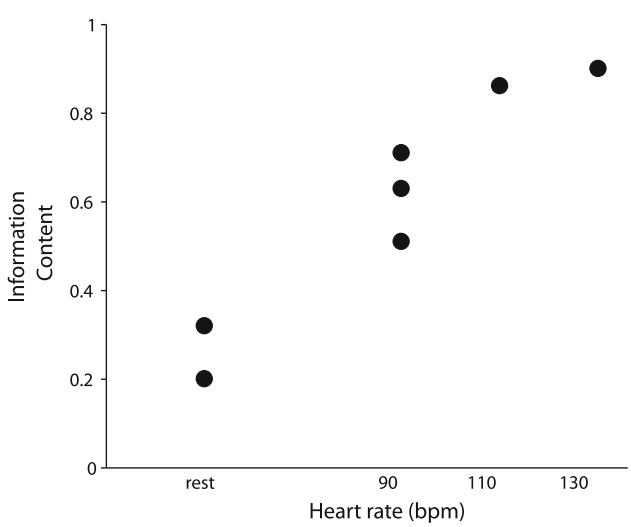

Fig. 8 The effect of heart rate on information content from datasets of published studies regardless of method used for AV optimization

In clinical practice, each individual physician can decide what level of precision is suitable in their context and can easily calculate the number of replicates required to achieve this as long as the information content of a single replicate of their local method is known. The number of replicates required for a range of such combinations is shown in Table 4.

For example, a clinician may wish to know the AV optimum with a 95% confidence interval of ±10 ms. How many replicates are needed depends on the heart rate at which optimization is to be carried out (Table 1). Studies at resting heart rate have found rather low information contents around 0.3.

If a confidence interval of optimization of ±10 ms is wanted in this context, from Table 4 it can be seen that the number of replicates needing to be conducted at each setting is 59.

At higher heart rates such as 90 bpm, information content is approximately 0.5–0.7 for several methods

(Table 1). Achieving a confidence interval of ±10 ms now only requires 11–25 replicates, as shown in Table 4, which might be achievable. At higher heart rates still, the number of replicates needed continues to fall (Fig. 8).

Adjustment of VV delay, in contrast, exerts a much smaller signal effect on physiological measurements than AV adjustment by a factor of about 5–7 fold [15]. Even if the variation in blood pressure is just fivefold smaller, the information content is roughly 25-fold smaller—because it is the variances (squared deviates) that matter. Therefore, even assuming a favourably elevated (90 bpm) heart rate, a favourable range of AV optimization information contents of 0.5–0.7, and a possible relative signal variance for VV of $(1/7)^2$ to $(1/5)^2$, the information content for VV would lie between 0.01 and 0.03. It can be seen from Table 4 that this necessitates well over 500 replicates at each setting to achieve the desired precision of optimization.

Although there are detailed descriptions of meticulous protocols [14] even putting together 1,500 beats of data does not give high information content. Multi beat averages reduce noise, but if signal is small, information content may still be small. High heart rate raises signal magnitude [12] and has allowed a higher information content to be obtained.

Clinical implications

No clinical optimization protocol currently specifies a number of replicates to be carried out, whilst giving a quantitative reasoning. This may be because the impact of noise has not been considered or measured. It may not be rational to conduct an optimization without ensuring adequate precision of the optimum. Although there may be a clinical imperative to be seen to be doing something, we should not necessarily give in to perceived pressure to conduct a placebo procedure. Worse still, if the apparent optimization is in fact no different to randomization amongst a constrained range, it is inescapable that half of all such procedures worsen cardiac efficiency rather than improve it.

If we want our optimization service to be delivering clinical valuable results, there are three generic steps we should take. First, we should have as large an underlying signal as possible. For blood pressure changes, it has been reported that the signal is larger in absolute terms at higher heart rates than at lower heart rates [12].

Second, we should have as small a noise as possible. We should not criticise operators for inadequate care when they may simply be correctly measuring biological variability. Instead we should design our measured variable and protocol to have a high information content.

Third, we can take averages of multiple replicate measurements of cardiac function at each AV delay setting. An *R*-fold replication will have the same beneficial effect as reducing the noise variance of individual measurements by *R*-fold. This can be applied to any measurement technique, but of course carries the cost of increased labour.

Realization of these inherent properties of optimization should encourage us to mandatorily report the noise and information content of our monitored variable in our hands. We should be able to therefore present the confidence interval with every optimization we carry out. This may be uncomfortable.

We emphasize that in this article we are not recommending one method of measurement (e.g. VTI or pressure) over another, nor suggesting whether measurement should be invasive or non-invasive. The choice of measurement modality for optimization should be prejudged by personal preference or based on whim, but rather selected on the balance of relevant properties. The most important property of optimization (a process that recommends small adjustments to pacemaker settings) is the precision with which the recommendation is given. This is a neutral article which simply provides a language to rationally evaluate, discuss and improve this precision.

Practical recommendations

This analysis is completely general to all optimization schemes which test a range of settings and select the one with the greatest measurement. Any laboratory conducting optimization can use Eq. 2 and Fig. 8 to calculate their typical information content. In concert with device physicians, who can recommend an acceptable confidence interval, the laboratory can see how many replicates are required.

Such an estimated number of replicates required only applies to an "average" patient in the population. The size of the signal may vary between patients. For example, one patient may have a particularly critical dependency on AV setting and another a below-average amount of dependency. The former would need fewer replicates to identify the optimum within a given size of confidence interval, and the later would need more. Similarly, one patient may have more noise for one of many reasons, including deeper respiration due to acute physiological distress; chronic lung disease that enhances ventilatory fluctuation in haemodynamics; obesity impairing image quality; agitation impairing probe position maintenance. This would necessitate more replicates.

But whilst individual patients may have different strict needs for replication, all patients will need more replicates if the optimization technique has poor information content. Any protocol document (which specifies and optimization technique) to be credible must at least give quantitatively sound guidance as to the number of replicates needed for an average patient to obtain optimization with a level of

precision widely considered reasonable. If a protocol does not give such guidance, clinical time pressures may lead to all patients having optimizations that are, on average, worthless (helping half slightly and harming half slightly).

Conclusions

Information content, the proportion of the observed differences in the measurements at different settings that is genuinely due to the change in settings, has an overwhelmingly important impact on the meaningfulness of the pacemaker optimization processes. Although easy to measure, it is rarely reported or commented on, and may be surprisingly low unless steps are taken to improve it.

Low information content leads to frequent misidentification of the optimum. However, worse than this, it inflates the apparent benefit of optimization: counter-intuitively, the worse the optimization method, the better it will superficially appear (unless one asks about information content).

Worst of all, because low information content makes apparent optima more variable, the poorer the optimization method, the more frequently one will feel compelled to re-optimize the patient (unless we ask about information content).

Information content is easy to improve for any technique. All that is needed is (a) to use a technique where the underlying difference between settings is as large as possible, (b) to use a technique with beat-to-beat variability as small as possible and (c) to make multiple measurements at each setting and calculate the average.

If, despite these steps, information content is still low, clinical resources could be saved by selecting a setting arbitrarily or even at random, with no additional loss to the patient's physiology. We do not make this suggestion for fun but to point out the seriousness of the present situation. Optimization is not optimization when it is roulette.

Acknowledgments Imperial College London has conducted research into pacemaker optimization funded by Medtronic Incorporated. The British Heart Foundation has funded several of the authors: DPF FS/10/038/28268, PP PG08/114, KW PG07/065. AH and DPF received funding support from NIHR biomedical research centre scheme and the British Heart Foundation Research Excellence Award scheme.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Valzania C, Eriksson M, Boriani G, Gadle F (2008) Cardiac resynchronization therapy during rest and exercise: comparison of two optimization methods. Europace 10:1161–1169
2. Scharf C, Li P, Muntwyler J, Chugh A, Oral H, Pelosi F, Morady F, Armstrong WF (2005) Rate-dependant AV delay optimization in cardiac resynchronization therapy. PACE 28:279–284
3. Barold SS, Ilercil A, Herweg B (2008) Echocardiograhic optimization of the atrioventricular and interventricular intervals during cardiac resynchronization. Europace 10:88–95
4. Gol M, Niazi I, Giudici M, Leman RB, Sturdivant JL, Kim MH, Yu Y, Ding J, Waggoner AD (2007) A prospective comparison of AV delay programming methods for haemodynamic optimization during cardiac resynchronization therapy. J Cardiovasc Electrophysiol 18:490–496
5. Kass DA, Chen CH, Curry C, Talbot M, Berger R, Fetics B, Nevo E (1999) Improved left ventricular mechanics from acute VDD pacing in patients with dilated cardiomyopathy and ventricular conduction delay. Circulation 99:1567–1573
6. Reiter MJ, Hindman MC (1982) Haemodynamic effects of acute atrioventricular sequential pacing in patients with left ventricular dysfunction. Am J Cardiol 49:687–692
7. Zhang Q, Fung JW, Chan YS, Chan HC, Lin H, Chan S, Yu CM (2008) The role of repeating optimization of atrioventricular interval during interim and long-term follow-up after cardiac resynchronization therapy. Int J Cardiol 124:211–217
8. Turcott RG, Witteles RM, Wang PJ, Vagelos RH, Fowler MB, Ashley EA (2010) Measurement precision in the optimization of cardiac resynchronization therapy. Circ Heart Fail 3:395–404
9. Porciani MC, Dondina C, Macioce R, Demarchi G, Cappelli F, Lilli A, Pappone A, Ricciardi G, Colombo PC, Padeletti M, Jelic S, Padeletti L (2006) Temporal variation in optimal atrioventricular and interventricular delay during cardiac resynchronization therapy. J Card Fail 12:715–719
10. Anselmino M, Antolini M, Amellone C, Piovano E, Massa R, Trevi G (2009) Optimization of cardiac resynchronization therapy: echocardiographic versus semiautomatic device algorithms. Congest Heart Fail 15:14–18
11. Auricchio A, Stellbrink C, Block M, Sack S, Vogt J, Bakker P, Klein H, Kramer A, Ding J, Salo R, Tockman B, Pochet T, Spinelli J (1999) Effect of pacing chamber and atrioventricular delay on acute systolic function of paced patients with congestive heart failure. Circulation 99:2993–3000
12. Whinnett ZI, Davies JE, Willson K, Chow AW, Foale RA, Davies DW, Hughes AD, Francis DP, Mayet J (2006) Determination of optimal atrioventricular delay for cardiac resynchronization therapy using acute non-invasive blood pressure. Europace 8:358–366
13. van Geldorp IE, Delhaas T, Hermans B, Vernooy K, Broers B, Klimusina J, Regoli F, Faletra FF, Moccetti T, Gerritse B, Cornelussen R, Settels JJ, Harry JGM, Crijns HJGM, Auricchio A, Prinzen FW. Comparison of a non-invasive arterial pulse contour technique and echo Doppler aorta velocity-time-integral on stroke volume changes in optimization of CRT. Europace 2010 (in press)
14. Auricchio A, Stellbrink C, Sack S, Block M, Vogt J, Bakker P, Mortensen P, Klein H (1999) The pacing therapies for congestive heart failure (PATH-CHF) study: rationale, design, and endpoints of a prospective randomized multicenter study. Am J Cardiol 11;83(5B):130D–135D
15. Whinnett ZI, Davies JE, Willson K, Manisty CH, Chow AW, Foale RA, Davies DW, Hughes AD, Mayet J, Francis DP (2006) Haemodynamic effects of changes in atrioventricular and interventricular delay in cardiac resynchronisation therapy show a consistent pattern: analysis of shape, magnitude and relative importance of atrioventricular and interventricular delay. Heart 92:1628–1634
16. Nishimura RA, Hayes DL, Holmes DR Jr, Tajik AJ (1995) Mechanism of hemodynamic improvement by dual-chamber pacing for severe left ventricular dysfunction: an acute Doppler and catheterization hemodynamic study. J Am Coll Cardiol 25:281–288