Europace (2014) **16**, 541–550
doi:10.1093/europace/eut257

Applicability of the iterative technique for cardiac resynchronization therapy optimization: full-disclosure, 50-sequential-patient dataset of transmitral Doppler traces, with implications for future research design and guidelines

Siana Jones[1]*, Matthew J. Shun-Shin[1], Graham D. Cole[1], Arunashis Sau[1], Katherine March[1], Suzanne Williams[1], Andreas Kyriacou[1], Alun D. Hughes[1], Jamil Mayet[1], Michael Frenneaux[2], Charlotte H. Manisty[1], Zachary I. Whinnett[1], and Darrel P. Francis[1]

[1]International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, 59–61 North Wharf Road, London W2 1LA, UK; and [2]School of Medicine and Dentistry, University of Aberdeen, UK

Received 10 April 2013; accepted after revision 27 July 2013; online publish-ahead-of-print 25 September 2013

Aims Full-disclosure study describing Doppler patterns during iterative atrioventricular delay (AVD) optimization of biventricular pacemakers (cardiac resynchronization therapy, CRT).

Method and results Doppler traces of the first 50 eligible patients undergoing iterative Doppler AVD optimization in the BRAVO trial were examined. Three experienced observers classified conformity to guideline-described patterns. Each observer then selected the optimum AVD on two separate occasions: blinded and unblinded to AVD. Four Doppler E-A patterns occurred: A (always merged, 18% of patients), B (incrementally less fusion at short AVDs, 12%), C (full separation at short AVDs, as described by the guidelines, 28%), and D (always separated, 42%). In Groups A and D (60%), the iterative guidelines therefore cannot specify one single AVD. On the kappa scale (0 = chance alone; 1 = perfect agreement), observer agreement for the ideal AVD in Classes B and C was poor (0.32) and appeared worse in Groups A and D (0.22). Blinding caused the scattering of the AVD selected as optimal to widen (standard deviation rising from 37 to 49 ms, $P < 0.001$). By blinding 28% of the selected optimum AVDs were $\leq$60 or $\geq$200 ms. All 50 Doppler datasets are presented, to support future methodological testing.

Conclusion In most patients, the iterative method does not clearly specify one AVD. In all the patients, agreement on the ideal AVD between skilled observers viewing identical images is poor. The iterative protocol may successfully exclude some extremely unsuitable AVDs, but so might simply accepting factory default. Irreproducibility of the gold standard also prevents alternative physiological optimization methods from being validated honestly.

Keywords Atrioventricular delay (AVD) • Doppler echocardiography • Optimization

Introduction

Biventricular pacing is an effective treatment for heart failure.[1–3] Expert recommendations for post-implant optimization of device settings include methods for optimizing the atrioventricular delay (AVD).[4–6] The iterative method for determining the optimal AVD, recommended in the guidelines[4–6] and the literature[7] and used in a landmark trial of biventricular pacing,[1] involves subjective analysis of transmitral flow patterns.

The original studies that developed the iterative method appear never to have been published. Nevertheless, as well as being extensively described in the literature, it has also been used as a reference

* Corresponding author. Tel: +44 207 594 1027; fax: +44 207 594 1706. *E-mail address*: siana.jones@imperial.ac.uk

What's new?

- This study provides a novel insight into the guideline recommended use of iterative atrioventricular delay (AVD) optimization in patients undergoing cardiac resynchronization therapy.
- First, we present an original classification of the actual behaviour of transmitral inflow pattern during AVD optimization, which highlights that over half of the patients do not follow the 'classic' behaviour described in the guidelines, which prevents it from identifying one particular AVD as the iterative optimum.
- Secondly, independent observers presented with the same sets of transmitral Doppler images do not agree well on the optima. The mean bias is low, but this only indicates that the observers' values for the group of patients have distributions with almost the same mean, and gives no indication of whether they are similar or different in individual patients. The level of agreement can be seen in *Figure 4*, and can be quantified by the confidence intervals, which are wide.
- Thirdly, agreement is even poorer when the observers are blinded to the AVD being tested suggesting that they may be accessing collateral information when making their decision during unblinded selection.
- Finally, datasets are presented in an online supplement to support future methodological testing.

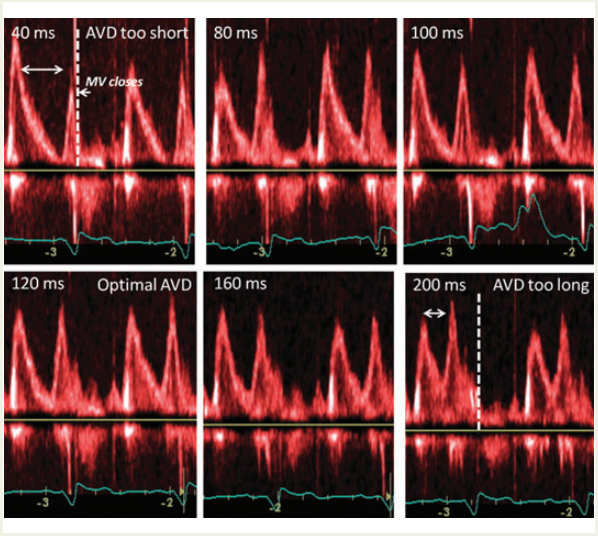

Figure 1 The iterative method of AVD optimization. Examples of transmitral inflow patterns at short, optimal, and long AVDs are given from a patient undergoing AVD optimization by the iterative method who showed transmitral filling pattern changes that matched the guideline expectations. AVD, atrioventricular delay.

standard against which novel algorithms such as SMART-AV[8,9] and QuickOpt[10] have been compared.

The iterative method relies on a predicted pattern of transmitral inflow Doppler (E- and A-wave) response to changing the AVD. When the AVD is too short (for most patients ~40–80 ms), the E- and A-waves are predicted to be widely separated, with A-wave truncation caused by ventricular contraction shutting the mitral valve early. As the AVD is lengthened, the E- and A-waves should move closer together until the A-wave is no longer truncated. When the AVD is too long, E- and A-waves are expected to merge, because passive diastolic filling is delayed by a much later ventricular contraction; the effect of this is to reduce the net ventricular filling by (i) reducing forward flow because of the merging and (ii) increasing backward flow due to pre-systolic mitral regurgitation. The AVD that gives distinct E- and A-waves, with no truncation of the A-wave, should give maximal atrial contribution to ventricular filling and minimal mitral regurgitation, and is considered optimal (*Figure 1*).

It is unknown in what proportion of real-world patients transmitral Doppler pattern changes with changes in AVD in the manner required to implement the iterative method as stated in the guidelines. Trials and review articles rarely describe cases in which such optimization is not achievable. Our first aim was to review the literature for the nature and extent of the descriptions given.

Our second aim was to examine, with experienced operators, and by using consecutive patients recruited into a randomized controlled trial of cardiac resynchronization therapy (CRT) optimization, (i) what proportion of patients show guideline-described Doppler patterns during AVD adjustment, (ii) the degree of agreement between operators on optimum from viewing identical digital datasets, and (iii) whether blinding to AVD affected the choice of optima.

Finally, we aimed to determine whether experienced operators are truly selecting the atrioventricular (AV) optimum based only on the information contained in the Doppler traces, or are biased to favour AVD values considered physiologically plausible a priori.

These findings will have important implications in the applicability of using the iterative method clinically and also as an optimization method for verification of novel optimization techniques.

Methods

Literature review

We conducted a literature review to determine the proportion of papers that provided: a written description of the iterative method, showed a cartoon representation of the iterative method, showed examples from real pulsed-wave Doppler traces, showed more than the bare minimum of three traces (too short, too long, and just right), and showed data from more than one example patient.

The review was performed systematically using Pubmed; details of the search criteria and methods for screening and excluding papers are contained in Supplementary material online, Appendix 1.

Clinical study

Patient demographics

We obtained transmitral Doppler AV optimization data from consecutive patients undergoing AVD optimization with echocardiography in the BRAVO randomized controlled trial (NCT01258829), until 50 patients in sinus rhythm and with Doppler images considered of suitable quality were obtained. All the patients had biventricular pacing set in the DDD or DDD-R mode. The Doppler images from the three patients not

considered suitable were also documented and shown in the Online Supplementary data.

All the patients entering the BRAVO trial were New York Heart Association (NYHA) Class II or above (although in practice almost all were NYHA II), had undergone CRT implantation >6 months previously and were >90% biventricular paced. Patients with cardiac events in the preceding 6 weeks were excluded. Age was 67 ± 9 years with average implant duration of 40 ± 21 months. All the patients showed some level of diastolic dysfunction, which has been reported based on E-/A-wave pattern. Other patient characteristics are shown in *Table 1*. All the patients gave written informed consent for the study, which was approved by the local Research Ethics Committee.

Image acquisition

All the patients had their devices programmed with rate-responsive AVD off, the left ventricular (LV)–right ventricular delay fixed at 0 ms and lower rate programmed to ensure atrial sensing. All the data for an individual patient were acquired in one session to minimize physiological variation and changes in heart rate. Pulsed-wave Doppler images were obtained using a Vivid I (Vivid I system, GE Healthcare) with a 1.5–3.6 MHz transducer, with simultaneous three-lead electrocardiogram acquisition to confirm biventricular pacing.

For each tested AVD, four to eight beats of mitral inflow data were acquired, with data from ectopic and post-ectopic beats excluded. Atrioventricular delay was programmed in 20 ms increments from 40 (or 60 depending on device interactions) until the programmer marked beats as entirely intrinsic ventricular conduction (i.e. beyond fusion) or, in cases of complete heart block, at 320 ms. After the changes, the sonographers waited for at least 12 beats before capturing the mitral inflow Doppler images.

Selecting atrioventricular delay optima

A set of slides (available to any reader on request) showing the transmitral Doppler E–A-wave traces for each AVD was prepared using Microsoft Powerpoint. Patients had anonymous study identifiers. Each slide showed, from one patient, between two and five clearly displayed, non-ectopic beats recorded for each tested AVD.

Three sonographers, trained and regularly conducting iterative optimization, independently examined each set of slides and, for each patient, selected the optimal AVD to be that at which the mitral inflow pattern best matched the desired pattern in accordance with the guidelines[4] (as shown in *Figure 1*). The programmed AVD value for each Doppler trace was visible with the trace.

At least 2 weeks later, the same sets of images were then re-presented to the same observers but with the AVD information removed, new identification numbers allocated, and the order of the patients changed. The observers again selected the optimal AVD, this time using only the Doppler traces, blinded to AVD.

Classification of patients' transmitral Doppler patterns

The 'classic' guideline-described behaviour of mitral inflow Doppler across a range of AVDs during iterative optimization (shown in *Figure 1*), with at least one trace with a truncated A-wave (AVD too short), one trace with E–A fusion (too long), and one that has neither (optimal), may not always be identifiable. We, therefore, devised a systematic classification to describe the spectrum of 'real life' behaviour in a better manner. Patients were categorized into one of four transmitral classes (*Figure 2*) depending on the behaviour of their mitral inflow Doppler patterns with AVD adjustment:

Table 1 Patient demographics of the 50 patients included in this analysis

Demographics		*n* (%)
Gender	Male	41 (82)
Aetiology	Ischaemic	20 (40)
NYHA functional class	II	49 (98)
	III	1 (2)
LV ejection fraction (%)	39 ± 13	–
Diastolic function at 120 ms AVD	E-wave dominant	19 (38)
	A-wave dominant	31 (62)
Medication	ACE/ARB inhibitors	49 (98)
	ß-Blockers	42 (84)
	Digoxin	4 (8)
	Furosemide	27 (54)
	Spironolactone	26 (52)

Transmitral Class A: E- and A-waves were merged at all AVDs, except when the AVD was so short that the A-wave was abolished (a situation not covered by the guidelines).

Transmitral Class B: E- and A-waves were always merged, but the time between peaks increased at a shorter AVD (not explicitly described by the guidelines, but manageable by implicit extension).

Transmitral Class C: E- and A-waves were separate at short AVDs and became merged at long AVDs (in accordance with the guidelines).

Transmitral Class D: E- and A-waves remained separate at all AVD settings (again not covered by the guidelines).

Two independent observers classified each of the 50 patients' mitral inflow behaviour patterns; a third independent observer resolved any discrepancies of classification.

Statistical analyses

Agreement on classification of transmitral Doppler patterns was measured using Cohen's kappa for two observers. Agreement regarding the optimal AVD was addressed in two manners: categorical and continuous. Light's kappa, a multi-rater multi-category version of Cohen's kappa,[11] was used to quantify this. Kappa scores can be intuitively considered a 'probability of two observers agreeing', with the proviso that chance agreement has been subtracted, as explained in Supplementary material online, Appendix 2. A kappa score of 0 indicates a purely chance level of agreement, <0.2 may be considered poor agreement, and >0.8 very good agreement.[12] Bootstrapped confidence intervals (CIs) were calculated.

For continuous analysis, Bland–Altman plots[13] were generated for each of the pairwise comparisons between the three observers. The limits of agreement at ± 1.96 SDs are shown in each experimental state: unblinded and blinded to the AVD. The standard deviation of the entire population of selected values was calculated in each experimental state. The within-patient standard deviation of the values selected by the three observers was calculated using analysis of variance techniques in each experimental state. To visualize the effect of blinding to AVD on selection of optima, histograms were generated of the relative frequency of selected AVD optima under each experimental state. Variances were compared using Fisher's *F*-test for equality of variances. Proportions were compared using Fisher's exact test.

Statistical calculations were performed using 'The R project for statistical computing',[14] with the 'psy' package[15] for calculating Light's kappa, and 'ggplot2' for generating graphs.[16]

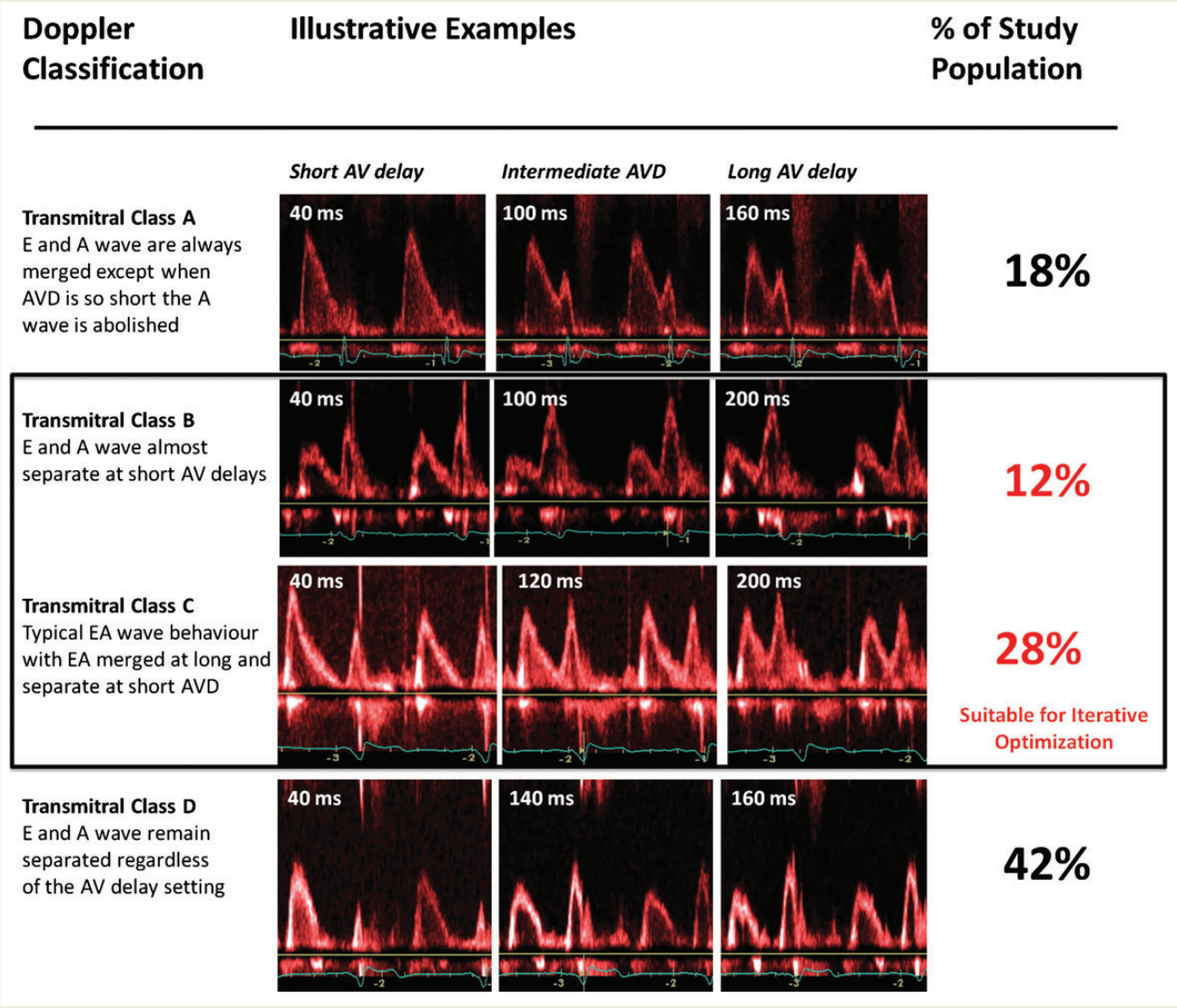

Figure 2 Classification of patients' Doppler traces into Transmitral Classes A–D. Example data from four patients showing three of the tested AVDs: the shortest delay (40 ms), an intermediate delay that might be optimal, and the longest AVD before intrinsic conduction. The percentage of the study population in each transmitral class and the implementability of the iterative method for optimization of these patients are given on the right. Only Transmitral Classes B and C (boxed) are suitable for the iterative method. AVD, atrioventricular delay.

Scientific integrity

The authors are committed to conducting and presenting research reliably. All the raw data are available to any reader. Without precondition, we welcome collaboration with others seeking to confirm, develop, or refute these findings. No data were deleted, nor re-measured, to favour one result over another.[17]

Results

Patient demographics

Patient demographics are shown in *Table 1*.

Systematic review of published patterns

The results of the systematic review are summarized in *Table 2*. Of the 392 publications identified, 19 publications met criteria for inclusion see Supplementary material online, Appendix 1).

Of the 19 papers[4,7–10,18–31] describing iterative optimization, 10 (53%) showed examples from real patient traces, 3 (16%) also had a cartoon representation. Only seven (37%) showed more than a bare minimum of three traces (too short, too long, and just right) and none showed data from more than one example patient. *Table 2* shows the detailed results.

Classification of transmitral inflow pattern behaviour

To obtain 50 patients with adequate images, the traces of 53 consecutive patients undergoing iterative optimization were screened for eligibility. One patient was excluded for very poor echocardiographic windows, and two for excessive atrial or ventricular ectopy resulting in unusually significant beat-to-beat variation in mitral inflow pattern. A summary document showing two-beat Doppler traces for all

Table 2 Descriptions of the iterative method in the literature identified by systematic review

First author	Year	Article type	Patients enroled	Written description	Example patients shown	Example AVD	Cartoon figure	Iterative impossible
Ellenbogen KA	2010	Trial	1014	Yes	0	0		0
Kamdar R	2010	Trial	26	Yes	0	0		0
Kedia N	2006	Trial	215	Yes	0	0		0
Pavlopoulos H	2010	Review	–	Yes	0	0		–
Stein KM	2010	Study design	–	Yes	0	0		–
Taha N	2010	Trial	44	Yes	0	0		0
Valzania C	2007	Trial	37	Yes	0	0		0
Vidal B	2007	Trial	100	Yes	0	0		0
Zuber M	2008	Trial	43	Yes	0	0		0
Jones RC	2010	Trial	63	Yes	1	2		0
Gorcsan J	2008	Guidelines	–	Yes	1	3	Yes	–
Levin V	2007	Trial	25	Yes	1	3	Yes	0
Barold SS	2008	Review	–	Yes	1	4	Yes	–
Rafie R	2012	Trial	34	Yes	1	4		–
Naqvi TZ	2010	Review	–	Yes	1	5		–
Waggoner AD	2008	Review	–	Yes	1	5		0
Cuoco FA	2012	Review	–	Yes	1	5		0
Hasan A	2006	Trial	22	Yes	1	8		0
Naqvi TZ	2006	Case studies	3	Yes	1	9		0

Publications are ordered by the number of patients' Doppler traces shown, and the number of tested AVDs presented for each patient. The number of patients who were reported by the authors as un-optimizable by the iterative method is given in the final column.
AVD, atrioventricular delay.

patients at each AVD, including the three excluded patients, are shown in the Online Supplementary data. The full dataset, for study replication, is available from the authors in Powerpoint format on request.

The 50 patients were categorized into four classes based on their transmitral flow patterns as follows: 9 (18%) in Transmitral Class A, 6 (12%) in B, 14 in C (28%), and 21 in D (42%) as shown in *Figure 2*.

Agreement on classification between the two initial observers was good: Cohen's kappa = 0.80 (95% CI 0.65–0.91). The observers agreed on classification in 43 patients, and in the remaining 7, the third observer agreed with one of the other two observers, and the patient was allocated to that class. Light's kappa for classification, covering all three observers, was 0.72 (0.60–0.83).

Inter-observer agreement on optimal atrioventricular delay using the iterative method

Categorical analysis

Agreement was poor between the observers as assessed using kappa, regardless of whether they were blinded to AVD (0.32, 95% CI 0.22–0.44), or unblinded (0.27, 95% CI 0.16–0.40) to the AVD (*Table 3*).

Continuous analysis

For each subject, the optima selected by different observers varied substantially regardless of whether blinded or unblinded to the AVD, this variability was consistent across the range of AVDs (*Figures 3* and *4*).

Bland–Altman plots for the pairwise comparisons between the three observers are shown in *Figure 4*. The mean bias between the observers was small when both were blinded (Ob1 vs. Ob2 3.2 ms, Ob1 vs. Ob3 1.6 ms, Ob2 vs. Ob3 −1.6 ms) and unblinded (Ob1 vs. Ob2 −3.6 ms, Ob1 vs. Ob3 −2.0 ms, Ob2 vs. Ob3 1.6 ms) to the AVD.

However, the limits of agreement (i.e. ±2 SDs) were clinically large when both were blinded (Ob1 vs. Ob2 −77 to 70 ms, Ob1 vs. Ob3 −95 to 91 ms, Ob2 vs. Ob3 −82 to 85 ms) and unblinded (Ob1 vs. Ob2 −52 to 59 ms, Ob1 vs. Ob3 −61 to 64 ms, Ob2 vs. Ob3 −71 to 68 ms) to the AVD.

The within-subject standard deviation for the AVD selected by the three operators was correspondingly significantly larger when the operators were blinded to the AVD than when unblinded (30.0 vs. 22.4 ms, $P = 0.002$).

That the mean difference between replicate measurements is close to zero is not reassuring,[32] since entirely random numbers drawn from the same distribution would be expected to have a mean difference of approximately zero. The relevant measure is the spread between different measurements in the same patient.

Effect of blinding vs. unblinding to atrioventricular delay on identification of the optimum

The relative frequency of the AVD selected as optimal is shown in *Figure 5*, under blinded (top), and unblinded (bottom) conditions.

Table 3 Inter-observer agreement within guideline and non-guideline recognized transmitral behaviour

Transmitral classification	Distribution, *n* (%)	Light's kappa score for observer agreement (95% CI)	
		Blinded to AVD	Unblinded to AVD
Transmitral Classes B and C	20 (40%)	0.41 (0.27–0.60)	0.32 (0.18–0.52)
Transmitral Classes A and D	30 (60%)	0.23 (0.10–0.39)	0.21 (0.08–0.39)
All patients	50 (100%)	0.32 (0.22–0.44)	0.27 (0.16–0.40)

AVD, atrioventricular delay.

When unblinded to the AVD, the observers were significantly more likely to select an AVD inside the physiologically plausible central range of 80–160 ms (83 vs. 72%, $P = 0.04$). Correspondingly, the standard deviation of the selected optimal delays was greater when the observers were blinded than unblinded (48.7 vs. 37.4 ms, $P = 0.001$).

Discussion

Clinical research and guidelines recommend iterative optimization, whereby the AVD at which the transmitral Doppler E- and A-waves are separated without truncation of the A-wave is selected as the optimum.[4,5,7] In this systematic full-disclosure study, we wanted to classify the patterns of change of transmitral Doppler flow with changes to the AVD. In the absence of any published classification system, we are obliged to devise a novel system for distinction between different Doppler pattern behaviours. All the raw datasets are shown in an Online Supplementary data so that the readers can review the raw data and propose and justify an alternative system if desired. The system designates each patient into one of four classes of transmitral flow behaviour with increasing AVD. Less than half of the patients had transmitral Doppler profiles that changed as described in the guidelines and schematic representations.

This has several implications: first, only 28% of the patients display the pattern changes commonly shown in the literature. A systematic review of the literature showed that all example traces displayed in published manuscripts were of Transmitral Class C as described in the guidelines, and never from Classes A, B, and D, despite these forming the majority of patients in our study. The tendency to select one 'best' example for publication is understandable but may induce unrealistic expectations in readers regarding the ease of performing iterative AVD optimization.

Secondly, the majority of patients, 60%, have Doppler patterns that do not conform to standard descriptions and schematic representations of the iterative method.

Thirdly, experienced observers, examining identical sets of transmitral inflow Doppler patterns, and following the guidelines for iterative optimization but forbidden from conferring, rarely agreed on the AVD optima to a level expected for a guideline recommended protocol in general clinical use: kappa was only 0.27 on a scale, where 1 indicates perfect agreement and 0 indicates agreement expected by chance alone.

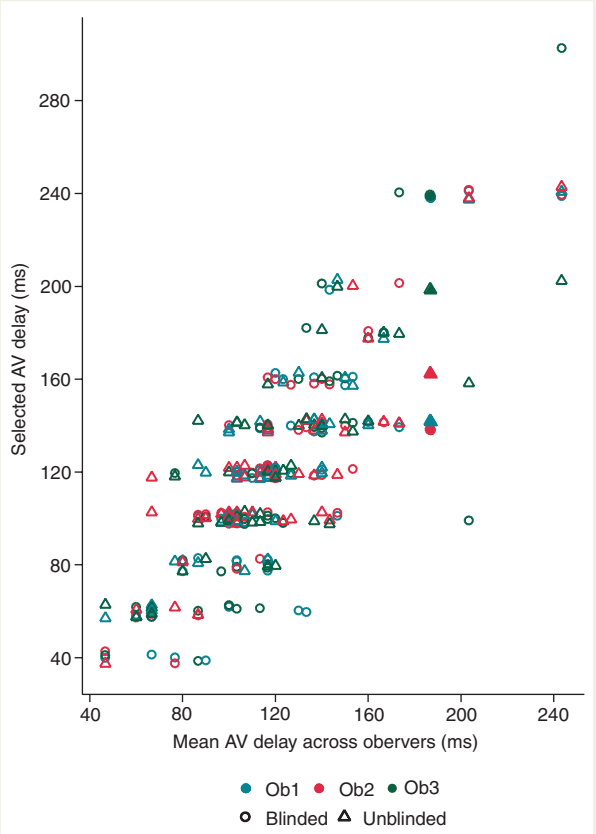

Figure 3 The spread of the AVD optimum identified by three observers, on two viewings, plotted against the best estimate of each patient's optimum. *Figure 3* visualizes the spread of all the six assessments vertically from a single patient. For each patient, the mean of the six assessments is taken as the best estimate of the optimum. The six individual assessments for that patient are plotted at that *x*-coordinate (small displacements have been added to increase visibility). One patient is shown with solid symbols as an example: the mean of his six evaluations was 187 ms, and the six evaluations are individually shown, ranging from 140 to 240 ms. Across the graph as a whole, the vertical spread gives an impression of the degree of discrepancy between the AVD optima assigned by different observers and the same observer on second viewing. Across all the patients, the average spread between the six evaluations is 22 ms. AVD, atrioventricular delay; Ob, observer.

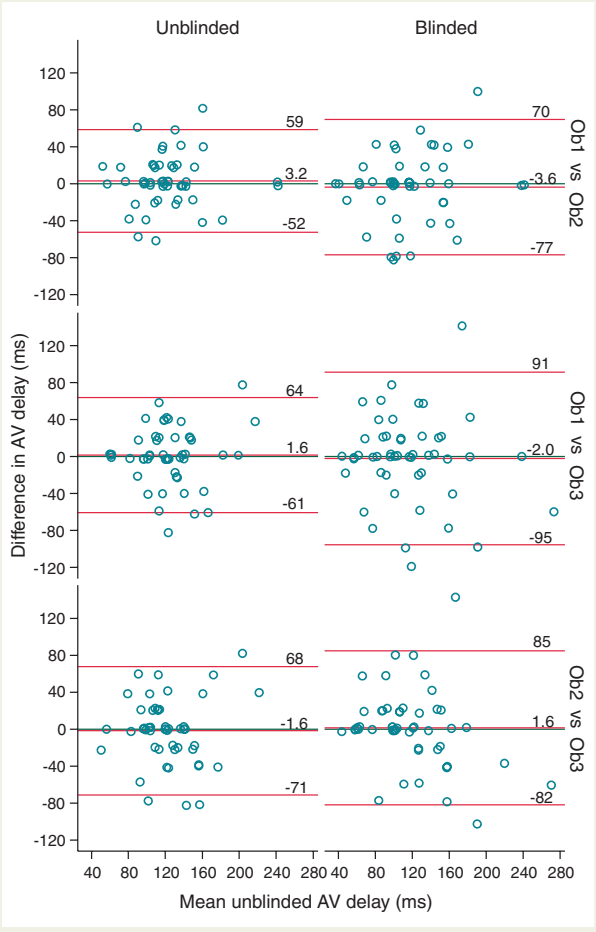

Figure 4 Bland–Altman plots showing inter-observer agreement of the AVD optima under unblinded (left) and blinded selection conditions (right), The three observers had no tendency to pick consistently longer or shorter optima than each other (no bias). The observers did not agree on the optimal AVD for each patient (wide limits of agreement) and their disagreement was greater when they were blinded to the AVD (right panels), the limits of agreement were wider than when unblinded (left panels), $P = 0.02$. Plotted points are displaced slightly so that multiple points, that would otherwise overlap, can be seen. AVD, atrioventricular delay; Ob, observer.

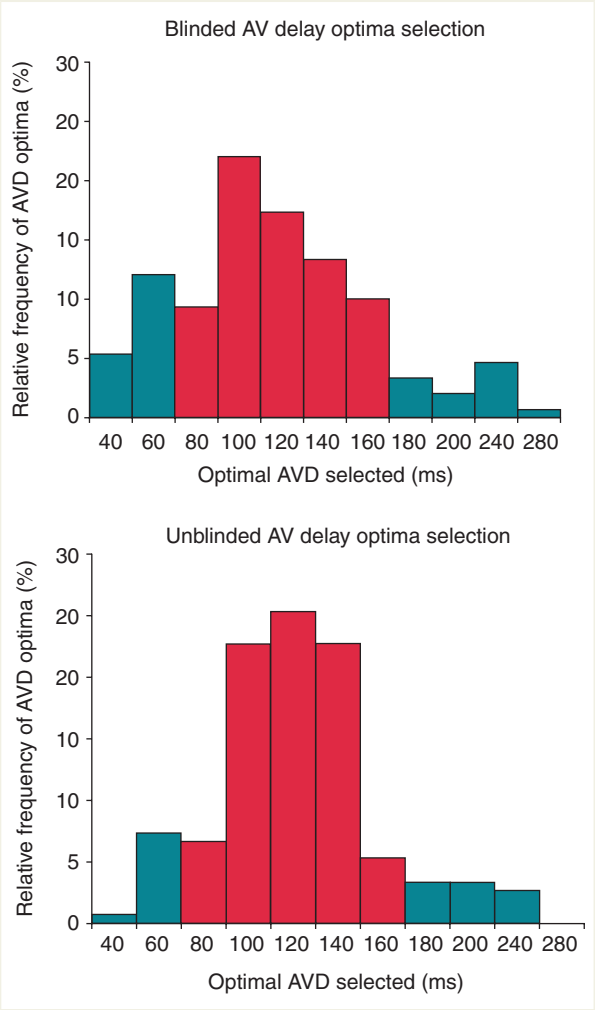

Figure 5 Relative frequency of the AVD selected as optimal by the operators who were shown the transmitral Doppler filling patterns with the corresponding AVD values (top) and blinded to the AVD values (bottom). When the operators were blinded to the AVD values corresponding to the transmitral Doppler filling patterns, they were more likely to select an optimum outside the usual physiological range of 80–160 ms. AVD, atrioventricular delay.

Identification of one optimal atrioventricular delay vs. a range that excludes some extremely unsuitable atrioventricular delays

Iterative optimization currently has several roles. First, it is recommended as a routine after device implantation. Secondly, it is an option that is sometimes offered to patients who have a disappointing symptomatic response after device implantation. Thirdly, it is considered a gold standard for validation of the accuracy of alternative techniques for AV optimization.

In the first role, it has been found to be not significantly better than factory default settings, in the well-conducted, well-powered, prospective, randomized controlled trial SMART-AV.[9] A meta-analysis by the Auger/Delgado group confirmed this.[33]

In the second role, it may be worthwhile to observe that the great majority of patients who report symptomatic response after implantation would have reported it without implantation,[34] suggesting that the classification does not reflect a meaningful physiological separation. Further evidence of the responder/non-responder dichotomy being inadvisable is the recognition that of the numerically independent markers of response, response on each correlates poorly with response on the others. Restricting iterative optimization to non-responders is therefore physiologically little different to restricting it to a random subset of ~30%, which might be a first step to halting its use completely.

In the third role, as the gold standard for creation of alternative methods, it is unable to provide a single value in the same patient when viewed by two observers or even the same observer twice, even in the artificially favourable circumstance of all observers viewing exactly the same Doppler traces.

Controversy

During the peer-review process it was highlighted to us that the continuous analysis of agreement between the observers showed the mean bias to be low (−3.6, −2.0, and 1.6 ms, respectively, between the three possible pairs) and that this should be reassuring that the iterative method is satisfactory (see Supplementary material online, Appendix 3). However, it is our opinion that this only shows that across all patients the mean of the distribution of values reported by Observer 1 is similar to the mean of the distribution of values reported by Observer 2, and again both are similar to the mean of the distribution reported by Observer 3. Bias would still be near zero even if each observer simply reported a random number drawn from their own pool, as long as all the three pools had the same mean. In our opinion, the readers should focus on the extent of the patient-by-patient disagreement between the observers, which may be expressed, for example, as the 95% limits of agreement. This was large (Ob1 vs. Ob2 −52 to 59 ms, Ob1 vs. Ob3 −61 to 64 ms, Ob2 vs. Ob3 −71 to 68 ms). Thus, once one observer had reported, in all cases, the report from the second observer looking at the identical sets of images would have a 95% range that is >100 ms wide. Whether this extent of variation in interpretation of a fixed set of images is clinically acceptable is the question on which individual clinicians have to decide. If it is considered acceptable, the next test would be blinded test–retest assessment, i.e. a fresh set of images acquired by a staff member unaware of the previous findings and under no pressure to conform. The limits of agreement between test–retest pairs would be wider. Ultimately, it is such test–retest reproducibility that it is the key requirement of clinical measurements intended for programming a patient's device over the long term. Whether a >100 ms wide band of uncertainty is clinically acceptable is for the clinical readership to decide individually.

We do not understand how any other AV optimization method can have been found to agree closely with the iterative method when the iterative method cannot agree with itself. The possibilities for this to occur include inadvertent unblinding of the echo operator to the results of the other optimization method, or vice versa.[35,36]

A reviewer also pointed out that it may be possible to use changes in A-wave size to guide selection of the best mitral inflow pattern when the patients have transmitral Classes A and D in the example profiles in *Figure 2* (see Supplementary material online, Appendix 4). We recognize that some experts may be able to assess mitral inflow patterns in this way, but we are concerned that applying the guidelines in this manner merely allows exclusion of very poor AVD settings without accurately pin-pointing the 'true' optimum. While one expert may be confident, there is no guarantee that different experts (or even the same expert on a separate blinded viewing) may not be equally confident that a different AVD is optimal. When we previously conducted a study with a small number of datasets and a large number of experts, we found that they all identified the optima confidently but nevertheless they extensively disagreed with each other—and with themselves when re-shown the same images.[37]

Thus, the degree of confidence of the observer in the optimum selected does not confer likelihood of other observers choosing the same optimum.

Unrecognized accessing of collateral information

The observers were significantly more likely to select an AVD inside the physiological range of 80–160 ms when the AVD was unmasked. This suggests that, unknowingly, the observers were drawing on this collateral information and prior beliefs, which may partially explain the marginally improved level of agreement between the observers while unblinded to the AVD. This habit may be widespread in clinical practice. We do not criticize clinicians for using common sense to act in the best interests of patients, but the observation raises a challenge to the true informativeness of Doppler optimization. If Doppler alone cannot reject extreme AVDs confidently, we do not understand how it can reliably make subtle distinctions within the plausible range.

Should the iterative method be used in association with a quantitative method?

In cases where the iterative method cannot distinguish the desirability of the Doppler patterns of two or three settings, it might be suggested that a quantitative method be used alongside it, for example, a Doppler assessment of blood flow through the heart, such as Velocity Time Integral (VTI), through the LV outflow tract, transmitral VTI, or timing of LV filling. Previous work suggests that mitral inflow VTI agrees well with invasive haemodynamic measurements during AVD optimization immediately post-implant,[38] however, this method of optimization is not recommended in the guidelines.[4–6] Our data and the previous work have found that there are also inherent difficulties with the quantitative echocardiographic methods of optimization[39] and because this study was focused on the iterative technique, we did not include any quantitative methods of assessing the AVD optima in this study.

Nevertheless, once a quantitative method has been selected, there are two simple tests to conduct before trying to combine them. First, for patients in whom the iterative method makes a clear recommendation, it is useful to test under blinded conditions whether the quantitative method consistently recommends the same AVD as the iterative method. If not, then it is unwise to attempt to combine them without further research into the origin of the discrepancy.

Secondly, if the quantitative method is used to resolve subtle differences irresolvable by the iterative method, an explanation is required for why the iterative method is needed at all.

Studies comparing the AV optima between methods, reporting the actual individual values in individual patients, are surprisingly scarce, with reports focusing most often merely on whether the averages are the same, or the resulting VTI's are the same. Indeed, even test–retest reproducibility of the optimum by any quantitative method, conducted under scientific conditions of blinding, is scarce. Our analysis suggests that reproducible optimization by these quantitative methods requires averaging a number of replicate measurements that is inhumanly large.[39]

No unambiguous protocol for using the iterative method in combination with a quantitative method has been published, and there is certainly no data to suggest that such a combination improves

irreproducibility of the optima produced by either method. Our data and the previous work[39] suggests that an attempt to collect such evidence would be misguided. Therefore, we did not include any quantitative methods of assessing the AVD optima in this study.

Study limitations

Our systematic classification is imperfect. Two observers disagreed on 7 of the 50 classifications. However, the ability to perfectly classify every patient is less important than the open recognition that different classes of behaviour exist and that this may partly account for the very poor level of agreement between the observers.

This was not a study of test–retest reproducibility of the iterative optimum, in which the two clinical assessments would need to be made: 'other day, other hands, other eyes'. Test–retest reproducibility is unlikely to be better than the repeatability determined using identical images and would almost certainly be worse because of unavoidable practical confounders such as differences in probe or patient position and physiological variation.

Prior to enrolment on this study we did not assess patients for factors such as LV diastolic function and LV dyssynchrony, which may affect transmitral Doppler filling pattern. We aimed to have a representative sample and iterative optimization is intended for use in patients who quite often have these features. We did not include or exclude individual patients on an *adhoc* basis, because of the danger of inadvertent bias.[35]

All the patients in this study had had their CRT devices implanted some time prior to data collection, but this passage of time should not make it less possible to carry out the guidelines, since the necessary elements—namely atrial and ventricular contraction, transmitral flow, and pacing—are still present. Furthermore, our patients had relatively mild symptoms, with almost all in NYHA Class II. Having milder symptoms and potentially having undergone favourable remodelling again, however, should not be expected to enhance the proportion of patients with a non-guideline-described Doppler pattern. None of the previous trials, which have original data from sequential patients collected soon after device implantation, had published consecutive unedited Doppler datasets and therefore we do not know for certain what the patterns at such an earlier time point, and in sicker patients, might be.

This is not a large study, but, with 50 patients, 300 assessments, and full disclosure of all Doppler traces, it is large enough to establish the approximate proportion of patients with different types of transmitral filling and is a useful public dataset. Moreover, it was designed to attempt to reduce bias[40] by using three mutually blinded observers who interpreted each dataset twice on separate occasions.

Conclusion

Only 28% of the CRT patients show the typical spectrum of transmitral Doppler filling patterns across a range of the AVDs described in the clinical guidelines and commonly published research protocols, and which allow confident application of the iterative method for AVD optimization.

Observers agree poorly when selecting an optimal AVD using the iterative method. Agreement is at its worst when the transmitral Doppler inflow patterns do not fit with the changes expected and reported schematically (or with data from only one example patient) in the guidelines and research protocols.

When the observers are unblinded to the tested AVD, in the majority of patients, the iterative method may successfully allow rejection of extreme AVDs and selection of an optimal from within a narrower range. However, we have shown that the observers are influenced by collateral information, such as knowledge of the AVD, when making their selections.

Therefore, using the iterative method as a reference to develop new methods may be unsafe because observers (and protocol planners) are not aware of these facts. A new method studied rigorously cannot agree very closely with the iterative result, since the iterative result does not agree very closely with itself. Conversely, studies conducted without rigour (i.e. lacking mutual blinding, or allowing several iterative optima to be offered and the best match accepted) may be the only ones showing strong associations.[35] Thus, paradoxically, reports of strong agreement between a new method and the iterative method may be a marker of poor study design rather than excellence of the new method. Genuine progress in the field of optimization has been minimal in its decade of life so far, and may be faster if the recommendation for the iterative method is rescinded.

Supplementary data

Supplementary material is available at *Europace* online.

Conflict of interest: none declared.

Funding

This work was supported by the British Heart Foundation (SP/10/002/28189).

References

1. Cleland JG, Daubert JC, Erdmann E, Freemantle N, Gras D, Kappenberger L *et al.* The effect of cardiac resynchronization on morbidity and mortality in heart failure. *N Engl J Med* 2005;**352**:1539–49.
2. Abraham WT, Fisher WG, Smith AL, Delurgio DB, Leon AR, Loh E *et al.* Cardiac resynchronization in chronic heart failure. *N Engl J Med* 2002;**346**:1845–53.
3. Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, De Marco T *et al.* Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *N Engl J Med* 2004;**350**:2140–50.
4. Gorcsan J 3rd, Abraham T, Agler DA, Bax JJ, Derumeaux G, Grimm RA *et al.* Echocardiography for cardiac resynchronization therapy: recommendations for performance and reporting—a report from the American Society of Echocardiography Dyssynchrony Writing Group endorsed by the Heart Rhythm Society. *J Am Soc Echocardiogr* 2008;**21**:191–213.
5. Vardas PE, Auricchio A, Blanc JJ, Daubert JC, Drexler H, Ector H *et al.* Guidelines for cardiac pacing and cardiac resynchronization therapy. The Task Force for Cardiac Pacing and Cardiac Resynchronization Therapy of the European Society of Cardiology. Developed in collaboration with the European Heart Rhythm Association. *Europace* 2007;**9**:959–98.
6. Daubert JC, Saxon L, Adamson PB, Auricchio A, Berger RD, Beshai JF *et al.* 2012 EHRA/HRS expert consensus statement on cardiac resynchronization therapy in heart failure: implant and follow-up recommendations and management. *Europace* 2012;**14**:1236–86.
7. Barold SS, Ilercil A, Herweg B. Echocardiographic optimization of the atrioventricular and interventricular intervals during cardiac resynchronization. *Europace* 2008; **10**(Suppl 3):iii88–95.
8. Stein KM, Ellenbogen KA, Gold MR, Lemke B, Lozano IF, Mittal S *et al.* SmartDelay determined AV optimization: a comparison of AV delay methods used in cardiac resynchronization therapy (SMART-AV): rationale and design. *Pacing Clin Electrophysiol* 2010;**33**:54–63.

9. Ellenbogen KA, Gold MR, Meyer TE, Fernndez Lozano I, Mittal S, Waggoner AD *et al.* Primary results from the SmartDelay determined AV optimization: a comparison to other AV delay methods used in cardiac resynchronization therapy (SMART-AV) trial: a randomized trial comparing empirical, echocardiography-guided, and algorithmic atrioventricular delay programming in cardiac resynchronization therapy. *Circulation* 2010;**122**:2660–8.
10. Kamdar R, Frain E, Warburton F, Richmond L, Mullan V, Berriman T *et al.* A prospective comparison of echocardiography and device algorithms for atrioventricular and interventricular interval optimization in cardiac resynchronization therapy. *Europace* 2010;**12**:84–91.
11. Light RJ. Measurement of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bull* 1971;**76**:365–77.
12. Altman DG. *Practical Statistics for Medical Research*. London: Chapman and Hall; 1991
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**i**:307–10.
14. Team RDC. *R: A Language and Environment for Statistical Computing*. Vienna; 2006. Available from: http://www.R-project.org/
15. Falissard B. *psy: Various Procedures used in Psychometry*. R package version 1.0; 2009. Available from: http://cran.r-project.org/package=psy
16. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer; 2009
17. Francis DP. How easily can omission of patients, or selection amongst poorly-reproducible measurements, create artificial correlations? Methods for detection and implications for observational research design in cardiology. *Int J Cardiol* 2013;**167**:102–13.
18. Kedia N, Ng K, Apperson-Hansen C, Wang C, Tchou P, Wilkoff BL *et al.* Usefulness of atrioventricular delay optimization using Doppler assessment of mitral inflow in patients undergoing cardiac resynchronization therapy. *Am J Cardiol* 2006;**98**:780–5.
19. Pavlopoulos H, Nihoyannopoulos P. Recent advances in cardiac resynchronization therapy: echocardiographic modalities, patient selection, optimization, non-responders—all you need to know for more efficient CRT. *Int J Cardiovasc Imaging* 2010;**26**:177–91.
20. Taha N, Zhang J, Ranjan R, Daneshvar S, Castillo E, Guillen E *et al.* Biventricular pacemaker optimization guided by comprehensive echocardiography-preliminary observations regarding the effects on systolic and diastolic ventricular function and third heart sound. *J Am Soc Echocardiogr* 2010;**23**:857–66.
21. Valzania C, Biffi M, Martignani C, Diemberger I, Bertini M, Ziacchi M *et al.* Cardiac resynchronization therapy: variations in echo-guided optimized atrioventricular and interventricular delays during follow-up. *Echocardiography* 2007;**24**:933–9.
22. Vidal B, Sitges M, Marigliano A, Delgado V, Diaz-Infante E, Azqueta M *et al.* Optimizing the programation of cardiac resynchronization therapy devices in patients with heart failure and left bundle branch block. *Am J Cardiol* 2007;**100**:1002–6.
23. Zuber M, Toggweiler S, Quinn-Tate L, Brown L, Amkieh A, Erne P. A comparison of acoustic cardiography and echocardiography for optimizing pacemaker settings in cardiac resynchronization therapy. *Pacing Clin Electrophysiol* 2008;**31**:802–11.
24. Jones RC, Svinarich T, Rubin A, Levin V, Phang R, Murillo J *et al.* Optimal atrioventricular delay in CRT patients can be approximated using surface electrocardiography and device electrograms. *J Cardiovasc Electrophysiol* 2010;**21**:1226–32.
25. Levin V, Nemeth M, Colombowala I, Massumi A, Rasekh A, Cheng J *et al.* Interatrial conduction measured during biventricular pacemaker implantation accurately predicts optimal paced atrioventricular intervals. *J Cardiovasc Electrophysiol* 2007;**18**:290–5.
26. Rafie R, Qamruddin S, Ozhand A, Taha N, Naqvi TZ. Shortening of atrioventricular delay at increased atrial paced heart rates improves diastolic filling and functional class in patients with biventricular pacing. *Cardiovasc Ultrasound* 2012;**10**:2.
27. Naqvi TZ. Echocardiography-guided biventricular pacemaker optimization. *JACC Cardiovasc Imaging* 2010;**3**:1168–80.
28. Waggoner AD, de Las Fuentes L, Davila-Roman VG. Doppler echocardiographic methods for optimization of the atrioventricular delay during cardiac resynchronization therapy. *Echocardiography* 2008;**25**:1047–55.
29. Cuoco FA, Gold MR. Optimization of cardiac resynchronization therapy: importance of programmed parameters. *J Cardiovasc Electrophysiol* 2012;**23**:110–8.
30. Hasan A, Abraham WT. Optimization of cardiac resynchronization therapy after implantation. *Curr Treat Options Cardiovasc Med* 2008;**10**:319–28.
31. Naqvi TZ, Rafique AM. Novel insights on effect of atrioventricular programming of biventricular pacemaker in heart failure—a case series. *Cardiovasc Ultrasound* 2006;**4**:38.
32. Patel V, Giesebrecht S, Burton AR, Cvejic E, Lemon J, Hadzi-Pavlovic D *et al.* Reliability revisited: autonomic responses in the context of everyday well-being. *Int J Cardiol* 2013;**166**:743–5.
33. Auger D, Hoke U, Bax JJ, Boersma E, Delgado V. Effect of atrioventricular and ventriculoventricular delay optimization on clinical and echocardiographic outcomes of patients treated with cardiac resynchronization therapy: a meta-analysis. *Am Heart J* 2013;**166**:20–9.
34. Sohaib SMA, Chen ZC, Whinnett Z, Francis DP, Manisty C. Systematic review of genuine symptomatic response to cardiac resynchronization therapy: acknowledging the contribution of spontaneous response. *Eur Heart J*. [Meeting Abstract] 2012; **33**:995
35. Shun-Shin M, Francis DP. Why are some studies of cardiovascular markers unreliable? The role of measurement variability and what an aspiring clinician scientist can do before it is too late. *Prog Cardiovasc Dis* 2012;**55**:14–24
36. Francis DP. How easily can omission of patients, or selection amongst poorly-reproducible measurements, create artificial correlations? Methods for detection and implications for observational research design in cardiology. *Int J Cardiol* 2013;**167**:102–13.
37. Raphael CE, Kyriacou A, Jones S, Pabari P, Cole G, Baruah R *et al.* Multinational evaluation of the interpretability of the iterative method of optimisation of AV delay for CRT. *Int J Cardiol* 2012; doi: 10.1016/j.ijcard.2012.09.097. [Epub ahead of print].
38. Jansen AH, Bracke FA, van Dantzig JM, Meijer A, van der Voort PH, Aarnoudse W *et al.* Correlation of echo-Doppler optimization of atrioventricular delay in cardiac resynchronization therapy with invasive hemodynamics in patients with heart failure secondary to ischemic or idiopathic dilated cardiomyopathy. *Am J Cardiol* 2006;**97**:552–7.
39. Francis DP. How to reliably deliver narrow individual-patient error bars for optimization of pacemaker AV or VV delay using a 'pick-the-highest' strategy with haemodynamic measurements. *Int J Cardiol* 2013;**163**:221–5.
40. Nijjer SS, Pabari PA, Stegemann B, Palmieri V, Leyva F, Linde C *et al.* The limit of plausibility for predictors of response: application to biventricular pacing. *JACC Cardiovasc Imaging* 2012;**5**:1046–65.