JACC: CARDIOVASCULAR IMAGING
© 2012 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION
PUBLISHED BY ELSEVIER INC.

VOL. 5, NO. 10, 2012
ISSN 1936-878X/$36.00
http://dx.doi.org/10.1016/j.jcmg.2012.07.010

The Limit of Plausibility for Predictors of Response: Application to Biventricular Pacing

Sukhjinder S. Nijjer, BSc, MBChB,* Punam A. Pabari, MBChB, PhD,†
Berthold Stegemann, PhD,‡ Vittorio Palmieri, MD, PhD,§ Francisco Leyva, MD,‖
Cecilia Linde, MD, PhD,¶ Nick Freemantle, PhD,# Justin E. Davies, BSc, MBBS, PhD,*
Alun D. Hughes, MBBS, PhD,** Darrel P. Francis, MA, MD**

London and Birmingham, United Kingdom; Maastricht, the Netherlands; Avellino, Italy; and Stockholm, Sweden

OBJECTIVES We sought a method for any reader to quantify the limit, imposed by variability, to sustainably observable R^2 between any baseline predictor and response marker. We then apply this to echocardiographic measurements of mechanical dyssynchrony and response.

BACKGROUND Can mechanical dyssynchrony markers strongly predict ventricular remodeling by biventricular pacing (cardiac resynchronization therapy)?

METHODS First, we established the mathematical depression of observable R^2 arising from: 1) spontaneous variability of response markers; and 2) test–retest variability of dyssynchrony measurements. Second, we contrasted published R^2 values between externally monitored randomized controlled trials and highly skilled single-center studies (HSSCSs).

RESULTS Inherent variability of response markers causes a contraction factor in R^2 of 0.48 (change in left ventricular ejection fraction [ΔLVEF]), 0.50 (change in end-systolic volume [ΔESV]), and 0.40 (change in end-diastolic volume [ΔEDV]). Simultaneously, inherent variability of mechanical dyssynchrony markers causes a contraction factor of between 0.16 and 0.92 (average, 0.6). Therefore the combined contraction factor, that is, limit on sustainably observable R^2 between mechanical dyssynchrony markers and response, is ~0.29 (ΔLVEF), ~0.24 (ΔESV), and ~0.30 (ΔEDV). Many R^2 values published in HSSCSs exceeded these mathematical limits; none in externally monitored trials did so. Overall, HSSCSs overestimate R^2 by 5- to 20-fold ($p = 0.002$). Absence of bias-resistance features in study design (formal enrollment and blinded measurements) was associated with more overstatement of R^2.

CONCLUSIONS Reports of $R^2 > 0.2$ in response prediction arose exclusively from studies without formally documented enrollment and blinding. The HSSCS approach overestimates R^2 values, frequently breaching the mathematical ceiling on sustainably observable R^2, which is far below 1.0, and can easily be calculated by readers using formulas presented here. Community awareness of this low ceiling may help resist future claims. Reliable individualized response prediction, using methods originally designed for group-mean effects, may never be possible because it has 2 currently unavailable and perhaps impossible prerequisites: 1) excellent blinded test–retest reproducibility of dyssynchrony; and 2) response markers reproducible over time within nonintervened individuals. Dispassionate evaluation, and improvement, of test–retest reproducibility is required before any further claims of strong prediction. Prediction studies should be designed to resist bias. (J Am Coll Cardiol Img 2012;5:1046–65) © 2012 by the American College of Cardiology Foundation

From the *Department of Cardiology, Hammersmith Hospital, London, United Kingdom; †Department of Echocardiography, St Mary's Hospital, London, United Kingdom; ‡Medtronic Bakken Research Center, Medtronic Inc., Maastricht, the

Biventricular pacing is thought to deliver benefit in heart failure through resynchronization of dyssynchronous cardiac mechanical function, hence the term *cardiac resynchronization therapy* (1–6). Some studies (7–9) demonstrate strong relationships (high coefficient of determination, R^2 values) between baseline mechanical dyssynchrony and echocardiographic outcome measures, whereas others (10–12) show much weaker relationships. Most guidelines for selecting patients for biventricular pacing emphasize electrical dyssynchrony manifested as wide QRS duration rather than mechanical dyssynchrony (13), although pressure is growing from increasing numbers of positive studies reporting an association between baseline dyssynchrony and ventricular response. One country's guidelines already include mechanical dyssynchrony in selection (14).

Tantalizing glimpses of reliable prediction of response continue to drive the search for mechanical dyssynchrony markers or multivariate combination algorithms to provide better prediction. But is this approach wise? Why do studies disagree? Reports of R^2 exceeding 1.0 would be recognized as incorrect, but is the real upper limit of *sustainably* observable R^2 really 1.0, or something lower? How can one calculate the highest plausible R^2 between dyssynchrony and response? How should we interpret this clinically, and does it affect how we design future research?

The ceiling on R^2 depends upon natural variability of dyssynchrony markers and of response markers. Blinded test–retest reproducibility data on mechanical dyssynchrony markers (15,16) and commonly used outcome markers of reverse remodeling are scarce. In this study we collate these and thereby calculate the true upper limit on plausible sustainably observable R^2 between dyssynchrony markers and echocardiographic response.

We evaluate the implications for design and interpretation of studies seeking clinically reliable markers of mechanical dyssynchrony in particular, and for studies making claims of individualized prediction of response to any intervention in general.

METHODS

Quantitative separation of device-mediated, versus spontaneous, changes in left ventricular ejection fraction (LVEF). Randomized trials of biventricular pacing are the best way to separate spontaneous changes from device-induced changes in cardiac function. Patients undergoing biventricular pacing have 2 drivers of pre-to-post change in the chosen echocardiographic outcome measure (e.g., change in left ventricular ejection fraction [ΔLVEF]). First, inherent phenomena unrelated to biventricular pacing, including true biological variability and measurement error, will contribute to individual patients' ΔLVEF. The variance (square of standard deviation [SD^2]) of ΔLVEF in the *control* patients in a randomized trial measures the size of this inherent scatter between successive measurements over time. Second, the device itself imposes an effect over and above the inherent variation. Because different patients (presumably) gain different amounts of effect from the device, ΔLVEF is more widely spread in the *device* patients than the control patients. The extra variance in ΔLVEF in the device patients is the variance caused by the device (Fig. 1). Only this extra variance, caused by the device, has any hope of being predicted by baseline dyssynchrony.

Meanwhile, baseline dyssynchrony markers also have inherent variability *within* a given patient over time. Only test–retest reproducibility studies reveal the extent of this. This is also true of multivariate combination algorithms used to score dyssynchrony because each component contributes something to inherent variability.

ABBREVIATIONS AND ACRONYMS

CI = confidence interval

EM-RCT = externally monitored randomized controlled trials

HSSCS = highly skilled single-center studies

LVEDV = left ventricular end-diastolic volume

LVEF = left ventricular ejection fraction

LVESV = left ventricular end-systolic volume

SD = standard deviation

Netherlands; §Department of Heart and Vessels, Azienda Ospedaliera di Rilevanza Nazionale e di alta specialità "S.G.Moscati," Avellino, Italy; ‖Centre for Cardiovascular Sciences, Queen Elizabeth Hospital, University of Birmingham, Birmingham, United Kingdom; ¶Department of Cardiology, Karolinska University Hospital, Stockholm, Sweden; #Department of Primary Care and Population Health, University College London, London, United Kingdom; and the **International Centre for Circulatory Health, Imperial College London, London, United Kingdom. Research by Dr. Nijjer is supported by the Medical Research Council (grant no. G110043). Research by Drs. Francis and Pabari is supported by the British Heart Foundation (grant nos. FS/10/038 and PG/08/114, respectively). Dr. Leyva has received financial support for research from Medtronic, St. Jude Medical, Sorin, and Boston Scientific. Dr. Linde is a consultant to Medtronic, is the principal investigator for the REVERSE trial, sits on the Data Safety and Monitoring Boards of Biopace study sponsored by St. Jude; and a consultant to CARDIOMEMS. Dr. Freemantle sits on the Data Safety and Monitoring Boards for Medtronic Trials and has received consultancy and travel fees for Medtronic Data Safety and Monitoring Boards. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received September 7, 2011; revised manuscript received July 5, 2012, accepted July 9, 2012.

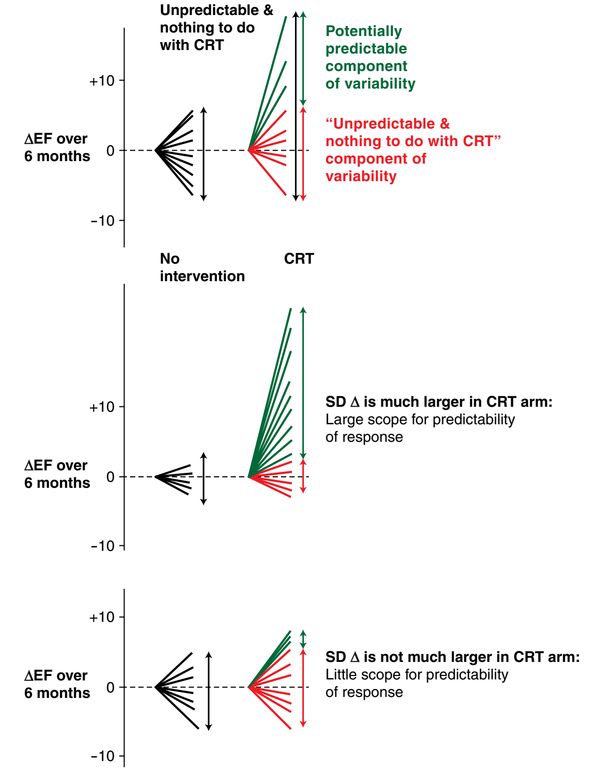
Unpredictable & nothing to do with CRT
Potentially predictable component of variability
"Unpredictable & nothing to do with CRT" component of variability
ΔEF over 6 months
+10
0
−10
No intervention
CRT
SD Δ is much larger in CRT arm: Large scope for predictability of response
SD Δ is not much larger in CRT arm: Little scope for predictability of response

Figure 1. Only Part of the Spectrum of Response in the Intervention Group Is Attributable to the Intervention

Control populations in randomized controlled trials of cardiac resynchronisation therapy (CRT) have changes in their outcome markers (e.g., left ventricular ejection fraction [LVEF], left ventricular end-systolic volume [LVESV], and left ventricular end-diastolic volume [LVEDV]) even without intervention. This change is inherent and unpredictable but can be measured by the variance (square of standard deviation [SD^2]) of the *change* in LVEF (ΔLVEF) **(top row, left panel)**. Those undergoing CRT have a further change in LVEF over and above inherent changes, widening the spread or variance in ΔLVEF **(top row, right panel)**. Only the incremental change above inherent change is attributable to the device and only this component of variance of ΔLVEF is predictable by any baseline marker. (Although it is the variances of ΔLVEF that matter, the figure displays SDs as a visual proxy.) When the SD of the ΔLVEF (SDΔ) is much larger in the biventricular pacing arm of a trial than in the control arm **(middle row)**, there is good scope for predicting response. When the SDΔ is not much larger in the biventricular pacing arm than in the control arm **(bottom row)**, the scope for predictability is much smaller.

When correlating (r) two variables, such as mechanical dyssynchrony and echocardiographic response, or determining the predictive value of one on the other (R^2), the measurement variability of both combine to depress the observable relationship strength (Fig. 2) (17). We term this the R^2 *contraction factor* because of the following relationship (Equation 1):

$$\text{Observed } R^2 = \text{Underlying } R^2 \times R^2 \text{ Contraction Factor} \quad [1]$$

where *underlying* R^2 is the potential correlation between the variables if all measurement noise could be eradicated. The online appendix shows full details. The R^2 contraction factor is also a ceiling on sustainably observable R^2 values because the underlying R^2 cannot exceed 1.0.

The R^2 contraction factor has 2 contributors: 1) contraction from response irreproducibility; and 2) contraction from dyssynchrony irreproducibility. Both are easy to calculate if data are available. Calculating the R^2 contraction factor arising from response irreproducibility requires the SD of the Δ in the outcome measure in both the control arm and device arm of a randomized control trial. It is not sufficient to know the distribution of the initial and final LVEFs. Rather, the distribution of *the change,* that is, the SD of Δ, is needed. This can be used in the following calculation (Equation 2):

R^2 contraction factor caused by inherent variability in response variable =

$$1 - \left[\frac{\mathrm{SD}(\Delta_{\text{control arm}})}{\mathrm{SD}(\Delta_{\text{intervention arm}})}\right]^2 \quad [2]$$

A similar formula is used for the mechanical dyssynchrony measure. The 2 contraction factors are then multiplied to determine the combined contraction factor (Equation 3).

$$\text{Observed } R^2 = \text{Underlying } R^2 \times \frac{\text{Combined } R^2 \text{ Contraction Factor}}{\begin{array}{c} R^2 \text{ contraction factor} \\ \text{imposed by} \\ \textit{Dyssynchrony Marker} \end{array} \times \begin{array}{c} R^2 \text{ contraction factor} \\ \text{imposed by} \\ \textit{Response Marker} \end{array}} \quad [3]$$

The MIRACLE-ICD II (Multicenter InSync ICD Randomized Clinical Evaluation II) trial (18) can be used as a worked example. In the control arm, ΔLVEF has an SD of 6.2; in the biventricular pacing arm, ΔLVEF has an SD of 8. Therefore, the contraction factor imposed on R^2 by the response marker LVEF is $1 - (6.2/8)^2 = 0.40$. Thus in populations like those in MIRACLE-ICD, even with an imaginary perfectly comprehensive and perfectly reproducible dyssynchrony marker, the highest R^2 that could be sustainably observed with ΔLVEF would be 0.40.

In reality, mechanical dyssynchrony markers or scores do not have perfect test–retest reproducibility and so impose their own contraction factor. If, for example, the dyssynchrony marker imposed a contraction factor of 0.50, then the combined contraction factor would be $0.40 \times 0.50 = 0.20$. This means that even if the marker

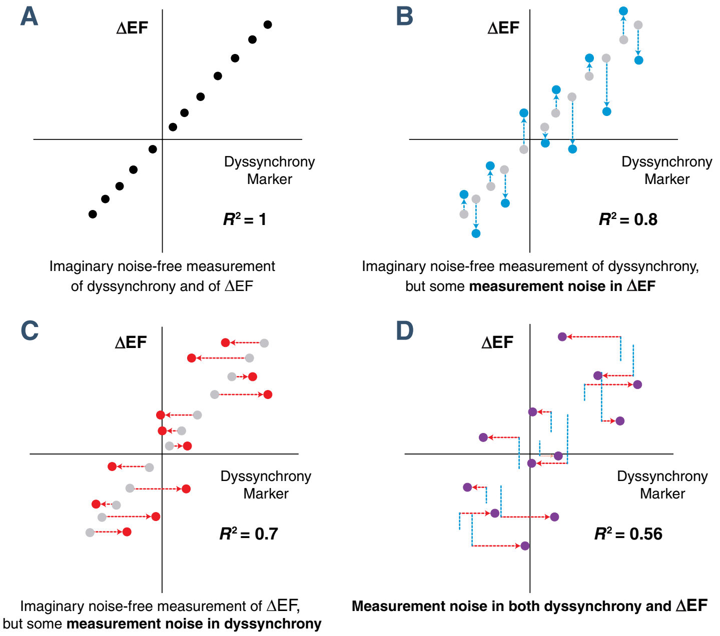

Figure 2. How Inherent Variability in 2 Measures Reduces the Maximum Achievable R^2 Between Them

Imagine a universally comprehensive dyssynchrony marker that can perfectly predict response, as long as measurement noise could somehow be eliminated **(A)**. In practice there is natural variability in measurements of ejection fraction **(B)** and of the dyssynchrony marker **(C)**. These noise properties combine together multiplicatively to depress the actually observable R^2 value **(D)**. In this example case, it is mathematically impossible for an R^2 value over 0.56 to be observed sustainably.

is completely comprehensive in describing all aspects of dyssynchrony (and there are no confounding features, e.g., scar or lead position), the maximum R^2 observable is still only 0.20.

Data extraction from published studies. A systematic review of studies assessing the response to biventricular pacing was performed using the EMBASE and MEDLINE databases (Fig. 3). The terms *cardiac resynchronization therapy, biventricular pacing,* and *dyssynchrony* were used and abstracts reviewed for relevance.

All published studies that assessed mechanical dyssynchrony markers against ΔLVEF, ΔLV end-systolic volume (ESV), and ΔLV end-diastolic volume (EDV) were analyzed and had R^2 data extracted (7–9) (Online Appendix References 1–40). R^2 was calculated where necessary by squaring the correlation coefficient between the mechanical dyssynchrony marker and outcome measure, using the published data in tabular, text, or graphic form. Weighted averages of the R^2 were calculated using the size of the study.

Studies reporting mechanical dyssynchrony markers were assessed to determine the test–retest variability of the markers and whether data were collected after formal enrollment with blinding. Studies that report SD within 1 patient and the SD across the population can have the R^2 contraction factor from the mechanical dyssynchrony marker calculated as $1 - (SD_{within\text{-}patient}/SD_{between\text{-}patient})^2$. Where test–retest variability is given as a correlation coefficient r, it was used as an estimate of the R^2 contraction factor imposed by the dyssynchrony marker.

The landmark externally monitored randomized controlled trials (EM-RCTs) of biventricular pacing were assessed to determine the SD of ΔLVEF, ΔLVESV, and ΔLVEDV in the control and intervention arms (Online Appendix References 41–58). This enabled calculation of contraction factors of these response markers from rigorously performed, formally recruited, externally monitored heart failure populations.

Statistics. Values are shown as mean (95% confidence interval [CI]), except where otherwise indicated. Comparisons between classes of study were made using the Student unpaired t test and the Mann-Whitney U test. A p value <0.05 was pre-defined as statistically significant. Stata/SE version 10.0 (StataCorp LP, College Station, Texas) was used to perform the statistical analysis.

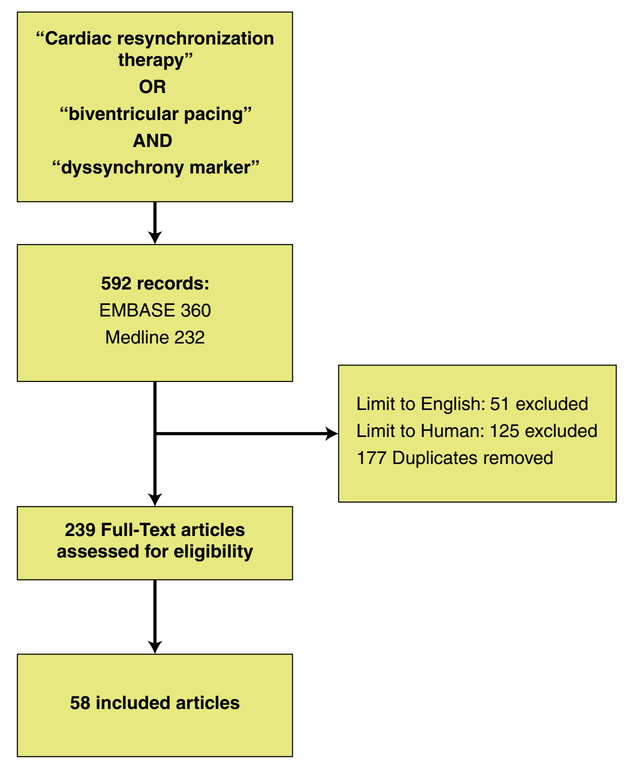
"Cardiac resynchronization therapy"
OR
"biventricular pacing"
AND
"dyssynchrony marker"
592 records:
EMBASE 360
Medline 232
Limit to English: 51 excluded
Limit to Human: 125 excluded
177 Duplicates removed
239 Full-Text articles assessed for eligibility
58 included articles

Figure 3. Systematic Search Strategy

A systematic review of the literature was undertaken to identify studies using baseline echocardiographic mechanical dyssynchrony markers to predict change in echocardiographic response markers (ΔLVEF, ΔLVESV, and ΔLVEDV). Abbreviations as in Figure 1.

RESULTS

Reported R^2 for echocardiographic response in EM-RCTs and highly skilled single-center studies. Fifty-eight reports were identified and assessed. The majority were retrospective cohort studies with or without a control group, performed in highly skilled single centers (HSSC) with specific interest in echocardiographic dyssynchrony markers and a track record of innovation in the field (Online Appendix references 41–58). The reported R^2 in these studies between individual dyssynchrony markers and echocardiographic response to biventricular pacing (ΔLVEF, ΔLVESV, or ΔLVEDV) are tabulated in Table 1.

R^2 values were weighted according to etiology of heart failure (ischemic heart disease vs. idiopathic); no statistically significant difference was found between the 2 groups ($p = 0.38$).

EM-RCTs establishing the use of biventricular pacing were assessed. Primary and secondary publications report a wide variety of potential R^2 between the outcome of biventricular pacing and baseline measures of dyssynchrony (Table 1) (Online Appendix references 41–58). The reported R^2 values found in EM-RCTs were significantly smaller than those found in the HSSC studies (HSSCSs) ($p = 0.02$) for response in ΔLVEF (0.40 vs. 0.07), ΔLVESV (0.24 vs. 0.06), and ΔLVEDV (0.53 vs. 0.01) (Fig. 4).

R^2 contraction factor arising from outcome variable. EM-RCTs provided data sufficient to estimate the R^2 contraction factor for the commonly used echocardiographic outcome measures (Table 2) and clinical response markers (Table 3). All 3 echocardiographic outcome variables (ΔLVEF, ΔLVESV, and ΔLVEDV) have sufficient variability in the control populations to give R^2 contraction factors that limit observed R^2 to modest values, even if the predictive dyssynchrony marker or combination algorithm was perfect and had no variability.

R^2 contraction factor arising from the dyssynchrony variable. We assessed the published variability of mechanical dyssynchrony markers between repeated echocardiograms in the same patient (test–retest reproducibility) (Table 4). Only 3 studies report the true test–retest variability needed to calculate the contraction factor for each dyssynchrony marker. In one, within-patient variation and between-patient variation was small (15). The second assessed test–retest reproducibility of tissue Doppler imaging mechanical dyssynchrony markers and presented the R^2 contraction factor directly when measured by 2 separate readers, giving an average value of 0.35 (16). The third randomized patients to biventricular pacing or medical therapy and reported the change in dyssynchrony indexes remeasured in the control population (70). The mean change and its SD were provided to us by the authors. Overall, the available contraction factors range from 0.16 to 0.92, averaging ~0.6.

Combined R^2 contraction factor. The combined R^2 contraction factor between echocardiographic response and a dyssynchrony marker is calculated by multiplication. We estimate that for ΔLVEF it is 0.29 (0.6 × 0.48); for ΔLVEDV, 0.24 (0.6 × 0.40); and for ΔLVESV, 0.30 (0.6 × 0.50). These are point estimates and may overestimate or underestimate the true combined contraction factor. Table 5 displays the likely values with the most likely region in boldface.

Comparing study design between HSSCSs and EM-RCTs. We assessed whether studies specified 3 key design features that limit bias: 1) predictive marker stated to be measured blinded to outcome; 2) outcome marker stated to be measured blinded to the patient's treatment with biventricular pacing; and 3) patients stated to be formally enrolled before the

Table 1. A Comparison of the Baseline Dyssynchrony Variables Found To Predict Response in Externally Monitored Randomized Controlled Trials With Those Found Highly Skilled Single-Center Studies

Response Measure/ Study/First Author/Year	Baseline Variable	N	Ischemia		DCM		Observed Correlation Coefficient*	R^2*	95% CI	R^2 Ischemia	R^2 DCM	p Value
			n	%	n	%						
							Highly Skilled Specialist Center Studies					
ΔLVEF												
Bax 2003	TDI septal-lateral delay	25	11	44	14	56	0.47	**0.22**	(0.01–0.53)			
Pitzalis 2005	SPWMD	51	11	22	40	78	0.69	**0.48**	(0.26–0.66)			
Marsan 2008	SDI	56	35	62	21	38	0.7	**0.49**	(0.29–0.66)			
Weighted average								**0.42**		0.40	0.40	0.59
%ΔLVEF												
Penicka 2004	Sum asynchrony	49	23	47	26	53	0.73	**0.53**	(0.32–0.70)			
Mele 2006	SPWMD	37	16	43	21	57	0.07	**0.01**	(0.00–0.15)			
	TPS-SD	37	16	43	21	57	0.86	**0.74**	(0.55–0.86)			
	SPWTD	37	16	43	21	57	0.53	**0.28**	(0.06–0.53)			
Lim 2011	SDI	189	63	33	126	67	0.45	**0.20**	(0.11–0.31)			
Weighted average								**0.29**		0.36	0.34	0.61
ΔLVEDV												
Pitzalis 2002	SPWMD	20	4	20	16	80	−0.73	**0.53**	(0.18–0.79)			
Weighted average								**0.53**		0.53	0.53	1.00
ΔLVESV												
Pitzalis 2002	SPWMD	20	4	20	16	80	−0.7	**0.49**	(0.14–0.76)			
Yu 2003	Ts-SD	30	12	40	18	60	−0.76	**0.58**	(0.30–0.77)			
Bax 2004	TDI septal-lateral delay	80	44	55	36	45	0.84	**0.71**	(0.58–0.80)			
Yu 2004	TDI-Ts-SD	54	22	41	32	59	−0.74	**0.55**	(0.35–0.71)			
	TDI-Ts-12	54	22	41	32	59	−0.6	**0.36**	(0.16–0.56)			
	SRI-Tsr-SD	54	22	41	32	59	−0.01	**0.00**	(0.00–0.08)			
	PSS-12	54	22	41	32	59	−0.28	**0.08**	(0.00–0.26)			
Yu 2005	Ts-SD-12-ejection	56	28	50	28	50	−0.61	**0.37**	(0.17–0.57)			
	Ts-SD-6-ejection	56	28	50	28	50	−0.52	**0.27**	(0.09–0.47)			
	Ts-12-ejection	56	28	50	28	50	−0.6	**0.36**	(0.16–0.56)			
	Ts-6-ejection	56	28	50	28	50	−0.53	**0.28**	(0.10–0.48)			
Porciani 2006	Ts-SD	59	30	51	29	49	−0.32	**0.10**	(0.00–0.28)			
	oExcT	59	30	51	29	49	−0.48	**0.23**	(0.07–0.43)			
Yu 2006	TDI-SD-12	55	28	51	27	49	−0.76	**0.58**	(0.38–0.73)			
	TDI-SD-6	55	28	51	27	49	−0.63	**0.40**	(0.19–0.59)			
	Diff-12	55	28	51	27	49	−0.64	**0.41**	(0.20–0.60)			
	Diff-6	55	28	51	27	49	−0.62	**0.38**	(0.18–0.58)			
	Sep-Lat	55	28	51	27	49	−0.54	**0.29**	(0.10–0.50)			
	Sep-Post	55	28	51	27	49	−0.49	**0.24**	(0.07–0.45)			
Marsan 2008	SDI	56	35	62	21	38	0.6	**0.36**	(0.16–0.56)			
Bank 2009	IVCT	64	40	63	24	37	−0.24	**0.06**	(0.00–0.21)			
	TT_{S-L} delay	64	40	63	24	37	−0.14	**0.02**	(0.01–0.14)			
	TVI_{S-L} delay	64	40	63	24	37	−0.12	**0.01**	(0.02–0.13)			
	SD Rad-6	64	40	63	24	37	−0.2	**0.04**	(0.00–0.18)			
Soliman 2009	SDI (3DE)	90	46	51	44	49	0.56	**0.31**	(0.16–0.47)			
Van Bommel 2010	LV dyssynchrony	361	221	61	140	39	0.38	**0.14**	(0.08–0.22)			
Miyazaki 2010	SPWMD	117	66	56	51	44	0.32	**0.10**	(0.02–0.22)			
	S-L delay	117	66	56	51	44	−0.05	**0.00**	(0.02–0.05)			
	Tv-SD	117	66	56	51	44	0.15	**0.02**	(0.00–0.10)			
	$T\varepsilon$-SD	117	66	56	51	44	0.38	**0.14**	(0.05–0.28)			
	$T\varepsilon$-dif	117	66	56	51	44	0.43	**0.18**	(0.07–0.32)			
Weighted average								**0.22**		0.25	0.28	0.76

Continued on next page

Table 1. Continued

Response Measure/Study/First Author/Year	Baseline Variable	N	Ischemia		DCM		Observed Correlation Coefficient*	R^2*	95% CI	R^2 Ischemia	R^2 DCM	p Value
			n	%	n	%						
%ΔLVESV												
Mele 2006	SPWMD	37	16	43	21	57	−0.42	**0.18**	(0.01–0.61)			
	TPS-SD	37	16	43	21	57	−0.01	**0.0001**	(0.00–0.11)			
	SPWTD	37	16	43	21	57	−0.73	**0.53**	(0.28–0.73)			
Delgado 2008	AS-P (radial strain)	161	92	57	69	43	0.41	**0.17**	(0.07–0.28)			
	SDt_{65s}	161	92	57	69	43	0.26	**0.07**	(0.01–0.16)			
Lim 2008	SD-TDI	65	25	39	40	61	0.2	**0.04**	(0.00–0.18)			
	T_{SL}	65	25	39	40	61	0.2	**0.04**	(0.00–0.18)			
	12SD-ε	65	25	39	40	61	−0.38	**0.14**	(0.02–0.33)			
	SDI	100	35	35	65	65	−0.69	**0.48**	(0.33–0.61)			
Wang 2010	RDI-6 basal segments	30	12	40	18	60	0.42	**0.18**	(0.00–0.46)			
	RDI-6 mid-LV segments	30	12	40	18	60	0.75	**0.56**	(0.29–0.76)			
	RDI-12 (combination)	30	12	40	18	60	0.62	**0.38**	(0.11–0.64)			
Lim 2011	12SD-ε	189	63	33	126	67	0.18	**0.03**	(0.00–0.10)			
	Global strain	189	63	33	126	67	0.2	**0.04**	(0.00–0.11)			
	SDI	189	63	33	126	67	0.61	**0.37**	(0.26–0.48)			
Weighted average								**0.18**		0.21	0.22	0.38
ΔNYHA												
No HSSCS reported *r* or R^2 of a dyssynchrony marker to ΔNYHA												
ΔQoL												
No HSSCS reported *r* or R^2 of a dyssynchrony marker to ΔQoL												
Externally Monitored Randomized Controlled Trials												
ΔLVEF												
CONTAK-CD: Marcus 2005	SPWMD	79	57	72	22	28	−0.11	**0.01**	(0.01–0.10)			
MADIT-CRT: Pouleur 2011	Transverse strain dyssynchrony	761	416	55	345	45	−0.29	**0.08**	(0.05–0.13)			
Weighted average								**0.07**		0.04	0.05	0.81
%ΔLVEF												
No EM-RCT reported this outcome												
ΔLVEDV												
MIRACLE: Cappola 2006	Mitral regurgitation index	776	463	60	313	40	0.0022	**0.000005**	(0.00–0.00)			
MIRACLE: Cappola 2006	QRS width	776	463	60	313	40	0.12	**0.01**	(0.00–0.04)			
CONTAK-CD: Marcus 2005	SPWMD	79	57	72	22	28	−0.14	**0.02**	(0.00–0.12)			
MADIT-CRT: Pouleur 2011	Transverse strain dyssynchrony	761	416	55	345	45	0.25	**0.06**	(0.03–0.10)			
Weighted average								**0.01**		0.02	0.004	0.6
ΔLVESV												
CARE-HF: Ghio 2009	IVMD	365	168	46	197	54	Not published	—				
CONTAK: Marcus 2005	SPWMD	79	57	72	22	28	−0.1	**0.01**	(0.02–0.10)			
REVERSE: Linde 2009	IVMD	419	236	56	183	44	Not published	—				
MADIT-CRT: Pouleur 2011	Transverse strain dyssynchrony	761	416	55	345	45	0.25	**0.06**	(0.03–0.10)			
Weighted average								**0.06**		0.03	0.04	0.8

Continued on next page

Table 1. Continued

Response Measure/ Study/First Author/Year	Baseline Variable	N	Ischemia n	Ischemia %	DCM n	DCM %	Observed Correlation Coefficient*	R^2*	95% CI	R^2 Ischemia	R^2 DCM	p Value
%ΔLVESV												
No EM-RCT reported this outcome												
ΔNYHA												
MIRACLE: St John Sutton 2003	IVMD	323	~162	50	~161	50	0.12	**0.01**	(0.00–0.05)			
ΔQoL												
MIRACLE: St John Sutton 2003	IVMD	323	~162	50	~161	50	0.19	**0.04**	(0.01–0.09)			
Weighted average								**0.03**		0.03	0.03	1.00

*Correlation coefficients and coefficient of determination (R^2) as reported by studies assessing the ability of baseline mechanical dyssynchrony markers to predict change in response markers are shown, arranged by study design. A weighted average for the R^2 values is reported, together with the calculated 95% confidence intervals of each R^2. R^2 values when weighted for heart failure etiology (ischemic or dilated cardiomyopathy [DCM]) were not significantly different. Boldface represents the R^2 value calculated from each study between stated dyssynchrony marker and response marker. Please see the Online Appendix for reference citations.

CARE-HF = Cardiac Resynchronization in Heart Failure study; CI = confidence interval; CONTAK-CD = CONTAK CD Device Approval Study; EM-RCT = externally monitored randomized controlled trials; HSSCS = highly skilled single-center studies; LVEDV = left ventricular end-diastolic volume; LVEF = left ventricular ejection fraction; LVESV = left ventricular end-systolic volume; MADIT-CRT = Multicenter Automatic Defibrillator Implantation Trial With Cardiac Resynchronization Therapy; MIRACLE = Multicenter InSync ICD Randomized Clinical Evaluation; NYHA = New York Heart Association classification; QoL = quality of life; REVERSE = Resynchronization Reverses Remodeling in Systolic Left Ventricular Dysfunction study; other abbreviations are listed in the original publications.

measurements were made (Table 6). The majority of the EM-RCTs report some degree of blinding; almost all of the HSSCSs did not (odds ratio: 70 [95% CI: 6 to 777; $p < 0.0001$] for response markers; odds ratio: 32 [CI: 3 to 306; $p < 0.01$] for dyssynchrony markers).

To determine the impact of publication bias a funnel plot was performed (Fig. 5). Study size had a weak but positive relationship with the publishing larger R^2 values (r^2 for this relationship = 0.25, $p < 0.01$) indicating the evidence of publication bias favoring positive studies.

To determine the impact of study design that affected R^2, we plotted R^2 against study size and the number of bias-resistance features (Fig. 6). Each study scored 0 or 1 point for each of 3 features. When the presence of a feature was unclear, no point was given. While smaller studies showed a tendency to report higher R^2 values than larger studies, the number of bias-resistance features showed a stronger influence. Studies designed from the outset to resist bias never showed high R^2 values. Very high R^2 values occurred exclusively in studies that described little or nothing done to resist bias.

DISCUSSION

This article shows how to determine the ceiling on the sustainable R^2 values between any baseline marker and subsequent response to any intervention, such as biventricular pacing. Readers can use this ceiling, or contraction factor, to judge whether an R^2 is credible. The greater the test–retest variability in any predictor and/or response—whether due to biological factors, measurement error, or

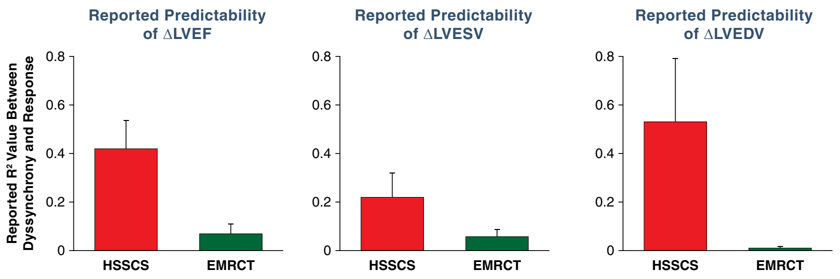

Figure 4. Comparison of the Reported R^2 Values Between Baseline Dyssynchrony Markers and Echocardiographic Response Markers by Different Study Design

Comparison of highly skilled single-center studies (HSSCSs) and externally monitored randomized controlled trials (EM-RCTs) for reported R^2 values between baseline dyssynchrony markers and echocardiographic markers of response to biventricular pacing. There is a significant difference between associations reported by HSSCSs and EM-RCTs ($p = 0.02$).

Table 2. Calculation of the Contraction Factor for ΔLVEF, ΔLVESV, and ΔLVEDV in Externally Monitored Randomized Controlled Trials Assessing Biventricular Pacing

		Breakdown of Variability		Mandatory Ceiling on R^2 Value Imposed Solely by Unpredictable Variability in Response Measure		
					Maximal Achievable R^2 Value*	
Response Measure/Study	**N**	**Unpredictable Element of Variability: SDΔ of Control Arm**	**Total Variability in Intervention Arm: SDΔ of Intervention Arm**	**Calculation**	**R^2**	**95% CI**
ΔLVESV						
MIRACLE-ICD II	153	57.0	77.0	$1-(57/77)^2$	**0.45**	(0.33–0.56)
CARE-HF	735	42.8	66.6	$1-(42.8/66.6)^2$	**0.59**	(0.54–0.63)
MADIT-CRT	1,366	16.3	31.2	$1-(16.3/31.2)^2$	**0.73**	(0.70–0.75)
REVERSE	487	23.4	29.5	$1-(23.4/29.5)^2$	**0.37**	(0.30–0.44)
RETHINQ†	142	5.1	5.61	$1-(5.1/5.61)^2$	**0.17**	(0.07–0.29)
RESPOND	55	44.1	46.9	$1-(44.1/46.9)^2$	**0.12**	(0.008–0.31)
Variance weighted average					***0.50***	
ΔLVEDV						
MIRACLE-ICD II	154	62.0	76.0	$1-(62/76)^2$	**0.33**	(0.21–0.45)
CARE-HF	735	50.7	74.5	$1-(50.7/74.5)^2$	**0.54**	(0.49–0.59)
MADIT-CRT	1,366	14.4	33.2	$1-(14.4/33.2)^2$	**0.81**	(0.79–0.83)
REVERSE	487	28.0	33.4	$1-(28/33.4)^2$	**0.3**	(0.23–0.37)
RETHINQ†	142	7.14	5.36	$1-(7.14/5.36)^2$	**−0.77**	
RESPOND	55	47.6	51.9	$1-(47.6/51.9)^2$	**0.16**	(0.02–0.36)
Variance weighted average					***0.40***	
ΔLVEF						
CONTAK-CD	490	10.3	10.4	$1-(10.3/10.4)^2$	**0.02**	(0.002–0.05)
MIRACLE-ICD II	153	6.2	8.0	$1-(6.2/8)^2$	**0.4**	(0.28–0.52)
CARE-HF	735	4.5	8.5	$1-(4.5/8.5)^2$	**0.72**	(0.68–0.75)
MADIT-CRT	1,366	3.0	5.0	$1-(3/5)^2$	**0.64**	(0.61–0.67)
REVERSE	487	6.5	9.3	$1-(6.5/9.3)^2$	**0.51**	(0.45–0.57)
RETHINQ†	142	0.99	1.22	$1-(0.99/1.22)^2$	**0.34**	(0.21–0.47)
RESPOND	55	11.7	15.7	$1-(11.7/15.7)^2$	**0.44**	(0.23–0.62)
Variance weighted average					***0.48***	

*The maximal possible R^2 for any parameter to the ΔEF in each study is presented for each study. †The ΔLVESV, ΔLVEDV, and ΔLVEF values in RETHINQ have distinctive properties making them unsuitable for this analysis. First, the test–retest reproducibility (Δ in controls) had a variance a magnitude smaller than the other studies. Second, the distributions of Δ for these echo markers were skewed with a curtailment of the tail trending toward deterioration, and extension of the tail trending toward improvement. This skew was not present in continuous outcome markers in which only a single measurement was possible, such as exercise duration (Online Appendix reference 55). The RETHINQ protocol blinded assessors to allocation but did not report blinding echocardiographers prior to measurements. It also permitted exclusion of extreme values, as per clinical convention. The published data suggest this latitude was exercised asymmetrically, favoring an appearance of improvement, which is common in clinical practice because the psychological impact of a deterioration in individual patients is important. However, from a pure research viewpoint, this skew of Δs makes the data unsuitable for evaluating the scientifically desired blinded test–retest reproducibility of variables (41–43). Boldface represents the calculated R^2 from each study.

MIRACLE-ICD II = Multicenter InSync ICD Randomized Clinical Evaluation II; RESPOND = Resynchronization in Patients with Heart Failure and a Normal QRS Duration; RETHINQ = Resynchronization Therapy in Normal QRS study; SD = standard deviation; other abbreviations as in Table 1.

random noise—the lower the limit on the sustainably observable R^2 (20,22).

This approach could be used to screen the plausibility of any claim of a baseline marker apparently predicting the effect of any intervention on any variable.

The ceiling to observed R^2. Spontaneous variability in both dyssynchrony markers and response markers conspire to limit observed R^2 to only low values. This happens with any etiology of heart failure, and for both single-variable and multivariate risk indexes. HSSCSs consistently report higher R^2 values than EM-RCTs, and frequently exceed the mathematical ceiling. This suggests that the HSSCS method is in error.

This study exposes the prerequisites for reliable prediction of individual response, which are challenging. Even a theoretical, perfectly comprehensive dyssynchrony marker (whether a single-variable or multivariate algorithm) that incorporates every facet of responsiveness to biventricular pacing will still have a low ceiling. This is because it will have spontaneous variability, as will the marker of response, and these 2 contraction factors multiply to limit R^2.

These findings are important because some studies forget the impact of variability within predictors and response markers, and so have unrealistic expectations. Our study shows that many published

Table 3. R^2 **Contraction Factor for Nonechocardiographic Markers of Response to Biventricular Pacing**

		Breakdown of Variability		Mandatory Ceiling on R^2 Value Imposed Solely by Unpredictable Variability in Response Measure		
					Maximal Achievable R^2 Value	
Response Measure/Study	N	Unpredictable Element of Variability: SDΔ of Control Arm	Total Variability in Intervention Arm: SDΔ of Intervention Arm	Calculation	Point Estimate R^2	95% CI
6MWD (m)						
COMPANION	592	93.0	96.0	$1 - (93/96)^2$	**0.06**	(0.03–0.10)
CONTAK-CD	444	103.8	104.8	$1 - (103.8/104.8)^2$	**0.02**	(0.002–0.05)
MIRACLE	171	98.0	109.0	$1 - (98/109)^2$	**0.19**	(0.09–0.30)
Peak VO_2 (ml/kg/min)						
CONTAK-CD	417	4.3	4.4	$1 - (4.3/4.4)^2$	**0.05**	(0.02–0.10)
MIRACLE	303	3.2	3.2	$1 - (3.2/3.2)^2$	**0.00**	
QoL score						
CONTAK-CD	459	2.0	2.0	$1 - (2/2)^2$	**0.00**	
MIRACLE	403	21.7	25.1	$1 - (21.7/25.1)^2$	**0.03**	(0.004–0.06)
%ΔQoL						
COMPANION	753	23.0	26.0	$1 - (23/26)^2$	**0.22**	(0.17–0.27)
V_E/V_{CO2}						
MIRACLE	144	5.2	6.2	$1 - (5.2/6.2)^2$	**0.30**	(0.18–0.43)

*The maximal achievable R^2 column estimates the maximum R^2 that any predictor could find when correlated against any of these non-echocardiographic outcome measures. Boldface represents the calculated R^2 from each study.

6MWD = 6-minute walk distance; COMPANION = Comparison of Medical, Pacing, and Defibrillation Therapies in Heart Failure Trial; SDD = standard deviation of difference (Δ); VO_2 = oxygen consumption; V_E/V_{CO2} = ratio of minute ventilation (V_E) and minute production of CO_2 (V_{CO2}), a measure of ventilatory response to exercise; other abbreviations as in Table 1.

studies markedly overestimate the predictive effect of dyssynchrony markers compared with studies having formal enrollment and blinded analysis.

This should not be mistaken as a criticism of workers' integrity, but rather a failing in all of us in underestimating the importance of aspects of study design that might appear superficially uninteresting or trivial. The time-honored approach of hypothesizing correlations and then finding confirmatory evidence in one's local clinical data is incorrect because it provides results that are not only too high but actually above the mathematical ceiling. Strong prediction is sustainable only if both contraction factors are almost 1. Not only are they nowhere near 1, but comparatively little effort has been put into establishing what they are.

Can the contraction factor be improved? Only genuinely reducing variability in both predictors and response markers can improve the contraction factor. Formal blinded test–retest reproducibility ("other day, other hands, other eyes") of the markers must be carried out. Methods should then be refined or rejected, and the cycle iterated, until a protocol is obtained that reliably delivers high reproducibility in independent, blinded hands. It is the measurement protocol, and not the operators, that is being tested. If wide test–retest variability is observed, then that is the result of the study design and it should be reported dispassionately. Operators should not be blamed for reporting the truth. Some planners mistake published data on remeasurement for test–retest reproducibility. Others collect it too late to change the study protocol. Worse still, some collect it only under pressure from journal reviewers after study completion, at which stage there is overwhelming pressure to report a narrow variability even if unrepresentative.

Only markers with strong test–retest reproducibility should even be considered for expensive trials of individualized prediction (Table 5). Unless very much narrower than the population distribution of these variables, the markers should be rejected or refined before initiating any major study. Effort expended on maximizing the ratio of signal (between-patient genuine variability) to noise (within-patient variability) is indispensable to improving prediction of response.

Opportunities and limits of replicate averaging. Increasing patient recruitment will not raise the ceiling on sustainably observable R^2; instead it enforces the same mathematical ceiling more firmly by reducing scope for fluke associations.

The impact of noise can, however, be reduced by making multiple replicate measures per patient and using the average (19,21). However, signal itself is

Table 4. **Test–Retest Variability of Dyssynchrony Markers Within Individuals, in Populations Who Are Candidates for Biventricular Pacemaker Implantation***

		Mandatory Ceiling on R^2 Arising From Variability in Dyssynchrony Marker			
Dyssynchrony Marker/Study	**Within-Patient SD**	**Between-Patient SD**	**Calculation**	**Limit on R^2**	**R^2 95% CI**
Interventricular Mechanical Delay					
Pulsed flow Doppler					
Burri et al. 2007	Not reported	26	Incalculable		
Duncan et al. 2006	Not reported	22	Incalculable		
De Boeck et al. 2008	Not reported	25	Incalculable		
Bordachar et al. 2010	Not reported	29	Incalculable		
Intraventricular Mechanical Delay					
Pulsed tissue Doppler					
Onset of systolic motion - 2 segment					
Burri et al. 2007	Not reported	20	Incalculable		
Onset of systolic motion - 4 segment					
Penicka et al. 2004	Not reported	37	Incalculable		
Onset of systolic motion - (S-L)					
Bleeker et al. 2007	Not reported	49	Incalculable		
Peak of systolic motion - 2 segment					
Burri et al. 2007	Not reported	34	Incalculable		
Onset of systolic motion					
Penicka et al. 2004	Not reported	37	Incalculable		
Burri et al. 2007	Not reported	29	Incalculable		
Peak of systolic motion					
Burri et al. 2007	Not reported	33	Incalculable		
Peak of systolic motion - 4 basal segments					
Bax 2004	Not reported	49	Incalculable		
Peak of systolic motion - 12 segments					
Yu et al. 2003	Not reported	17	Incalculable		
Onset of systolic motion - 2 basal segments (S-L)					
Soliman et al. 2007	Not reported	48	Incalculable		
Peak of systolic motion - 2 basal segments (S-L)					
Soliman et al. 2007	Not reported	44	Incalculable		
Onset of systolic motion - 2 basal segments (AL-IS)					
Palmieri et al. 2010†	32	42	$1 -(32/42)^2$	**= 0.42**	(0.13–0.68)
Onset of systolic motion - 2 basal segments (A-I)					
Palmieri et al. 2010†	22	27	$1 - (22/27)^2$	**= 0.34**	(0.07–0.62)
Onset of ejection					
Foley et al. 2011	29	76	$1 - (29/76)^2$	**= 0.85**	(0.70–0.93)
Color tissue Doppler					
Peak of systolic motion - 2 basal segments (S-L)					
Van Bommel et al. 2010	Not reported	51	Incalculable		
Peak of systolic motion - 2 basal segments (S-L)					
Conca et al. 2009	Not reported	50	Incalculable		
Yu et al. 2007	Not reported	54	Incalculable		
Faletra et al. 2009	Not reported	75	Incalculable		
Peak of systolic motion - 2 opposite segments					
Yu et al. 2007	Not reported	32	Incalculable		
Peak of systolic motion - 2 segment					
Shanks et al. 2010	Not reported	42	Incalculable		
Conca et al. 2009	Not reported	50	Incalculable		
De Boeck et al. 2008	Not reported	57	Incalculable		
Veseley et al. 2008	Not reported	not reported	—	**0.35**	(0.01–0.72)

Continued on next page

Table 4. Continued

		Mandatory Ceiling on R^2 Arising From Variability in Dyssynchrony Marker			
Dyssynchrony Marker/Study	Within-Patient SD	Between-Patient SD	Calculation	Limit on R^2	R^2 95% CI
Peak of systolic motion - 4 segment					
Veseley et al. 2008	Not reported	Not reported	—	**0.46**	(0.06–0.78)
Peak of systolic motion - 6 segment					
Notabartolo et al. 2004	Not reported	137	Incalculable		
Conca et al. 2009	Not reported	41	Incalculable		
Peak of systolic motion - 12 segment					
De Boeck et al. 2008	Not reported	16	Incalculable		
Deplagne et al. 2009	Not reported	9	Incalculable		
Yu et al. 2007	Not reported	39	Incalculable		
Peak of systolic motion - 12 segments SD					
Conca et al. 2009	Not reported	15	Incalculable		
Yu et al. 2004	Not reported	13	Incalculable		
Faletra et al. 2009	Not reported	19	Incalculable		
Yu et al. 2007	Not reported	15	Incalculable		
Van de Veire et al. 2007	Not reported	16	Incalculable		
Peak of systolic motion - 2 basal segments (AL-IS)					
Palmieri et al. 2010†	23	81	$1 - (23/81)^2$	**= 0.92**	(0.83–0.96)
Peak of systolic motion - 2 basal segments (A-I)					
Palmieri et al. 2010†	29	105	$1 - (29/105)^2$	**= 0.92**	(0.83–0.96)
Peak of systolic motion - 12 segments SD					
Palmieri et al. 2010†	10	22	$1 - (10/22)^2$	**= 0.79**	(0.60–0.90)
Peak of diastolic motion - 2 segment					
Shanks et al. 2010	Not reported	49	Incalculable		
Tissue synchronization imaging					
Gorcsan et al. 2004	Not reported	158	Incalculable		
Conca et al. 2009	Not reported	15	Incalculable		
Faletra et al. 2009	Not reported	35	Incalculable		
M-Mode					
Septal posterior wall motion delay (SPWMD)					
Bleeker et al. 2007	Not reported	119	Incalculable		
Pitzalis et al. 2002	Not reported	92	Incalculable		
Pitzalis et al. 2005	Not reported	96	Incalculable		
Diaz-Infante et al. 2007	Not reported	113	Incalculable		
Sassone et al. 2007	Not reported	46	Incalculable		
Foley et al. 2011	91.7	99.8	$1 - (91.7/99.8)^2$	**= 0.16**	(0.00–0.45)
Lateral wall postsystolic displacement (LWPSD)					
Sassone et al. 2007	Not reported	24	Incalculable		
3D systolic dyssynchrony index					
Marsan et al. 2008	Not reported	0	Incalculable		
Faletra et al. 2009	Not reported	3	Incalculable		
Conca et al. 2009	Not reported	1	Incalculable		
Deplagne et al. 2009	Not reported	0	Incalculable		
Liodakis et al. 2010	Not reported	0	Incalculable		
Soliman et al. 2009	Not reported	0	Incalculable		
Transverse strain					
Time to peak transverse strain - 12 segments SD					
Pouleur et al. 2011	Not reported	63	Incalculable		
Longitudinal strain					

Continued on next page

Table 4. Continued

		Mandatory Ceiling on R^2 Arising From Variability in Dyssynchrony Marker			
Dyssynchrony Marker/Study	**Within-Patient SD**	**Between-Patient SD**	**Calculation**	**Limit on R^2**	**R^2 95% CI**
Time to peak strain - 12 segments SD					
Lim et al. 2011	Not reported	35	Incalculable		
Strain delay index - 16 segments					
Lim et al. 2011	Not reported	12	Incalculable		
LV pre-ejection period					
Duncan et al. 2006	Not reported	16	Incalculable		
De Boeck et al. 2008	Not reported	26	Incalculable		
Bordeachar et al. 2010	Not reported	41	Incalculable		
RV pre-ejection period					
Duncan et al. 2006	Not reported	10	Incalculable		
Combined Interventricular and Intraventricular Dyssynchrony					
Pulsed tissue Doppler					
Penicka et al. 2004	Not reported	65	Incalculable		

*Most studies did not report test–retest reproducibility. Where test–retest reproducibility is available, the R^2 contraction factor applied by variability in the dyssynchrony marker can be found. †In the Palmieri study, the data for SD within 1 patient are from their Table 3, and the SD across the population is from their Table 1. Boldface represents the calculated R^2 from each study. Please see the Online Appendix for reference citations.

not increased by replication and so sustainably observable R^2 can rise as high as the underlying R^2 but no higher.

When performing replicates for averaging, researchers must avoid the natural temptation to choose replicates that appear similar, ignoring apparent outliers. Measurements are best made without reference to each other, to maximize the statistical advantages of replicate averaging (22,23).

Researchers should also resist the clinically natural temptation to choose the replicate most representative of the patients' full clinical status, because doing so may innocently but powerfully bias the result toward confirming whatever association the researcher believes (24). Clinical practice is heavily dependent on such application of common sense, but unfortunately this is why routinely acquired unblinded clinical measurements are unsuitable for testing whether clinically believed associations are true. The stronger the belief, and the wider the range of values available (22,24), the greater the danger of self-deception.

In practice, replicates are often made a few seconds apart, but this does not capture variability over hours and days, or sensitivity to imperceptible differences in probe position (23). Paradoxically, an average of replicates would be more consistent over time if the measurements aggregated within

Table 5. The True Limit to the Observed R^2 for the Correlation Between ΔLVEF and Dyssynchrony Markers Is a Product of Their Two R^2 Contraction Factors*

R^2 Contraction Factor Imposed by Variability in Response Marker	**R^2 Contraction Factor Imposed by Variability in Dyssynchrony Marker**								
	0.1	0.2	0.3	0.4	0.5	**0.6**	0.7	0.8	0.9
0.1	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.2	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18
0.3	0.03	0.06	0.09	0.12	0.15	0.18	0.21	0.24	0.27
0.4	0.04	0.08	0.12	0.16	0.20	**0.24**	0.28	0.32	0.36
0.5	0.05	0.10	0.15	0.20	0.25	**0.30**	0.35	0.40	0.45
0.6	0.06	0.12	0.18	0.24	0.30	0.36	0.42	0.48	0.54
0.7	0.07	0.14	0.21	0.28	0.35	0.42	0.49	0.56	0.63
0.8	0.08	0.16	0.24	0.32	0.40	0.48	0.56	0.64	0.72
0.9	0.09	0.18	0.27	0.36	0.45	0.54	0.63	0.72	0.81
1.0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90

*The potential R^2 contraction factor arising from the variability in dyssynchrony markers and response markers is shown. The resulting ceiling on observable R^2 values is modest, with the values in **bold** being the most likely.

Table 6. Differences in Study Design Between EM-RCTs and HSSCSs*

Study	EF/LVESV/LVEDV Measurements Stated to Be Blinded	Dyssynchrony Measurement Stated to Be Blinded	Study Measurements Made Only After Formal Enrollment
EM-RCTs			
CONTAK (Marcus 2005)	**Yes**	**Yes**	**Yes**
MIRACLE (Sutton 2003)	**Yes**	**Yes**	**Yes**
CARE-HF (Ghio 2009)	**Yes**	**Yes**	**Yes**
REVERSE (Linde 2008)	**Yes**	**Yes**	**Yes**
MADIT-CRT (Solomon 2010)	**Yes**	**Yes**	**Yes**
MADIT-CRT (Pouleur 2011)	**Yes**	**Yes**	**Yes**
RETHINQ (Beshai 2007)	No	No	**Yes**
RESPOND (Foley 2011)	**Yes**	**Yes**	**Yes**
Proportion (%)	**86**	**86**	**100**
HSSCSs			
Pitzalis et al. 2002	No	No	**Yes**
Bax et al. 2003	No	**Yes**	Unknown
Yu et al. 2003	No	No	**Yes**
Gorcsan et al. 2004	No	No	**Yes**
Notabartolo et al. 2004	No	No	**Yes**
Penicka et al. 2004	**Yes**	**Yes**	Unknown
Yu et al. 2004	No	No	Unknown
Pitzalis et al. 2005	No	No	**Yes**
Yu et al. 2005	No	No	**Yes**
Mele et al. 2006	No	No	Unknown
Porciani et al. 2006	No	No	Unknown
Suffoletto et al. 2006	No	No	**Yes**
Yu et al. 2006	No	No	**Yes**
Gorcsan et al. 2007	No	No	**Yes**
Soliman et al. 2007	No	No	Unknown
Yu et al. 2007	No	No	**Yes**
Delgado et al. 2008	No	No	Unknown
Jansen et al. 2008	No	No	**Yes**
Lim et al. 2008	No	No	**Yes**
Marsan et al. 2008	No	No	**Yes**
Van de Veire et al. 2007	No	No	Unknown
Bank et al. 2009	No	No	**Yes**
Deplagne et al. 2009	**Yes**	**Yes**	Unknown
Soliman et al. 2009	No	No	**Yes**
Park et al. 2010	No	No	**Yes**
Bordachar et al. 2010	No	No	Unknown
Kaufmann et al. 2010	No	No	Unknown
Miyazaki et al. 2010	**Yes**	**Yes**	**Yes**
Norisada et al. 2010	No	No	Unknown
Van Bommel et al. 2010	No	No	Unknown
Wang et al. 2010	No	No	Unknown
Shanks et al. 2010	No	**Yes**	Unknown
Lim et al. 2011	No	**Yes**	**Yes**
Proportion (%)	**9**	**18**	**≥50**

*Studies assessing the relationship between echocardiographic dyssynchrony markers and outcome markers were assessed for design features that resist bias. Specific statements regarding the blinding of measurements (both baseline dyssynchrony markers and outcome markers) were sought. Statements regarding formal enrollment of patients into a specific study prior to measurements being made were also sought. Please see the Online Appendix for reference citations.

Abbreviations as in Tables 1 and 2.

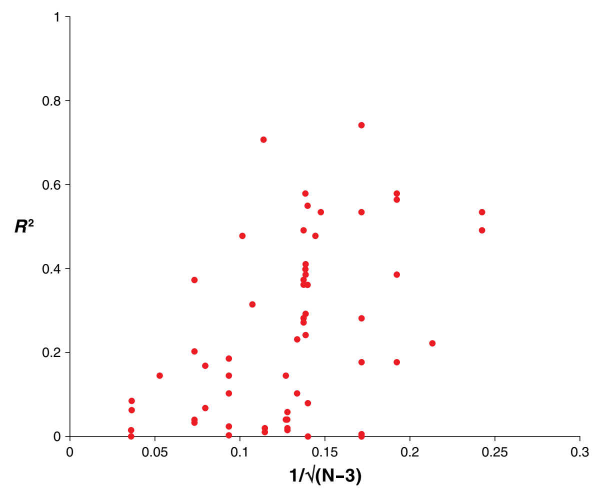

Figure 5. Evidence of Publication Bias

Funnel plot showing R^2 against reciprocal of uncertainty of R^2, calculated as $1/\sqrt{(N - 3)}$, where N is the sample size. Smaller studies show larger R^2 values (r^2 for this relationship $= 0.25$; $p < 0.01$) indicating the evidence of publication bias favoring positive studies. The combined size-and-design plot in Figure 5 permits comparison of the relative impact of study design and sample size.

each average were done on separate days (i.e., their internal variability had been maximized) rather than on successive beats (23). Averaging can only reduce the influence of variability the full spread of which is captured among the data points averaged.

The final advantage of replicate measurement, if "conducted on another day, acquired by other hands, viewed by other eyes," is that it exposes irreproducible markers for early dismissal.

Why some HSSCSs report higher R^2 (and higher than mathematically sustainable limits) than do EM-RCTs. Several factors may have contributed to HSSCSs' reporting significantly higher R^2 values than the sustainable values found in EM-RCTs (Fig. 4).

CHANCE ASSOCIATION. High R^2 values may have been found by statistical chance and then published with preferential enthusiasm. This could occur as submission bias from research groups and/or acceptance bias from journals.

'RUSSIAN-DOLL' PUBLICATION. Successive HSSCS publications from the same site may have overlapping patient cohorts. Patients might understandably be added to a growing database from which publications naturally arise. High R^2 occurring by chance in early cohorts would repeatedly contribute to subsequent publications.

PREFERENTIAL RECRUITMENT OF PATIENTS. Selection of extra patients who have unusually severe or mild mechanical dyssynchrony, or who have unusually large changes in the response variable, will artificially magnify R^2.

LACK OF BLINDING. The R^2 between mechanical dyssynchrony markers and response markers can reliably inform prospective clinical practice only if each measurement is performed by observers blinded to the other relevant measurements in that patient. Mechanical dyssynchrony should be measured without knowledge of the LVEF, and vice versa. Dyssynchrony markers are sensitive to adjustment of cursor position, and operators might inadvertently "dial in" (25) the expected dyssynchrony if unblinded. Ventricular function assessment is similarly sensitive to choices during acquisition and during analysis. Clinicians are generally right to preferentially select plausible rather than implausible values. Unfortunately, applying this habit in research is dangerous, because if the clinician already believes the hypothesis, even minor and innocent influence will raise R^2 dramatically (24). Concealment of electrocardiography (which shows biventricular pacing spikes) is essential during analysis if unbiased ΔLVEF is sought. The majority of the EM-RCTs report blinding of dyssynchrony and response measurements (Table 6); almost all of the HSSCSs did not.

SELECTIVE INCLUSION OR EXCLUSION OF PARTICULAR PATIENTS. HSSCSs may receive unusual referral patterns distorting the distribution of dyssynchrony markers away from the pattern typically seen by future clinical practice and in EM-RCTs. Finally, HSSCSs, if done without the advantage of formal, sequentially numbered, prospective enrollment of patients, may end up unintentionally analyzing an incomplete subset of the population at that center (24); patients with notably strong concordance between physiological expectation and clinical response are especially unlikely to be forgotten, and their preferential recollection would persistently bias R^2 upward.

Why do inflated reports gain circulation, and how can recurrences be prevented? It is tempting to blame publication bias (i.e., failure to publish negative studies). However, the standard funnel plot (Fig. 5) shows that publication bias is a minority contributor. The overwhelming determinant is the vulnerability of the study design to bias, as shown on the combined design-and-size plot (Fig. 6).

The responsibility may lie more properly with us as an audience for several weaknesses in application of normal scientific critique.

SOUND-BITE SUSCEPTIBILITY. Our community accepted uncritically the term *cardiac resynchronization therapy*. With repetition it became obvious that quantifying mechanical dyssynchrony (which can only refer to ventricular timings because atrium and ventricle should not be synchronous) must quantify degree of benefit. Obvious, but not necessarily true. With experimental investigation of the therapy's mechanism of action still at an early stage, we might reduce cognitive distortion by using a neutral term such as *biventricular pacing* (26).

AUTHORITY AWE. Physical science audiences judge a scientific finding by the precise nature of the experiment, the attention to detail, and the track record of previous claims being verified by others. Cardiologic audiences may not apply the same level of scrutiny (in particular, bias resistance is rarely debated) and may apply the availability heuristic (judging the credibility of sources from public visibility rather than track record of reliability). Audiences could usefully restore habits from their earlier scientific training.

ANNUAL AMNESIA. Hearing each year of novel predictive markers with progressively more excellent predictive capacities, cardiologic audiences forget to ask what happened to markers of years past. If 2 different markers predict excellently, they must agree almost perfectly; if the latter is not the case, the former is not credible. Enhanced audience memory would help resist successive overstatements.

PRACTICAL PARALYSIS. In physical science, any reported efficacious new approach is rapidly tested in small experiments by the audience. Cardiologic audiences may feel unable to do this. Yet simple experiments taking only minutes can quickly reject some claims. One example is the evaluation of blinded test–retest reproducibility and the application of the formulas in this paper. Another is adjustment of interventricular delay across a wide range in a single biventricular pacemaker patient, with blinded measurement of mechanical dyssynchrony; if this does not show a clear minimum in this highly controlled environment, it cannot work across a population (27,28).

WISHFUL THINKING. We all want our specialty of echocardiography to be relevant. Reports of successful application are therefore intrinsically popular. But this is failure to separate our individual skill as echocardiographers from the ability of an echocardiographic technique to deliver what is claimed. Distinguishing falsity of a hypothesis from personal inadequacy requires courage but is the hallmark of science.

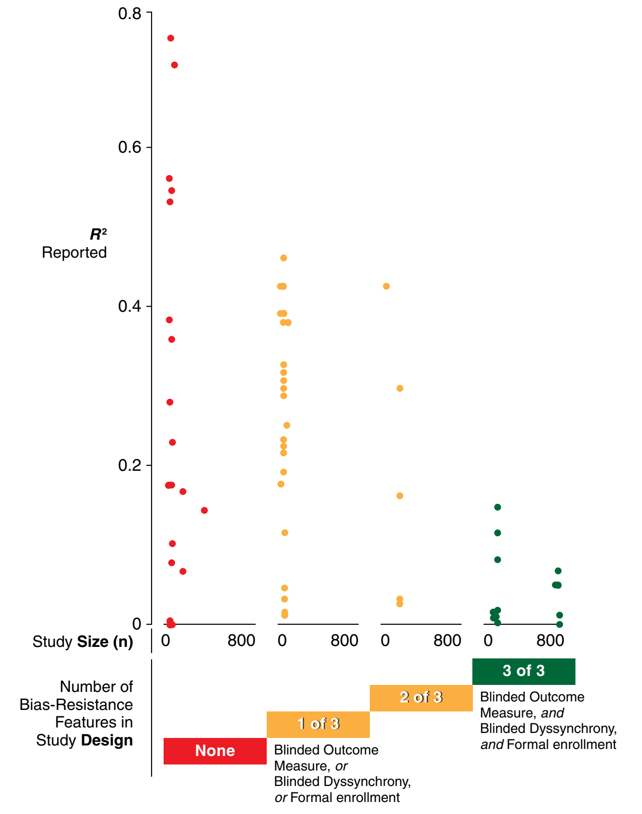

Figure 6. Size-and-Design Funnel Plot Showing Impact of Study Size and Bias Resistance of Study Design on R^2 Reported by the Study

The size and study design of studies affect the reported size of the R^2 between baseline mechanical dyssynchrony markers and change in echocardiographic response markers (e.g., ΔLVEF, ΔLVESV, and ΔLVEDV). Three study design features that resist bias were sought: 1) statements of blinded assessment of the outcome measure; 2) blinded assessment of baseline mechanical dyssynchrony; and 3) formal enrolment of patient into specific study. Studies with no bias-resistance features **(None, red)** report a wide spectrum of R^2 values. Studies designed with multiple bias-resistance features **(3 of 3, green)**—especially if large—report R^2 values close to 0. Abbreviations as in Figure 1.

CRYPTIC COMMENTARY. Even when experts carefully review available methods and tabulate that dyssynchrony markers are intensely vulnerable to noise and sometimes choice of measurement location so that there is risk for "dialing in" any desired level of dyssynchrony (25), they may be too polite to explain the quantitative implication for claims of response prediction (29).

BIAS BLINDNESS. We frequently confuse bias (which arises from study design) with chance (which is addressed by p values). Larger study sizes cause even minor systematic biases to become *more* statistically significant. This confusion afflicts even expert audiences, such as those writing guidelines, who often consider observational studies (if *large*) to be the same level of evidence ("B") as a randomized controlled trial. Yet ironically, larger observational study size *increases* susceptibility to bias, making them *less* reliable guides to therapeutic decisions.

TEST-RETEST TABOO. We frequently confuse remeasurement of identical digital images with genuine test–retest reproducibility, entirely ignoring the majority of variability, which occurs between beats. Test–retest variability can be readily checked by clinicians (30).

RETRACTION RELUCTANCE. In science, a finding beyond the bounds of plausibility, such as faster-than-light travel (31), is highlighted as suspicious by the authors even with a p value $\sim 2 \times 10^{-9}$. The error was found to be unrecognized measurement bias, and the scientific record corrected (32). Our field could encourage timely retraction of clinical reports that are discovered to be unrepresentative, giving credit to authors who report what went wrong in their own studies.

Clinical implications for mechanical dyssynchrony. This paper provides a simple method for clinicians to calculate the ceiling on plausible claims of predictability of ventricular response. It may seem surprising that variability in repeated measures can matter so much because such variability does not seem to impede normal clinical practice or trials addressing group mean effects. However, even small variability inevitably prevents accurate *individualized* prediction of response.

This approach may seem somewhat mathematical. However, publications stating an R^2 or correlation are making a mathematical assertion of association strength. This paper shows how the same mathematics that underlie R^2 calculations also demarcate the upper limit of plausibility, which is far below 1.0, exposing some assertions as anomalous.

Some may suspect that if ischemic scar or imperfect left ventricular lead positioning could be excluded, mechanical dyssynchrony markers might provide good prediction. These factors make prediction more difficult, that is, the ceiling is even lower than we describe here. Our calculations show that even if scarring, mispositioning, and all other confounders could be eliminated, the highest sustainable R^2 value would still be low.

Nor can multivariable prediction by composite markers evade these difficulties. Spontaneous variability in response markers remains and may be worse for composites of poorly reproducible components. Thus composite markers will likely have an even lower ceiling on R^2.

The search for predictive markers stems from a desire to optimize the resource cost of biventricular pacing. However, resources expended in identifying predictors unreliably would be better expended first screening predictors and response markers for blinded test–retest reproducibility. Early exposure of poor reproducibility would forestall reports of prediction that are destined not to stand the test of time.

Implications for research into mechanical dyssynchrony. Our analysis rejects not the concept of mechanical dyssynchrony, but rather the value of unblinded, informally enrolled studies of prediction (Fig. 5). Outcomes are known to be better in those device recipients who have greater dyssynchrony but observational study design cannot distinguish whether the better outcome would have happened without the device, or occurred as a result of the device. The continuous-variable analysis of the CARE-HF (Cardiac Resynchronization in Heart Failure) randomized controlled trial shows both mechanisms occurring simultaneously (33). Nondevice patients had progressively better outcomes the more dyssynchrony they had at baseline (33).

Approaches that have worked for addressing *group mean* effects cannot be uncritically expected to determine which *individual* patients benefit the most from biventricular pacing. This would need measurements of the effect of biventricular pacing on *individual* patients that have narrow within-individual error bars. Symptoms or outcomes assessed in the conventional way are not suitable, but quantitative physiological measurements could be developed to deliver this (33–35).

Wider implications for cardiologic research. Study design can overwhelmingly determine study results. In this example, a "perfect storm" of excellent plausibility—a clear survival benefit from the devices, clinical enthusiasm, unnoticed poor test–retest variability, underestimated impact of unblinded measurement, and lack of community awareness of ceilings on predictability—has made the literature as a whole unreliable.

But similar overstatements may be occurring elsewhere, unnoticed. Our cardiologic community should improve its ability to challenge claims. We must recognize that study size alone does not guarantee reliability (36); bias must be actively removed by careful planning. We should be emboldened to question early pioneering work because history shows it is often discredited later (37–39). Negative studies may seem superficially unexciting to journals, but carefully designed studies replicable by readers, contradicting prevailing beliefs, are the lifeblood of genuine science. We should prize not extreme claims but reliable experiments that can be checked by readers. When results are incompatible, groups should collaborate to understand why.

Few real-life outcomes are overwhelmingly determined by 1 variable; almost always, multiple features (including those that cannot currently be quantified) matter and interact in a way that cannot be captured in a single diagnostic marker. Therefore, we should react with surprise if a single variable is reported to predict any outcome with high certainty.

Finally, this recent "bubble market" in mechanical dyssynchrony should inoculate our community with skepticism for claims of association strength and encourage the examination of the track record of whether previous claims have stood the test of time.

Study limitations. This study is limited by the haste with which reports of prediction of biventricular pacing response arrive in the literature before independent blinded evaluations of their test–retest reproducibility of their methods. No study seems to have performed a series of blinded measurements of dyssynchrony to permit true reproducibility SD to be evaluated. We have had to use the SD of difference between just two, which is an imperfect estimate of this.

Furthermore, few reports of claimed prediction of response present sufficient information to determine the distribution of the change induced by biventricular pacing, or test–retest reproducibility of either predictors or outcome measures. We have used rigorously performed EM-RCTs, which have control populations to assess the inherent variability of response markers over time periods over which response to device implantation is normally measured. However, these control populations are only similar (through randomization) to their corresponding intervention populations, rather than being identical. Therefore, the estimates of the R^2 contraction factor may slightly under- or overestimate the true R^2 contraction factor. Individual studies may give an estimate <0, especially if the true contraction factor is near zero and/or the statistical characteristics of the Δ are non-normally distributed, as might occur when an analysis is unblinded.

These studies had different types of patients with differing etiologies, echocardiographic criteria, and outcome measures. This would limit the strength of a conventional meta-analysis, but it increases the generalizability of the findings from our study. No marker was strongly predictive when tested in bias-resistant designs, despite covering many spectra of patient populations. There was no relationship of the R^2 values to the proportion of patients who had ischemic etiology of heart failure.

Many different mechanical dyssynchrony markers have been assessed, some basic and some sophisticated. They are not all directly comparable. For openness and completeness, we fully report all of the markers for which R^2 ceiling was calculable (Table 1). Most have only modest values, including the newer, more sophisticated ones. Markers involving multiple steps to measure have more sources of variation and are likely to have a worse R^2 contraction factor.

Some studies recently report not R^2 but area under curve or a sensitivity analysis. The underlying variability remains, and an equivalent of the R^2 contraction factor affects all measures of association.

New studies have chosen to measure alternative methods of echocardiographic response, such as global strain. We specifically chose to review the commonly used echocardiographic response markers (LVEF, LVESV, and LVEDV). This is because, first, these markers are commonly measured by echocardiographic laboratories. Second, they are the most familiar to clinicians. Third, it has been accepted that improvement in these is related to outcome (40). Finally, they have been tested in formally enrolled EM-RCT settings allowing reliable assessment of spontaneous variability. Our findings, however, are applicable to all future predictors and response markers. Each must have low test–retest variability for the contraction factor to be sufficient for any chance of clinically impressive prediction of response.

CONCLUSIONS

No scientifically conducted study will give reliable individual patient prediction (e.g., $R^2 > 0.5$), by any current baseline marker of mechanical dyssynchrony, of any current marker of response to biventricular pacing, across a representative range of patients, with current measurement protocols. Unsustainably high R^2 values arise through honest efforts in HSSCSs not because the studies are small but because they unknowingly have inherently unreliable designs.

It may be time to critically reassess the utility of HSSCS literature on the prediction of ventricular response to mechanical dyssynchrony. The overstatement of relationship strength by 5- to 20-fold, and the large proportion of results exceeding the mathematically possible ceiling, indicate that unblinded studies without formal enrollment have failed us entirely.

It is not credible to attempt to usefully predict ventricular response in *individual* patients, or to embark on further research into such predictive power, unless 2 substantial methodologic advances arrive: 1) protocols for measuring mechanical dyssynchrony that reliably give high test–retest reproducibility in the hands of multiple centers beyond their originators, under formal blinded conditions; and 2) methods for quantifying "response" in a way that, in patients who undergo no intervention, shows minimal *within-individual* change over time, in blinded externally monitored analysis, over time periods similar to those over which biventricular pacing response is typically measured.

The latter may be biologically impossible, in which case reliable individualized prediction of response is impossible.

To prevent future bubble markets of ineffective diagnostics, we should not take seriously any more unblinded, unenrolled studies that make mathematically implausible claims.

Reprint requests and correspondence: Dr. Sukhjinder S. Nijjer, Department of Cardiology, Hammersmith Hospital, Du Cane Road, London W12 0HS, United Kingdom. *E-mail: s.nijjer@imperial.ac.uk*.

REFERENCES

1. Cazeau S, Leclercq C, Lavergne T, et al., for the Multisite Stimulation in Cardiomyopathies (MUSTIC) Study Investigators. Effects of multisite biventricular pacing in patients with heart failure and intraventricular conduction delay. N Engl J Med 2001;344:873–80.
2. Abraham WT, Fisher WG, Smith AL, et al., for the MIRACLE study group. Cardiac resynchronization in chronic heart failure. N Engl J Med 2002;346:1845–53.
3. Gras D, Leclercq C, Tang AS, Bucknall C, Luttikhuis HO, Kirstein-Pedersen A. Cardiac resynchronization therapy in advanced heart failure: the multicenter InSync clinical study. Eur J Heart Fail 2002;4:311–20.
4. Young JB, Abraham WT, Smith AL, et al., for the Multicenter InSync ICD Randomized Clinical Evaluation (MIRACLE ICD) Trial Investigators. Combined cardiac resynchronization and implantable cardioversion defibrillation in advanced chronic heart failure: the MIRACLE ICD Trial. JAMA 2003;289:2685–94.
5. Bristow MR, Saxon LA, Boehmer J, et al., for the Comparison of Medical Therapy, Pacing, and Defibrillation in Heart Failure (COMPANION) Investigators. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. N Engl J Med 2004;350:2140–50.
6. Cleland JG, Daubert JC, Erdmann E, et al., for the Cardiac Resynchronization-Heart Failure (CARE-HF) Study Investigators. The effect of cardiac resynchronization on morbidity and mortality in heart failure. N Engl J Med 2005;352:1539–49.
7. Bax JJ, Marwick TH, Molhoek SG, et al. Left ventricular dyssynchrony predicts benefit of cardiac resynchronization therapy in patients with end-stage heart failure before pacemaker implantation. Am J Cardiol 2003;92:1238–40.
8. Yu CM, Fung JW, Zhang Q, et al. Tissue Doppler imaging is superior to strain rate imaging and postsystolic shortening on the prediction of reverse remodeling in both ischemic and nonischemic heart failure after cardiac resynchronization therapy. Circulation 2004;110:66–73.
9. Bleeker GB, Mollema SA, Holman ER, et al. Left ventricular resynchronization is mandatory for response to cardiac resynchronization therapy: analysis in patients with echocardiographic evidence of left ventricular dyssynchrony at baseline. Circulation 2007;116:1440–8.
10. Marcus GM, Rose E, Viloria EM, et al., for the VENTAK CHF/CONTAK-CD Biventricular Pacing Study Investigators. Septal to posterior wall motion delay fails to predict reverse remodeling or clinical improvement in patients undergoing cardiac resynchronization therapy. J Am Coll Cardiol 2005;46:2208–14.
11. Chung ES, Leon AR, Tavazzi L, Sun JP, et al. Results of the predictors of response to CRT (PROSPECT) trial. Circulation 2008;117:2608–16.
12. Miyazaki C, Redfield MM, Powell BD, et al. Dyssynchrony indices to predict response to cardiac resynchronization therapy: a comprehensive prospective single-center study. Circ Heart Fail 2010;3:565–73.
13. Hawkins NM, Petrie MC, Burgess MI, McMurray JJ. Selecting patients for cardiac resynchronization therapy: the fallacy of echocardiographic dyssynchrony. J Am Coll Cardiol 2009;53:1944–59.
14. National Institute for Health and Clinical Excellence. Cardiac resynchronisation therapy for the treatment of heart failure. TA120. London, UK: National Institute for Health and Clinical Excellence; 2007.
15. Palmieri V, Russo C, Buonomo A, et al. Test-re-test reproducibility of Doppler echocardiography for assessment of electromechanical dyssynchrony: implications for heart failure clinic. J Cardiol 2010;56:271–9.
16. Vesely MR, Li S, Kop WJ, et al. Test-retest reliability of assessment for intraventricular dyssynchrony by tissue Doppler imaging echocardiography. Am J Cardiol 2008;101:645–50.
17. Gustafson P. Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments. Boca Raton, FL: Chapman and Hall/CRC Press; 2003.
18. St John Sutton MG, Plappert T, Abraham WT, et al., for the Multicenter InSync Randomized Clinical Evaluation (MIRACLE) Study Group. Effect of cardiac resynchronization therapy on left ventricular size and function in chronic heart failure. Circulation 2003;107:1985–90.

19. Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. BMJ 2010;340:c2289.
20. Francis DP, Coats AJ, Gibson DG. How high can a correlation coefficient be? Effects of limited reproducibility of common cardiological measures. Int J Cardiol 1999;69:185–9.
21. Pabari PA, Willson K, Stegemann B, et al. 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. Heart Fail Rev 2011;16:277–90.
22. 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.
23. 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 2012 [E-pub ahead of print].
24. 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 2012 [E-pub ahead of print].
25. Abraham TP, Dimaano VL, Liang HY. Role of tissue Doppler and strain echocardiography in current clinical practice. Circulation 2007;116:2597–609.
26. Kyriacou A, Pabari PA, Francis DP. Cardiac resynchronization therapy is certainly cardiac therapy, but how much resynchronization and how much atrioventricular delay optimization? Heart Fail Rev 2011 [E-pub ahead of print].
27. Pabari PA, Kyriacou A, Moraldo M, et al. CRT optimisation: improving echocardiographic techniques by accommodating biological variability within different echocardiographic parameters. Heart 2011;97:A66–67.
28. Lumens J, Leenders GE, Cramer MJ, et al. Mechanistic evaluation of echocardiographic dyssynchrony indices: patient data combined with multiscale computer simulations. Circ Cardiovasc Imaging 2012;5:491–9.
29. MacRoberts MH, MacRoberts BR. The negational reference: or the art of dissembling. Soc Stud Sci 1984;14:91–4.
30. Finegold JA, Manisty CH, Cecaro F, Sutaria N, Mayet J, Francis DP. Choosing between velocity-time-integral ratio and peak velocity ratio for calculation of the dimensionless index (or aortic valve area) in serial follow-up of aortic stenosis. Int J Cardiol 2012 [E-pub ahead of print].
31. The Opera Collaboration. Measurement of the neutrino velocity with the OPERA detector in the CNGS beam. Available at: http://arxiv.org/abs/1109.4897. Accessed June 17, 2012.
32. Reich ES. Embattled neutrino project leaders step down. Available at: http://www.nature.com/news/embattled-neutrino-project-leaders-step-down-1.10371. Accessed June 17, 2012.
33. Richardson M, Freemantle N, Calvert MJ, Cleland JG, Tavazzi L, for the CARE-HF Study Steering Committee and Investigators. Predictors and treatment response with cardiac resynchronization therapy in patients with heart failure characterized by dyssynchrony: a pre-defined analysis from the CARE-HF trial. Eur Heart J 2007;28:1827–34.
34. Stegemann B, Francis DP. AV/VV delay optimization and response quantification in biventricular pacing (CRT): arrival of reliable clinical algorithms and research protocols, and how to distinguish them from unreliable counterparts. Europace 2012. In press.
35. Bogaard MD, Meine M, Tuineburg AE, Maskara B, Loh P, Doevendans PA. Cardiac resynchronization therapy beyond nominal settings: who needs individual programming of the atrioventricular and interventricular delay? Europace 2012 [E-pub ahead of print].
36. Parolari A, Tremoli E, Cavallotti L, et al. Do statins improve outcomes and delay the progression of non-rheumatic calcific aortic stenosis? Heart 2011;97:523–9.
37. Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2:e124.
38. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. JAMA 2005;294:218–28.
39. Hudes ML, McCann JC, Ames BN. Unusual clustering of coefficients of variation in published articles from a medical biochemistry department in India. FASEB J 2009;23:689–703.
40. Yu CM, Bleeker GB, Fung JW, et al. Left ventricular reverse remodeling but not clinical improvement predicts long-term survival after cardiac resynchronization therapy. Circulation 2005;112:1580–6.
41. Fisher RA, Tippett LHC. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society 1928;24:180–90.
42. Clark CE. The greatest of a finite set of random variables. Operations Research 1961;9:145–162.
43. Gumbel EJ. Statistical theory of extreme values and some practical applications: a series of lectures. Volume 33 of Applied mathematics series, United States. National Bureau of Standards. U.S. Government Printing Office, 1954.

Key Words: biventricular pacing ■ dyssynchrony ■ echocardiography ■ ejection fraction ■ statistics.

▶APPENDIX

For appendix, please see the online version of this article.