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

Judith A. Finegold *, Charlotte H. Manisty, Fabrizio Cecaro, Nilesh Sutaria, Jamil Mayet, Darrel P. Francis

International Centre for Circulatory Health, National Heart and Lung Institute, London, UK

ARTICLE INFO

Article history:
Received 15 October 2011
Received in revised form 10 April 2012
Accepted 14 April 2012
Available online xxxx

Keywords:
Aortic stenosis
Doppler
Reproducibility
Echocardiography

ABSTRACT

Background: It remains unclear which echocardiographic measure is most suitable for serial measurement in real-world aortic stenosis (AS) follow-up. We determine whether the dimensionless index (DI) between aortic valve and left ventricular outflow tract velocities is measured more consistently using velocity-time-integral (VTI) or peak velocities (Vpeak) in real life.

Methods: Serial echocardiograms acquired within 6 months in subjects with AS were analysed with blinding, to compare the variability over time of DI calculated using Vpeak, with that of DI calculated using VTI.

Results: Paired echocardiograms, acquired on average 72 days apart, were analysed from 70 patients with a range of severities of AS (50% severe). DI, calculated using either Vpeak or VTI, did not significantly change over this short time. Coefficient of variation was significantly better when DI was calculated using Vpeak than VTI (12.6 versus 25.4%, p < 0.0001). The variabilities of mean and peak trans-aortic valve 4V² and left ventricular outflow tract VTI were no better: 26.9%, 19.1% and 22.1% respectively.

Conclusions: Serially-followed variables require minimal noise to maximise detection of genuine change. For AS surveillance, calculating DI – or effective orifice area – from the ratio of Vpeak rather than VTIs would reduce 95% confidence intervals from ±51% to a still-disappointing ±25%. Guidelines recommend noisy surveillance measures, causing conscientious echocardiographers to ‘peak’ at previous values, and impairing clinicians’ faith in echocardiographically-observed changes when making clinical decisions. For us in echocardiography to improve our ability to contribute to AS follow-up requires us to first acknowledge and discuss this honestly.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Since the pivotal Ross and Braunwald report [1] that across multiple post-mortem studies of aortic stenosis, mortality was 100% and almost all had symptoms in the final years of life, clinicians have relied on symptoms as the primary guide to appropriate timing of valve replacement. Despite modern technological advances in equipment and measurement, echocardiography is still subordinate in status to clinical judgement during follow-up decision making.

For echocardiography to contribute meaningfully to follow-up requires exquisite test–retest reproducibility: narrow error bars within individuals. Wide error bars cause three harms. First, patients may falsely appear to have deteriorated. Second, true deteriorations may not be reflected in measurements and hence remain undetected. Third, clinicians may reject even substantial detected deteriorations due to a lack of confidence in the measures (defeating the purpose of the scan), or schedule excessively frequent visits in an effort to reduce the influence of measurement variability (draining resources).

As echocardiographers, to enhance our reliance to aortic stenosis follow-up, we should start by actively selecting the index we report for serially assessing disease severity based on the narrowness of its within-individual error-bars. In real life clinical practice, patients may have scans by different operators who may not only acquire images slightly differently, but also make measurements (including tracing Doppler envelopes) slightly differently. In the UK, the majority of echocardiography is performed in centres which use more than one type of scanner and conduct scans under relentless time pressure. In contrast, most of the published studies of reproducibility in aortic stenosis were conducted within a research environment with a single operator acquiring images who had unlimited time to concentrate on maximising the measurement reproducibility.

Guidelines recommend measuring AS severity using several haemodynamic features. AS jet peak velocity (Vpeak), peak instantaneous pressure drop 4V² peak and mean 4V² form one family of variables [2]. To counter confounding from changes in stroke volume, a second family of variables uses the ratio of velocities between the left ventricular outflow tract and the aortic valve: the ‘dimensionless

Abbreviations: AS, aortic stenosis; Vpeak, AS jet peak velocity; DI, dimensionless index; LVOT, left ventricular outflow tract; VTI, velocity-time-integral.

* Corresponding author at: Office of Dr Francis, International Centre for Circulatory Health, 59 North Wharf Road, National Heart and Lung Institute, London, UK. Tel.: +44 207 594 1093; fax: +44 208 082 5109.
E-mail address: JudyFinegold@doctors.org.uk (J.A. Finegold).

0167-5273/$ – see front matter © 2012 Elsevier Ireland Ltd. All rights reserved.
index’ (DI). This DI is often made into an Effective Orifice Area by multiplying it by the cross sectional area of the left ventricular outflow tract (LVOT): the continuity equation [3,4].

How should we measure DI during aortic stenosis follow-up to deliver narrow within-patient error-bars? The two options are the ratio of velocity-time-integral (VTI), or ratio of peak velocities (Vpeak), between LVOT and aortic valve. This study evaluates the test–retest reproducibility of measuring DI using VTI or Vpeak in aortic stenosis, using data from real-world clinical practice.

2. Methods

2.1. Subjects

Clinical echocardiographic data from consecutive patients with aortic stenosis who underwent serial echocardiography with 2 scans acquired within 6 months of each other between November 2007 and July 2011 in our hospital were retrospectively reviewed.

In total, 548 patients with aortic stenosis were identified who had undergone echocardiography in this time period, of whom 70 patients had had repeat echocardiograms within a 6 month timeframe. The severity of the aortic stenosis was defined in the subjects at their first echocardiogram: 5 (7%) had mild stenosis (DI > 2), 24 (34%) had moderate stenosis (DI 2–4) and 41 (59%) had severe stenosis (DI > 4). The average age at first visit was 79 years (range 53–92 years). Sixty-three percent were female. The time between repeat echocardiograms averaged 72 (standard deviation 59) days.

2.2. Difference in echocardiographic Doppler measurements of aortic stenosis between two scans

Echocardiography was conducted according to standard clinical guidelines [4] in the manner conventional in our hospital and most other hospitals. The separate visits were often conducted by different operators, and using machines made by different manufacturers. Images had been acquired and stored in digital format. For this study they were analysed offline by a single observer (JF), with measurements of each study blinded to the results of the paired study.

Standard Doppler measurements of flow in the left ventricular outflow tract (LVOT) and aortic valve were recorded from multiple windows to calculate the peak velocity across the valve and the velocity time integral (VTI) as recommended by guidelines [5,6]. We then calculated the DI using the ratio of measurements for aortic and LVOT, using both peak velocities (Vpeak) and VTI flow data (Fig. 1). The mean trans-aortic pressure drop was calculated automatically by the echocardiography machines, by averaging the instantaneous gradient over the period of flow. The peak instantaneous trans-aortic pressure drop was calculated as:

\[
\text{Peak instantaneous trans-aortic pressure drop} = 4 \left( \frac{V_{\text{peak, aortic valve}}}{C_{\text{16,17}}} \right)^2
\]

2.3. Statistics

The original intention had been to subtract the change in mean between the two visits, so as to expose the variability distinct from what we expected might be a group trend towards deterioration. In practice because there was no significant change in mean in this short period of time, this subtraction step was not required.

Changes in the average values of the DI and the mean and peak trans-aortic valve pressure drops between visit 1 and visit 2 were assessed using the paired Student t test. The relative proportions showing progression versus regression of these variables between visits were tested for difference from chance alone using Fisher’s exact test.

Variability was calculated using the coefficient of variation (standard deviation of differences divided by the mean). The spread of variability between visit 1 and visit 2 was compared for the paired echocardiograms using the F test. The results were also assessed using the Bland–Altman method [7].

A p value of < 0.05 was considered significant. Statistical analysis was performed using Prism software (version 5.0).

All authors confirm that the study was designed to make measurements without bias, to be held responsible for procedural deficiency, and to retract the paper if any are suspected. Patient data were selected only by the method described. Measurements were made blinded and uniformly. No data were deleted, nor re-measured to favour one result over another [8].

3. Results

3.1. Progression of aortic stenosis during the inter–test interval

Across the group of patients as a whole there was no statistically significant evidence of progression of aortic stenosis during the

![Fig. 1. Calculation of the dimensionless index. Calculation of the dimensionless index using either (left panel) the ratio of the left ventricular outflow tract aortic valve velocity time integral or (right panel) the ratio of the peak velocities.](image-url)
using instead $V_{\text{peak}}$ showed that 41.4% of patients had a lower DI on their second visit, and 58.6% a higher DI ($p = 0.40$ by Fisher’s exact test) as shown in Table 2 and Fig. 2.

### 3.2. Effective test–retest reproducibility in the dimensionless index calculated using VTI and peak velocities

The coefficient of variation was significantly better when calculating the DI using $V_{\text{peak}}$ (12.6%) than VTI (25.4%); $p < 0.0001$ using the F test, as shown in Fig. 3. Comparisons between DI measured using VTI and $V_{\text{peak}}$ are illustrated as Bland–Altman plots in Fig. 4.

### 3.3. Effective test–retest reproducibility in mean and peak aortic valve pressure drop and LVOT VTI

The mean and peak trans-aortic valve pressure drops and the LVOT VTI were calculated at visit 1 and visit 2. Over the timeframe measured (~72 days), there were no significant changes in any of these additional measures of aortic stenosis severity (Table 1). Comparison of mean trans-aortic valve pressure drops during the inter-test interval showed that 44.4% of patients had a lower pressure drop on their second visit than their first, and 55.6% had a higher pressure drop ($p = 0.70$ by Fisher’s exact test). Comparison of peak trans-aortic valve pressure drops during the inter-test interval showed that 44.3% of patients had a lower pressure drop on their second visit than their first and 55.7% had a higher pressure drop ($p = 0.61$ by Fisher’s exact test). Similarly, analysis of the serial changes in LVOT VTI found that 52.9% of patients had a lower value on their second visit than their first and 47.1% had a higher value ($p = 1.0$ by Fisher’s exact test) as shown in Table 2.

The coefficient of variation was smaller for peak instantaneous trans-aortic valve pressure drop (19.1%) than for the mean trans-aortic valve pressure drop (26.9%). The coefficient of variation was also high for LVOT VTI (22.1%), as shown in Fig. 3, $p < 0.0001$ for all comparisons.

### 4. Discussion

In order for changes in valve severity to be detected and acted upon in a timely fashion in aortic stenosis, echocardiographers and clinicians require robust and reproducible markers, which they trust to accurately reflect genuine physiological changes. This real-world clinical study shows that narrow within-patient error bars, which are essential for markers used in serial follow-up, are more achievable when DI is calculated using a ratio of peak velocities than when calculated using velocity time integral ratios. Although this study does not achieve the narrow within-patient error bars achieved by Otto et al. [23], it does support in real-world routine practice the Otto et al. recommendation for $V_{\text{peak}}$ over VTI for the serial monitoring of aortic stenosis severity.

Dimensionless index by peak velocity should not be considered to be a shortcut to the “real” DI, but rather the preferred marker for evaluating DI in the serial follow-up of aortic stenosis. Serial measurements of mean and peak trans-aortic valve pressure drops and the LVOT gradient VTI have wider within-patient error bars than those of DI by peak velocity. This makes it understandable that clinicians

---

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th>Proportion of patients appearing to show an improvement at visit 2</th>
<th>Proportion of patients appearing to show deterioration at visit 2</th>
<th>$p$ Value (Fisher’s test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensionless index by VTI</td>
<td>48.6%</td>
<td>51.4%</td>
<td>1.00</td>
</tr>
<tr>
<td>Dimensionless index by $V_{\text{peak}}$</td>
<td>41.4%</td>
<td>58.6%</td>
<td>0.40</td>
</tr>
<tr>
<td>Mean trans-aortic pressure drop (mm Hg)</td>
<td>44.4%</td>
<td>55.6%</td>
<td>0.70</td>
</tr>
<tr>
<td>Peak instantaneous trans-aortic pressure drop (mm Hg)</td>
<td>44.3%</td>
<td>55.7%</td>
<td>0.61</td>
</tr>
<tr>
<td>LVOT VTI (cm)</td>
<td>52.9%</td>
<td>47.1%</td>
<td>1.00</td>
</tr>
</tbody>
</table>

* All comparisons $p < 0.0001

**Fig. 3.** Comparison of the coefficient of variation for dimensionless index calculated using peak velocities, with the other echocardiographic parameters measured in the serial follow-up of patients with aortic stenosis. When the dimensionless index was calculated using peak LVOT and aortic velocities, the coefficient of variation was significantly smaller ($p < 0.0001$) than for all other echocardiographic parameters assessed.
may be sceptical of basing clinical decisions solely on small changes in any of these single markers, but rather use clinical symptoms despite their lack of quantitative resolution.

4.1. Within-patient error bars masquerading as progression — or as its mirror image

In aortic stenosis, where spontaneous regression is physiologically implausible, the proportion of patients whose measurements of stenosis appear to improve from one visit to the next is an index of real-life variability in the measurement. In our cohort, when using VTI or V_peak to calculate the dimensionless index, DI was just as likely to appear to decrease as it was to appear to increase. If we had assumed that all the numerical changes observed were genuine changes in biological status, then we would be obliged to conclude that the rate of change in the valve ranged from a deterioration of 3.3 DI units to an improvement of 3.5 DI units within this ~72 day period. Previous reports that discuss progression of AS quote only average changes in echocardiography measurements but do not analyse the direction of change: the writers may have under-recognised variability in echocardiography measurements — although clinical sonographers in day-to-day clinical practice rarely do. Our study recognises these apparent declines in severity as evidence of spontaneous variability, which (since the nature of measurement scatter is to be symmetrical unless there is some reason for bias) must be causing just as many false-positive deteriorations as it is false appearances of improvement. Since they occurred in a typical cohort of patients being scanned in an experienced and accredited laboratory, and numerically analysed by one operator, it is not plausible that routine clinical practice in similar patients in a general hospital will show substantially narrower reproducibility.

Fig. 5 illustrates this further. Our results in Fig. 2 resemble panel 4 of Fig. 5. The balanced numbers, and extent, of apparent improvement and apparent deterioration, are evidence of the underlying change being very small during this ~72 day period.

Cardiac catheterization is not used in our hospital for serial evaluation of aortic stenosis. Whilst it is almost always carried out prior to valve surgery for detection of coronary artery disease, most angiographers, including those authoring this paper (CHM, JM, DPF), do not measure invasive transvalvular pressure drop in the expectation of matching echocardiographic 4V_peak, for several reasons. First, validity of echocardiographic measurements of aortic stenosis with invasive data is already published to be so intense [9–13] that — given the mathematical limits on it arising from error bars in each [14] — the underlying r² required to explain this already exceeds 1.0. Second, the single arterial puncture used in our catheter laboratory permits measurement of non-simultaneous peak-to-peak pressure drops which are systematically smaller than echocardiographically-calculated peak instantaneous drops. Third, each reintroduction of the catheter from aorta to left ventricle incurs risk [15] and so it is not clinically attractive to carry out the multiple measurements required to obtain error bars narrow enough to give additional diagnostic information above echocardiography.

4.2. Need for test–retest variability studies in a real-world environment

When designing a test for serial follow-up of a condition a key criterion is minimal test–retest variability. In this study we have found higher variability when calculating the DI using VTI than when using V_peak. We propose that the reason for this is measurement error arising from the requirement to trace around the Doppler valve curves to find the VTI. In early and late systole the Doppler tracing of the aortic valve and the LVOT has a steep incline. If during tracing there is either over- or underestimation in this steep area of the curve, it will lead to large measurement error. The peak velocity, in contrast, is generally more easily determined precisely.

Using the same principle of quantifying test–retest variability, we compared values between visits for the mean and peak trans-aortic valve pressure drop. The mean trans-aortic valve pressure drop is calculated by dividing the integral of the instantaneous gradients (4V²) over the entire period of flow, by the duration of that flow. It performed particularly poorly, with a coefficient of variation of 26.9%, most likely due to the incorporation of a new variable, the duration, which provides an additional opportunity for measurement error.

Table 3 shows the clinical implication of the variability exposed here. When a patient is given a measurement, that value is not perfectly precise but has an uncertainty. For a large group of patients, the uncertainty of the mean of all those patients can be described by a distribution whose standard deviation is the standard error of the mean. However for an individual patient, the uncertainty for that single patient is given by a distribution whose standard deviation is simply the standard deviation of repeated measurements. That distribution, of individual uncertainty, is no narrower for a patient in a very large cohort, than for a patient in a small cohort.

A confidence interval can therefore be given for an individual patient’s value. Because each patient only has two measurements, we cannot reliably state the degree of uncertainty from the data of one patient alone, but instead make the assumption that uncertainty is shared by all patients, and therefore that the standard deviation of between-test differences across all patients can be used as an index of an individual patient’s between-test variability. Values that are higher will tend to have greater absolute variability than values that are lower, and therefore we used coefficient of variation, or percentage variability.

From an example starting value for each measurement of aortic stenosis at a previous visit in column 2, the table shows the range of
real-life measurements that occur without being clinically meaningful. Using variation coefficients calculated in this study, column 4 shows the within-patient 95% confidence intervals as the percentage by which each variable can change between visits, within the range of natural variability. For example, for the DI calculated using VTI the coefficient of variation is 25.4% and so, using a representative value of 3.0 for the DI at the first visit, the percentage by which the variable can change between visits is ±25.4%×1.96=±49.8%. Finally, the range of values possible at the second current visit within natural variability is calculated in column 3 using this 95% confidence interval as 3.0×(1−0.498)=1.51 and 3.0×(1+0.498)=4.49. This illustrates the challenge facing the clinician attempting to use these measurements to guide management: a large spread of apparent changes between successive visits during the serial monitoring of patients arises from spontaneous variability alone.

4.3. Need for honest test–retest reproducibility data and opportunity to use successive routine visits with de-biased, blinded reanalysis

Unfortunately some reports of reproducibility are just re-analysis of the same acquired images by the same (intra-observer) or different (inter-observer) observers who are aware that they are conducting a reproducibility study. This shows only a part of the clinically important biological and measurement variability which is important in the serial follow-up of patients with valve disease. The paucity of reproducibility studies conducted formally on separate days with blinded analysis, may be a result of several factors:

- the expense and administrative burden of conducting this as a formal research study
- the lack of a facade of advanced technical novelty
- fear that it can only make a department appear to be unreliable
- not noticing that existing reports do not provide this information
- believing that it does not affect clinical practice
- fatalism that nothing can be done to improve the situation if test–retest reproducibility is formally discovered to be poor
- fear of harming the credibility of echocardiography.

These can all be resolved, as follows.

Formal trials with third party monitoring of enrolment and blinded externally-conducted measurements do eradicate bias but are expensive. In this study we have shown how data acquired under standard clinical conditions can provide test–retest reproducibility information, if some precautions are taken. First, use a time window short enough to distinguish between the situations in the four lower panels except for the rare conditions in which the noise is much larger than any biological deterioration (panel 4, darker shading) which is recognizable by the distribution of changes being centred on 0. Statistically this would manifest as similar frequencies of apparent improvement and apparent worsening, or a mean change very close to zero.
relentless uncomfortable re-testing of assumptions and a willingness to reverse them in the light of more-representative information.

As for what constitutes reproducibility, ISO 5725 [16], or even brief consideration of where variability might arise, reveals that that re-measurement of an identical image is not relevant.

If operators are making measurements that they believe do not affect clinical practice, they might be guided to question this undesirable situation has arisen. Uncritical recommendation for measurements known by clinicians to have poor reproducibility may be the reason serial echo measurements are increasingly sidelined in follow-up decision making.

If reproducibility is poor, it can always be improved. Either technology can support individual measurements being more representative, or multiple measurements can be averaged to reduce scatter. But both approaches require us to know where the variability arises; denying its existence makes improvement impossible.

4.4. Lack of progression of aortic stenosis during the inter-test interval

Between visits in our study there was no statistically significant change in the dimensionless index across the patient group in this ~72 day time period. Previous studies [17–25], addressing trends over time rather than within-patient error bars, have used longer inter-test intervals and found progression of aortic stenosis severity with an average reduction of aortic valve area of 0.1 cm²/year, or a rise in peak aortic velocity of 0.1–0.4 m/s per year [21,26]. There are factors in addition to the shorter follow-up period that may explain the differences between our data and that of previous studies into the progression of aortic stenosis.

First, most previous reports appeared not to have taken special steps (such as evaluating within-patient error bars) to distinguish between random variation and genuine underlying deterioration, and may instead have assumed that all increases in measured severity were signal and not noise, and all decreases were noise and not signal. Analysing our data that way shows an average “deterioration” of aortic valve velocity by 0.1 DI units/year using Vpeak, and by 0.2 DI units/year using VTI. Only some of the previous studies reported both deterioration and improvement of valve severity during the inter-test period [17,18,21–24], and even then in only one of the studies was this variability addressed [21].

Second, some previous reports arranged for the same echocardiographers and same equipment to be used on successive visits. Although this will improve reproducibility, it may not reflect real-world clinical situations in many institutions. Honestly understanding what can be achieved without that luxury may yet be clinically useful for the typical end-user.

Third, and most importantly, in most studies operators making measurements on the repeat visit were not blinded to the results of the first visit, which may have resulted in an understandable tendency [8] to question and correct any instances that suggested a reduction in severity over time.

4.5. Clinical implications

If the purpose of echocardiographic assessment during serial follow-up of AS is to identify whether there has been a deterioration of the valve, to support the clinical consultation, then current arrangements may be suboptimal. We do not have clear, experimentally-secured guidance on what aspect of serial echocardiographic measurements to place the most weight on. Without reproducible echocardiographic measurements, rational clinicians’ only guide to decide on the timing of aortic valve replacement will be symptoms.

We need guidance regarding which variable we should rely on during serial measurements. Current European guidelines suggest that a change in peak velocity of >0.3 m/s/year in a heavily calcified aortic valve or ‘other evidence of haemodynamic progression’ should prompt surgical intervention [4]. However, the guidelines do not state what level of blinded test–retest reproducibility should be expected by a laboratory before using measurements for these clinical decisions.

Nor is it explained why the time interval is taken into consideration in this way. For example, a 10-year increment of 2.9 m/s is considered unalarming, whilst a 1-week increment of >0.006 m/s – which would occur in approximately half of all such repeat measurements – should apparently prompt valve surgery. In the absence of credible data that the increment between two measurements reliably predicts the next increment, and in the presence of wide reproducibility scatter, why should clinical action today be mandated by the date of the scan before the present one?

Is this why we are encouraged to do more frequent scans in patients who we suspect may be beginning to experience symptoms? To create more opportunities and a greater magnification factor (1/time interval) for chance variation to provide apparent physiological justification for valve surgery?

Moreover, determining valve progression purely on the basis of peak valve velocities risks confounding by increases in stroke volume, which might occur for many reasons, and makes determination of valve severity in the failing left ventricle difficult. Whilst good reproducibility is not sufficient to select a variable for use in follow-up, poor reproducibility is certainly sufficient grounds to deselect it. On this basis, dimensionless index, measured using peak velocities, appears to be currently the most experimentally plausible marker for the serial monitoring of aortic stenosis severity.

The Effective Orifice Area is a commonly-used variable for assessing the severity of aortic stenosis which may seem more practically-applicable for clinicians. The large variability in the measurement [27,28] of the LVOT diameter however means that during serial follow-up any change in area may be dominated by noise. If clinicians desire serial effective orifice area values, to make this a more reliable measurement the operator could calculate the LVOT cross-sectional area on the patient’s first echocardiogram and then use this area as a constant at each follow-up visit to re-calculate the orifice area. Although this is essentially the same as tracking the DI, it presents the results in a format that may be more widely palatable to cardiologists and trainees. It does, however, magnify the dependence on the first measurement of LVOT. An alternative, which maximises the reliability of today’s AVA estimate, would be to calculate the average of all LVOT dimensions over the years, and use it today for AVA and for the re-evaluation of previous effective orifice areas. However some clinicians may find this process, necessitating updating of past effective orifice area measurements, disconcerting.

4.6. Implications for clinical study design

In addition to the importance for guiding the clinical care of individual patients, accurate assessment of the progression of aortic stenosis is also crucial for determining the effects of clinical interventions on the disease as a whole. This can be seen from the studies investigating the effect of statin medication of disease progression in aortic stenosis. Initial studies [29–33] suggested a beneficial effect, however this was contradicted in the subsequent randomised prospective studies and meta-analysis [34–38]. A contributor to this discrepancy may have been that the inherent noise within the measurement variables resulted in operators mistrusting their measurements encouraging them to ‘peek back’ at previous measurements in the unblinded, non-randomised retrospective studies to ensure that they were not reporting implausible reductions in valve lesion severity.

4.7. Study limitations

This study used real-life data with no mandate to keep operator or machine constant between scans, or blind them to prior values. Measurements for this study were made from these scans by one blinded...
observer. In real clinical practice, observers vary and this means the full test–retest variability (another day, other hands, other eyes) may be still higher. Our study was intentionally limited to a short inter-test time period. Over this mean interval of over two months, genuine deterioration was negligible. A special prospective study in which patients consent to make special visits would eliminate all doubt, but might not reflect routine time-pressured clinical practice. Such an exercise might be worth the investment, and inconvenience to patients of additional research visits, once a marker has been identified which is inexpen-

itive, blinded evaluation has smaller test–retest variability than DI by $V_{peak}$.

We have not studied every possible variable that could be used to track aortic stenosis patients over time. For example, E/E' and BNP have been identified as prognostic markers recently in this population [39]. We only set out to address a very limited question of what is the reasonable expectation of reproducibility between tests, for the main recommended markers of aortic stenosis severity.

Further studies are needed, but in rational steps. First, we need to identify markers which have satisfactory blinded test–retest reproducibility in neutral hands under routine clinical conditions; this could be retrospective if done without bias. Second, the effectiveness of these in distinguishing clinically severe aortic stenosis apart from mild or moderate disease can then be compared head-to-head, amongst these test–retest reproducible markers defined before analysis begins; again this might be achievable retrospectively. Finally, a shortlist of very few plausible markers which might undergo a prospective observational study of hard outcomes might be considered. However any marker associated with age or comorbidity will have an advantage in predicting events even if such events are not valve-related; the marker’s real decision making power may therefore be less than its raw prognostic power. Moreover patients with echocardiographically severe AS often undergo special scrutiny for symp-
toms which might trigger valve replacement, which might produce artfactually high predictive power in the markers most trusted by local clinicians.

This is not a randomised trial. A randomised controlled trial, studying differences in mortality achieved by one severity-monitoring variable against another, would have overwhelming impact. But such a trial is unrealistic. For example, to obtain 80% power to detect a difference at the 5% significance level for a hypo-
thesised hazard ratio of 0.90, even if we could follow-up almost all pa-
tients to an event, we would need to randomise >2800 patients. For a hazard ratio of 0.95, we would need ~12,000 patients. Such trials would cost tens of millions of dollars. If clinicians made decisions exclusively using results of endpoint tri-
als, they would be paralysed for all individual decisions that have per-
centage impacts on mortality in single figures. Smaller studies, if done with care to minimise bias, can help us make small but reliable steps on the way to refine our diagnostic and therapeutic armamentaria.

5. Conclusions

Serial monitoring of valvular heart disease such as aortic stenosis ought to use variables with minimal noise, to protect the patient from the inconvenience of unnecessary repeat investigations and the health service from the associated costs. In this study dimensionless index calculated from peak velocities performed best amongst the common echocardiographic markers used for follow-up of aortic stenosis. It is the best-placed amongst these tested echocardiographic variables to be trusted by clinicians as a follow-up tool.

Even still, real-life 95% confidence intervals in our hands for indi-

vidual patients are a disappointing ±25%. If readers’ own local, blinded other day, other hands, other eyes data are much narrower than this, we need to learn from them how to reduce our variability (without peaking). If not, technical developments that improve on this would be particularly welcome — if consistently verifiable in independent hands under clinically-realistic conditions [8]. The habit of trying to arrange the same operator to do both scans and to “peek” at prior values is clinically understandable, although in other settings might be considered cheating. It might be flagged up in guidelines as a temporary response to incomplete technical develop-

ment rather than a long term solution, lest we permanently surrender opportunities to make reliable contributions to aortic stenosis follow-

up. The search for reliable methods to measure routine serial follow-

up values continues, unglamorously.

Funding sources and disclosures

DPF (FS/10/083) was supported by the British Heart Foundation.

Acknowledgements

The authors thank the NIHR Biomedical Research Centre scheme. The authors of this manuscript have certified that they comply with the Principles of Ethical Publishing in the International Journal of Cardiology [40].

References

[3] Bonow RO, Carabello BA, Chatterjee K, et al. ACC/AHA 2006 guidelines for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (writting Committee to Revise the 1998 guidelines for the management of patients with valvular heart disease developed in collaboration with the Society of Cardiovascular Anesthesiologists endorsed by the Society for Cardiovascular Angiog-


[5] Lang RM, Bierig M, Devereux RB, et al. Recommendations for Chamber Quantifica-
tion: a report from the American Society of Echocardiography’s Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr 2005;18:1440–67.
[8] Francis DP. How easily can omission of patients, or selection amongst poorly-

[10] Smith MD, Dawson PL, Eilon JL, et al. Correlation of continuous wave Doppler ve-


lism after retrograde catheterisation of the aortic valve in valvular stenosis: a pro-

[16] ISO/TR 22971. Practical guide to ISO 5725 Part 2: Basic method for the deter-


Shewan LG, Coats AJ. Ethics in the authorship and publishing of scientific articles. Int J Cardiol 2010;144:1–2.