

# A manifesto for cardiovascular imaging: addressing the human factor<sup>†</sup>

Alan G. Fraser<sup>1,2,3\*</sup>

<sup>1</sup>School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; <sup>2</sup>Department of Cardiology, University Hospital of Wales, Heath Park, Cardiff CF14 4XW, UK; and <sup>3</sup>Division of Cardiovascular Imaging and Dynamics, Department of Cardiovascular Sciences, KU Leuven, Herestraat 49, 3000 Leuven, Belgium

Received 15 August 2017; editorial decision 17 August 2017; accepted 11 September 2017; online publish-ahead-of-print 28 September 2017

Our use of modern cardiovascular imaging tools has not kept pace with their technological development. Diagnostic errors are common but seldom investigated systematically. Rather than more impressive pictures, our main goal should be more precise tests of function which we select because their appropriate use has therapeutic implications which in turn have a beneficial impact on morbidity or mortality. We should practise analytical thinking, use checklists to avoid diagnostic pitfalls, and apply strategies that will reduce biases and avoid overdiagnosis. We should develop normative databases, so that we can apply diagnostic algorithms that take account of variations with age and risk factors and that allow us to calculate pre-test probability and report the post-test probability of disease. We should report the imprecision of a test, or its confidence limits, so that reference change values can be considered in daily clinical practice. We should develop decision support tools to improve the quality and interpretation of diagnostic imaging, so that we choose the single best test irrespective of modality. New imaging tools should be evaluated rigorously, so that their diagnostic performance is established before they are widely disseminated; this should be a shared responsibility of manufacturers with clinicians, leading to cost-effective implementation. Trials should evaluate diagnostic strategies against independent reference criteria. We should exploit advances in machine learning to analyse digital data sets and identify those features that best predict prognosis or responses to treatment. Addressing these human factors will reap benefit for patients, while technological advances continue unpredictably.

## Keywords

cardiovascular imaging • evidence-based medicine • metacognition • diagnostic error • clinical guidelines

## Introduction

I was privileged to meet Inge Edler. When we invited him to the first EuroEcho conference, which was held in Prague in 1997, he described how he and his colleagues had paused to think about the clinical problem that faced them (how to select patients for closed mitral valvotomy, without exacerbating mitral regurgitation) and whether they could tackle it using radar or sonar.<sup>1</sup> Through a series of contacts and coincidences, he met the physicist Hellmuth Hertz and they were able to start clinical studies using reflected ultrasound in October 1953. More than 40 years later, I asked Professor Edler if he had ever guessed at the future of the technique that they initiated. He declined to answer, stating that he never

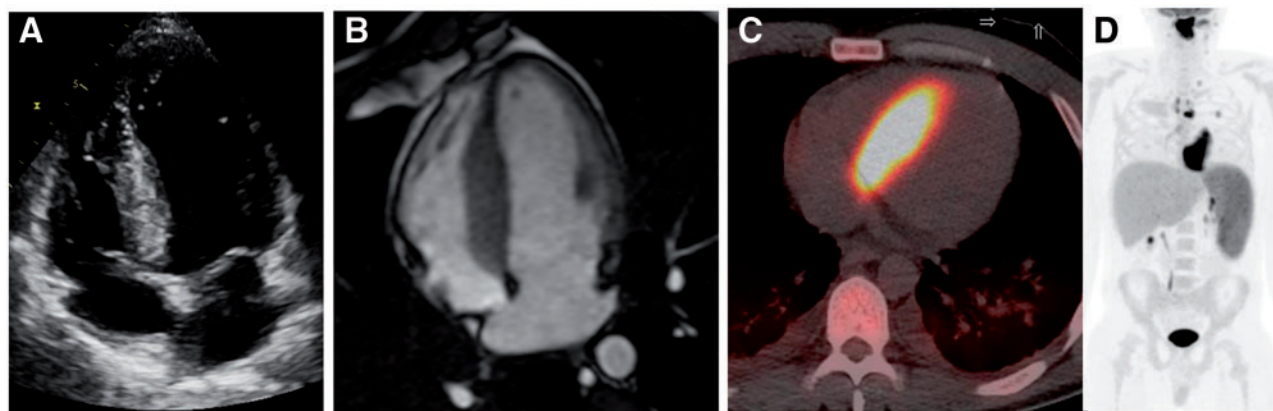
speculated, because it was impossible to predict technological developments.

That is still the situation today. We have a wonderful array of imaging tools that already outstrip our capacity to exploit their potential or apply them appropriately. We should challenge engineers with new objectives, but we can also leave innovative technological advances to take care of themselves. A greater concern is that new methods are sometimes implemented without a clear concept of their clinical role and some become established without ever having been evaluated properly. Arguably, we adopt more complex imaging modalities before we know if they have a positive impact on outcomes, instead of relying on the best or the most efficient test for answering each diagnostic question (*Figure 1*). The biggest challenge

\* Corresponding author. Tel: +44 29 2074 4986; Fax: +44 29 2074 4473. E-mail: fraserag@cf.ac.uk

<sup>†</sup> This manuscript is based on the Edler Lecture delivered at EuroEcho Imaging 2016, the annual conference of the European Association of Cardiovascular Imaging (EACVI), on 7 December 2016.

© The Author 2017. Published by Oxford University Press on behalf of the European Society of Cardiology. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)



**Figure 1** An example of multimodality imaging. Diagnostic images obtained in a 34-year-old man who presented in an electrical storm, having recurrent monomorphic ventricular tachycardia, and with complete heart block. He had previously had chemotherapy for an extranodal natural killer T-cell lymphoma. Transthoracic echocardiography (A, apical four-chamber view) and MR imaging (B) showed focal thickening and impaired function of the mid and basal ventricular septum. Hybrid CT and positron emission tomography (PET) (fused short-axis view, image C) demonstrated avid uptake of fluorodeoxyglucose in the septum. The PET whole-body scan (D) confirmed recurrence of lymphoma in the nasopharynx with high uptake also in the enlarged spleen. The diagnosis was apparent from the first bedside echocardiographic study, and in this patient, the subsequent imaging did not alter the outcome.

for specialists in imaging is how we think about, evaluate, and use the tools that are already available. Each has advantages and limitations, and each depends to some degree on expert interaction to acquire optimal images and to analyse and report them. It is this human factor that now becomes paramount and which I want to address.

A manifesto is a written statement of beliefs and policies. In that context, I offer here my personal perspective of what I think should be our priorities for improving the quality of diagnostic imaging and its contribution to clinical decision-making and outcomes for patients with cardiovascular disease.

## Reducing diagnostic error

The recent classification lists almost 70 000 diagnostic codes,<sup>2</sup> so it is clearly impossible for an individual physician even to be aware of all diseases. Accurate diagnosis is the foundation stone of evidence-based practice, but errors occur in 10–15% of cases.<sup>3</sup> A recent report from the Institute of Medicine in the USA stated that ‘all of us will likely experience a meaningful diagnostic error in our lifetime’.<sup>4</sup> Surprisingly, the commonest explanation in more than two-thirds is not poor knowledge but faulty cognition and synthesis.

Most cardiologists can recall individual patients who were given a wrong diagnosis or whose diagnosis was missed or delayed, but there have been few systematic studies of diagnostic errors in cardiovascular imaging. In 2008, a retrospective analysis of more than 50 000 echocardiographic studies performed in children identified 87 errors (or 0.17%), of which 34% could be attributed to cognitive error and 31% to technical factors.<sup>5</sup> Inaccuracy was more common in rare diseases (odds ratio 9.2) or when studies were performed in the recovery room after cardiac surgery (odds ratio 7.9). In a later analysis, faulty cognition still accounted for 37% of 254 diagnostic errors.<sup>6</sup> In other paediatric studies, 44% of diagnoses made by non-specialists but only 3% made by experts were inaccurate when compared with surgical findings.<sup>7</sup>

In adult practice, a 12% error rate was reported in one study of echocardiography in 170 patients with prosthetic heart valves, a few of whom died after unnecessary reoperation.<sup>8</sup> Grading of mitral regurgitation was inaccurate in 10% of reports made by 152 observers, especially if the quality of images was suboptimal.<sup>9</sup> Among 89 patients diagnosed by general cardiologists after magnetic resonance (MR) imaging to have arrhythmogenic right ventricular cardiomyopathy, 73% did not meet the criteria for the disease when they were reinvestigated in a specialist centre.<sup>10</sup> In a multicentre study of non-invasive coronary angiography using computed tomography (CT), significant luminal narrowing was diagnosed erroneously in 5% of segments when compared with quantitative coronary arteriography (QCA).<sup>11</sup> Cardiovascular diseases such as myocardial infarction, pulmonary embolism, and acute aortic syndromes had been misdiagnosed in 18.7% of 970 autopsies.<sup>12</sup>

Whatever is the true incidence of cardiovascular diagnostic errors—and we need prospective studies to investigate that question—there is consensus that the major cause is faulty thinking. We can learn from psychologists how we might mitigate this. According to Daniel Kahneman, we default to intuitive or ‘System 1’ thinking because of its cognitive ease.<sup>13</sup> It is fast and may be dominant in diagnostic imaging since we learn to recognize visual patterns, relying on mental shortcuts or heuristics,<sup>14</sup> but it is also prone to error. Analytical or ‘System 2’ thinking, on the other hand, is deductive and deliberate; by reviewing options and choosing the most appropriate decision, a consistent diagnosis can be reached that is more scientific and less prone to error.<sup>15</sup> By training ourselves to pause, we can over-ride type 1 in favour of type 2 thinking and avoid jumping to a conclusion that might be wrong.

Diagnostic imaging is also subject to fashion and to cognitive biases.<sup>16</sup> A current example may be left ventricular non-compaction (LVNC), which is a normal stage of myocardial development during foetal life. Three different echocardiographic criteria have been proposed, which show poor concordance.<sup>17</sup> In 199 unselected patients with systolic heart failure, 24% fulfilled at least one criterion and >90%

had apical segments that were hypertrabeculated.<sup>18</sup> In another study that used MR, 5 LV segments were non-compacted in >70% of 45 healthy volunteers.<sup>19</sup> Investigators who proposed original criteria for LVNC detected more cases up to 2006, but almost 50% of the subjects who were diagnosed recently were asymptomatic and none had any clinical events during follow-up.<sup>20</sup> They were probably healthy but overdiagnosed as LVNC because it was suspected, which would be an example of confirmation bias. It is likely that some hypertrabeculation can persist in healthy hearts and that LVNC as a distinct disease is rare. The wide range of mutations that have been associated with it, without a clear answer emerging, would support this.<sup>21</sup>

To err is human, as we were reminded by the Institute of Medicine in 2000, but at that time few professional societies had demonstrated a visible commitment to reducing errors in health care and to improving patient safety.<sup>22</sup> We are all liable to make mistakes and some misdiagnoses may be unavoidable—but ‘normalization of deviance’ is avoidable. Errors can be tolerated as long as we work to minimize their occurrence and consequent risk through individual and collective action at a personal and an institutional level. There are many effective interventions (Table 1).<sup>23,24</sup> We can practise and teach metacognition (thinking about thinking, or specifically in this context, about clinical reasoning), and we can apply strategies that are effective for debiasing.<sup>25,26</sup> We can study and avoid common pitfalls, perhaps using diagnostic checklists.<sup>27</sup> We can recognize common artefacts.<sup>28</sup> We can learn from errors if we report them routinely (in a voluntary and blame-free system), investigate their causes, and share our experience with others. That practice is well established in diagnostic radiology, with discrepancy meetings,<sup>29</sup> but it has not yet been specified in standards from the European Association of Cardiovascular Imaging (EACVI) relating to quality control and audit.<sup>30</sup>

## Selecting the right diagnostic target

A common approach is to consider each diagnostic error as an isolated individual case (or an active error made by one person), but sometimes our collective failure to understand the pathophysiological background to a question or the physical properties of the imaging modality means that we apply faulty concepts and reach a mistaken diagnosis in most cases (which would be an example of a latent or systems error). Often this is related to a morphological feature that seems to have little or no functional significance.

In the 1980s, there was a surge in the diagnosis of mitral valve prolapse soon after the introduction of cross-sectional echocardiography, before it was appreciated that the mitral annulus is not planar in shape.<sup>31</sup> For example, 35% of healthy children aged 10–18 years were reported to have prolapse in apical four-chamber views.<sup>32</sup> These were false-positive diagnoses because that imaging plane crossed the mitral annulus at its low points, and the prevalence fell to 1% if prolapse was required to be visible in other imaging planes too.

Elderly subjects may have prominent LV hypertrophy that is confined to the basal or outlet ventricular septum. This has been named variously as a sigmoid septum or septal bulge or discrete upper septal hypertrophy, and it has been diagnosed as a variant of hypertrophic cardiomyopathy although a causative mutation is rarely found. Now we know that hypertensive LV hypertrophy starts at this site, because it has the greatest regional wall stress<sup>33</sup> and that impaired function is

**Table 1** Priorities for improving diagnostic analysis and reporting

Minimizing diagnostic error
Teach and practise metacognition and analytical thinking
Implement debiasing strategies
Use objective rather than subjective tests
Develop and consult diagnostic checklists
Use normative reference populations to establish normal values
Consider pre-test probability when ordering and reporting a test
Consider reference change values in routine clinical practice
Document and analyse diagnostic errors
Hold regular discrepancy meetings
Implementing smarter information technology
Develop automated analyses
Develop and use decision-support software
Adjust for physiological status and risk factors
Apply cut-points for clinical decisions that have implications for clinical outcomes
Use machine learning to analyse complex data sets
Establish open-access imaging research databases
Develop interactive consensus documents and guidelines

related to high late systolic pressure rather than peak systolic pressure.<sup>34</sup> Upper septal hypertrophy is probably a manifestation of adverse ventricular-arterial coupling. In a definitive report of 3562 subjects in the Framingham Heart Study, upper septal hypertrophy correlated with age and blood pressure, but after adjusting for risk factors it was unrelated to cardiovascular disease or mortality at 15 years.<sup>35</sup>

There is a huge literature on carotid intima-media thickness (cIMT) as a marker of vascular disease, perhaps because it is a very accessible target. In large epidemiological studies, it predicts outcomes, but its added value is small.<sup>36</sup> In the Rotterdam study of 3580 non-diabetic subjects aged 55–75 years, cIMT improved prediction of cardiovascular risk only in women, of whom 8% were reclassified.<sup>37</sup> There are almost no studies in which a reduction or slower rate of progression of cIMT as a surrogate end point has been matched by fewer clinical events.<sup>38,39</sup> It seems that increased cIMT may be a diagnostic epiphenomenon rather than an early stage of disease, unlike atherosclerotic plaque which is more predictive.<sup>40,41</sup> In the RISC study of 627 healthy subjects, increased cIMT appeared to be an adaptive response to ageing, obesity, blood pressure, and low density lipoprotein cholesterol.<sup>42</sup> Recommendations for measuring cIMT in individuals have been taken out of clinical practice guidelines.

Myocardial strain by speckle tracking is popular for quantifying LV long-axis function, because it is quick to perform and more reproducible than alternatives. Global longitudinal strain has been applied to measure responses to exercise<sup>43</sup> but that takes no account of the basic principle that strain is preload dependent.<sup>44,45</sup> Its response is biphasic, increasing at low workloads, and then declining at peak when the end-diastolic volume declines.<sup>46,43</sup> Myocardial ischaemia is better detected by changes in strain rate, because it is not significantly influenced by preload<sup>47,48</sup> but that needs deformation imaging with much faster frame rates. Speckle tracking is also used to study left atrial (LA) function,

ignoring the evidence that changes in LA strain can largely be explained by LV function<sup>49,50</sup> except during atrial contraction.<sup>51</sup>

There could be other examples. The message is that we need to choose the right imaging test to answer each question, by understanding the basic mechanisms of the disease and the technical capacity of the imaging tools, and by applying the most relevant research. David Sackett wrote in 1994 that 'the ultimate criterion for the usefulness of a diagnostic test is whether it adds information that leads to a change in management that is ultimately beneficial to the patient'.<sup>52</sup> According to the 'Grading of Recommendations, Assessment, Development and Evaluation' (GRADE) collaborators, the best way to assess any diagnostic strategy is a trial in which investigators randomize patients to experimental or control diagnostic approaches and measure mortality, morbidity, symptoms, and quality of life.<sup>53</sup> Too often, diagnostic studies are performed to compare one test against another in subjects who are already known to have the disease. Instead, we should evaluate new tests against independent reference criteria and, if possible, assess their impact on clinical decision-making rather than on surrogate end points.

## Defining normality and preventing overdiagnosis

Remarkably, there seems to be no consensus within diagnostic imaging about how we should define normal subjects to establish reference ranges. For instance, should we apply statistical cut points in normative random samples that are representative of the whole population, or should we consider as healthy controls only people who do not have risk factors?

The question has been considered by geneticists who selected 'hypercontrols' with a low pre-test probability of disease in order to minimize the inclusion of false-positive subjects in mapping studies.<sup>54</sup> Applying this approach more generally would be difficult, because, to some extent, we are all abnormal. We share >99% of our genomes, but each of us may have >2800 unique variants in our exomes.<sup>55</sup> Many mutations may be benign or of uncertain significance, but 0.5% of 870 apparently healthy subjects had pathogenic variants for cardiomyopathy or arrhythmia.<sup>55</sup> The more we investigate controls, the less normal they will become.

Genetic factors account for only 1–3% of variance in common measurements such as LV ejection fraction and LV mass, in predominantly Caucasian populations.<sup>56</sup> There are stronger associations of LV long-axis function with ageing, physiological variables such as height and body mass, and risk factors such as hypertension and metabolic syndrome or diabetes.<sup>57–59</sup> These factors are normally distributed, and they progress from subclinical phenotypes to established disease, which leads to what can be called a paradox of precision: the more precise the measurement, the more difficult it becomes to discriminate health from disease.

Selected diagnostic cut points should have a sound statistical basis. Normal LV ejection fraction (EF), for example, has been quoted at >50% but an individual patient meta-analysis of echocardiographic data derived from >22 000 subjects participating in 43 studies quoted the lower limit of normal (fifth centile) of EF for a 30-year-old European man at 49%, whereas for an East Asian woman aged >50 years it was 57%.<sup>60</sup> These values can also be questioned since subjects with EF <30%, blood pressure >140/90

**Table 2** Requirements for evidence-based diagnostic imaging

Establishing diagnostic performance
Measure accuracy <i>in vitro</i> against standard phantoms
Compare software <i>in silico</i> against common digital data sets
Test validity against external, independent reference criteria
Measure <i>inter-observer</i> reproducibility <i>in vivo</i> with good sample size:
• for independent acquisition of images
• for independent analysis of images
• for biological/temporal variability
• in the target clinical population
Perform comparisons with similar and alternative diagnostic approaches
Document feasibility in routine clinical practice
Determine diagnostic utility in populations with varying pre-test probabilities
Opportunities for developing regulatory governance
Transparent reporting of performance (accuracy) against imaging phantoms
Open-access logs of software iterations
Public availability of reproducibility data and reference change values
Industry support for clinical end point studies of diagnostic technologies
Post-market surveillance and registries of diagnostic imaging
Integration of clinical decision support into diagnostic reporting systems

mmHg, serum creatinine >1.5 mg/dL, and other factors were excluded. Using a statistical approach, it would make sense to analyse a large random sample of the whole population including subjects with disease to define cut points corresponding to values more than, say, 2 SDs beyond the mean ( $Z$  score >2 or <-2). An example of this approach was the study of diastolic function using the mitral  $E/A$  ratio in one MONICA cohort.<sup>61</sup> There should be different reference ranges for male and female subjects of all ages and for different ethnic groups.

Ultimately, the logical way to define health is by relating diagnostic measurements to clinical outcomes. A good example comes from the International Database on Ambulatory Blood Pressure in relationship to Cardiovascular Outcomes (IDACO) Investigators. In 1991, data from 23 studies suggested that normal daytime blood pressure (BP; 95% confidence intervals) was 101–146/61–91 mmHg.<sup>62</sup> Later, IDACO used associations between ambulatory BP and 10-year outcomes to define optimal daytime BP as 122/79 mmHg and ambulatory hypertension as >140/85 mmHg; above these thresholds, risk was increased.<sup>63</sup> Another much larger analysis of 958 074 participants aged 40–89 years in 61 studies found increased risk at pressures above 115/75 mmHg.<sup>64</sup> There would be many practical difficulties but no theoretical reasons why we could not undertake similar studies for measurements of cardiac function.

Measurements of function are more important than structure. We know that visual assessment of coronary arteriography is inaccurate and subject to bias but quantitative QCA can also be misleading. In 4086 coronary stenoses observed in 2986 patients, the sensitivity, specificity, and diagnostic accuracy of a diameter stenosis by QCA of  $\geq 50\%$  for predicting fractional flow reserve  $\leq 0.80$  (the functional test and reference criterion) were only 61%, 67%, and 0.64, respectively, and the tests were discordant in 35% of all cases.<sup>65</sup>

We need to guard against unthinking application of inappropriate diagnostic cut-points that do not take account of subclinical changes such as adaptations to regular exercise or that are unrelated to outcomes, as this may cause 'non-disease' which is a serious disservice to patients. Overdiagnosis is more common when sensitive diagnostic tests are used for screening. It can be suspected when the incidence of a disease increases, but its mortality does not change. After CT angiography was introduced in the USA, the incidence of pulmonary embolism increased by 81% while there was minimal change in total mortality and a 36% decrease in case fatality, presumably because more small pulmonary emboli were being detected.<sup>66</sup> In an earlier study, missed pulmonary embolism was the most frequent diagnostic error.<sup>67</sup> It is impossible to know without further research whether treating small pulmonary emboli amounts to overdiagnosis or else might prevent some patients from developing chronic thromboembolic pulmonary hypertension. With the extra treatment, however, there was a 71% increase in complications mostly related to anticoagulation.<sup>66</sup>

Any use of new technology with improved resolution may demonstrate findings that are unusual but harmless, known as incidentalomas. In 1426 research imaging examinations, 39.8% had at least one incidental finding; their discovery was associated with clear medical benefit in only 1% of cases and with a definite burden in 0.5%.<sup>68</sup> CT imaging of the thorax revealed at least one unexpected abnormality in 55% of patients. We need strategies to identify which chance findings merit further investigation and which are unimportant.

## Establishing the 'basic science' of diagnostic imaging

Rigorous pre-market evaluation of any new diagnostic method is essential before it can be adopted widely. This should be comprehensive, as listed in Table 2, including *in vitro* testing against standard imaging phantoms. Methods have been developed for testing software *in silico* against common datasets.<sup>69</sup> Reproducibility should be investigated in clinical studies with good statistical power. These should include independent reacquisition of images and not only (as is commonly the case) independent reanalysis of the same images or digital data.

Reproducibility studies should assess if measurements are stable over a specified time period. The performance of the Agatston score for measuring coronary arterial calcification on CT images was studied in 104 subjects who were reinvestigated after 2 weeks. Inter-scan/inter-observer variability—the appropriate test for extrapolating to usual clinical practice—was  $14.5 \pm 21.8\%$  and  $12.5 \pm 21.8\%$ , compared with intra-scan/intra-observer variability of  $1.7 \pm 5.9\%$  and  $1.3 \pm 3.7\%$ .<sup>70</sup> Positron emission tomography is considered as a reference method for measuring myocardial perfusion, but the coefficient

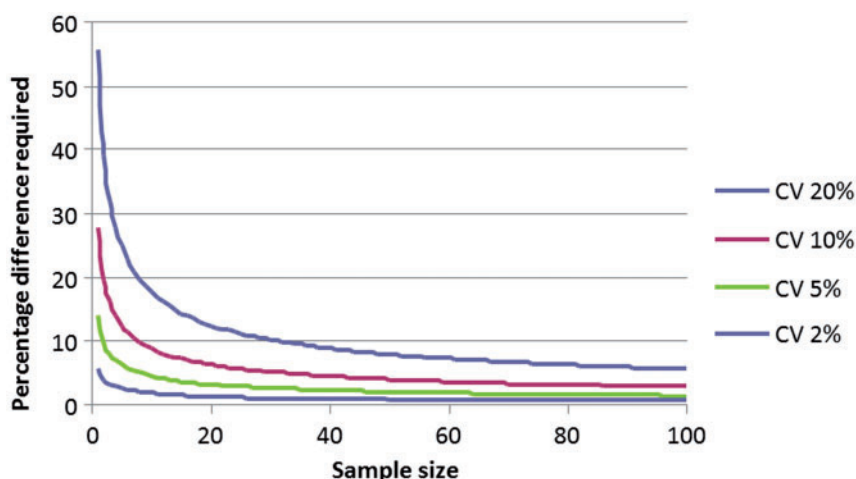
of variation (CV) for serial measurements, which was 10% when scanning was repeated within minutes, was only 21% when the tests were repeated after a median interval of 16 days.<sup>71</sup>

The variability of a measurement should also be assessed in the target population where it is going to be used, rather than in healthy controls. Isovolumic acceleration (IVA), which is a load-insensitive indicator of myocardial contractile function, could be determined in 97% of healthy subjects, with a CV for intra-observer reproducibility of 12%, but the measurement was feasible in only 82% of patients with heart failure in whom the CV for inter-observer reproducibility was 28%.<sup>72</sup> The first result made IVA appear suitable for use in individual patients, whereas the appropriate evaluation showed that it would be better as a research tool.

Another aspect of diagnostic performance that may be overlooked is inter-machine variability. Here, another paradox becomes apparent: the more advanced the imaging modality, the less likely it becomes that diagnostic systems from different manufacturers will give the same results. Each manufacturer processes digital data in its own way, using its own software to measure the same features that are reported by other manufacturers using their software. Substantial inter-vendor differences (with correlations between 0.23 and 0.72) have been reported for global and segmental longitudinal strain measured by echocardiographic speckle tracking<sup>73–75</sup> and for comparisons using software for CT and MR imaging.<sup>76–78</sup> These results indicate that serial studies in individuals should be performed using the same machine and software, if the measurements are to be compared. The initiative of the EACVI and the American Society of Echocardiography (ASE) with all the major vendors, to standardize strain measurements, is an excellent example of how to address this generic problem. Implementing common protocols<sup>79</sup> has halved inter-vendor variability.<sup>80,81</sup>

All results of validation and reproducibility studies and inter-vendor comparisons should be widely available, so that they can be considered in clinical practice. An important concept is the reference change value, which is the difference between consecutive measurements that needs to be exceeded before it can be concluded that an observed difference is a real change rather than a chance variation (Figure 2). A CV of 20% can be acceptable for a research study when each group has  $>50$  subjects, as an inter-group difference of  $>8\%$  will be significant, but for a single patient a CV of 20% means that any difference  $<55\%$  might occur by chance (see left end of each curve in Figure 2). In one individual, reference change values for measurements with coefficients of variation of 10%, 5%, and 2% are  $\pm 28\%$ ,  $\pm 14\%$ , and  $\pm 6\%$ , respectively. Only some imaging tests have CVs  $< 10\%$ , but differences between tests in one patient that are within the range of  $\pm 28\%$  are commonly and erroneously considered as significant. Reference change values have been developed and reported for laboratory assays,<sup>82,83</sup> but they need to become commonplace in quantitative diagnostic imaging. Repeatability limits have been defined for some specific applications of cardiovascular CT.<sup>84,85</sup>

In the 18th century, Thomas Bayes attempted 'to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances'.<sup>86</sup> We are familiar with the principles of a Bayesian approach to diagnosis, and experienced clinicians balance probabilities all the time, but we could do more to make this process more robust. A simple nomogram for applying Bayes' theorem is available (Figure 3).<sup>87</sup> If the pre-test probability and



**Figure 2** The reference change value. Plots demonstrating the differences between consecutive measurements that need to be exceeded before confidently stating that an observed change is likely to be a real effect rather than perhaps the consequence of chance variation related to the imprecision of the measurement. CV, coefficient of variation. Courtesy of Professor Frank Dunstan, Cardiff University.

the likelihood ratio (of true positives to false positives) are known, then it is easy to predict the post-test probability. For example, if the pre-test probability is 5%, which would not be unusual, then for the post-test probability to exceed 50%, the false-positive rate must be below 5%, which is uncommon. Diagnostic tests have limited added value in patients with low or high pre-test probabilities, but that does not stop us from using them even when we may not believe the results. A analysis of dobutamine stress echocardiography in 3259 subjects illustrated how tests are most useful at intermediate pre-test probabilities.<sup>88</sup>

We should understand the technical features of the diagnostic imaging machines that we use, and they should be maintained regularly. Worryingly, 40% of ultrasonic transducers in routine daily use in Swedish hospitals were faulty, with deficiencies such as broken crystals or delaminated cables that meant that they should already have been replaced.<sup>89</sup>

## Producing evidence-based diagnostic recommendations

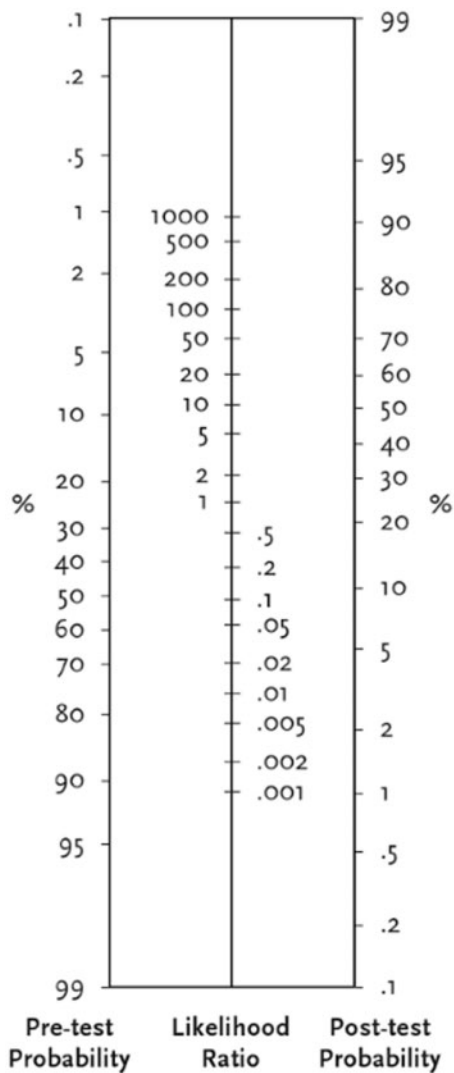
Coping with cognitive overload is an almost insurmountable task. It is inconceivable that any individual can keep up with the literature even in a specific field, such as cardiovascular imaging,<sup>90</sup> so reading and synthesis must be a collective endeavour, just as diagnosis should also be a collaborative effort.<sup>4</sup> This places a heavy responsibility on the authors of expert consensus statements and clinical guidelines. Contributors should stand aside if they have academic conflicts of interests such as patents or primary publications for a new diagnostic method.

The categories of evidence that should be available from studies of diagnostic imaging have been recognized for many years,<sup>91</sup> and standard tools are available to judge their quality,<sup>92</sup> but in cardiovascular diagnostic imaging the 'evidence base' from which we start is not

impressive. In a review of guidelines by the American College of Cardiology and the American Heart Association, published in 2009, the authors identified 333 recommendations concerning echocardiography, not one of which was linked to evidence.<sup>93</sup> All 84 recommendations relating to the use of radionuclide imaging were supported by evidence but was it was at level A for only 4.8%. Guidelines need to be based on comprehensive systematic reviews and meta-analyses and their processes and judgements must be transparent so that their conclusions are re-testable. They could then be debated like any scientific hypothesis. Two recent publications can illustrate the need for reform.

The EACVI and ASE revised their recommendations for evaluating diastolic function.<sup>94</sup> They are based on 8 variables, giving a total of 40 320 possible combinations (8 factorial), but specific combinations are described so some subjects will be unclassifiable. A major criterion is the ratio of the velocity of early diastolic mitral inflow to the early diastolic velocity of long-axis lengthening of the LV ( $E/e'$ ), although a meta-analysis concluded that  $E/e'$  cannot reliably estimate LV filling pressure in heart failure with preserved ejection fraction (HFpEF).<sup>95</sup> In the Stanislas cohort in France, applying the new recommendations compared with those that they replaced would reduce the prevalence of diastolic dysfunction from 6.3% to 1.1% in subjects aged 40–60 years and from 12.9% to 3.1% in those aged >60 years.<sup>96</sup> In the SABRE cohort in London, 89% of 1395 subjects with a mean age of 69 years would now be abnormal on the basis of their LV long-axis early diastolic velocities (Alun Hughes, personal communication), probably because the recommended cut points are unrelated to age. Any criterion that defines most of the target population as abnormal will not have discriminant value.

The second example relates to the revised guideline on heart failure from the European Society of Cardiology.<sup>97</sup> This starts with the preamble that 'Guidelines summarize and evaluate all available evidence on a particular issue at the time of the writing process', but then states that 'the main terminology used to describe heart failure is historical'



**Figure 3** Nomogram for applying Bayes' Theorem. From New England Journal of Medicine, Fagan TJ, Letter: nomogram for Bayes theorem, 293:257. Copyright © 1975, Massachusetts Medical Society. Reprinted with permission.<sup>87</sup>

and that normal LVEF is 'typically considered as  $\geq 50\%$ '. It continues: 'Patients with an LVEF in the range of 40-49% represent a 'grey area', which we now define as heart failure with mid-range ejection fraction (HFmrEF)'. Subdividing EF in this way and creating a new disease is irrational when EF is normally distributed and has considerable measurement variability. Unsurprisingly, repeated measurements often lead to reclassification. Using the previous cut point of 50%, then during a mean follow-up of 5.1 years, 39% of patients classified as heart failure with reduced EF (HFrEF) had an EF  $> 50\%$  and 39% of patients with HFpEF had an EF  $< 50\%$ .<sup>98</sup> In a recent study that applied the new recommendation, within 1 year of their diagnosis as HFmrEF 44% of patients had transitioned to HFpEF and 16% to HFrEF.<sup>99</sup> In some patients in both studies the changes may have been real, due to progression of disease or response to treatment, but in others the redesignation can be explained by variation within the reference change value.

These documents were surely the result of considerable study and deliberation, but the key point is that both recommendations have been revised without clear evidence being presented to support the changes and without it being possible to understand why they were made. Interactive electronic guidelines would allow others to study the same data and reassess the conclusions. Most importantly, new recommendations that represent hypotheses should be evaluated prospectively before they are included in guidelines. In diagnostic imaging, the use of appropriateness criteria<sup>100</sup> may be helpful, but they are a substitute for developing evidence, and in some circumstances, their introduction has had little impact on clinical practice.<sup>101</sup>

## Using machine learning and decision support

A provocative summary could be that we take diagnostic imaging tests that are objective, quantifiable, and continuously related to disease and that have established sensitivity, specificity, and positive and negative predictive values, but then we use them in patients without considering the pre-test probability and by interpreting the results subjectively rather than objectively. We prefer to apply single cut points rather than age- and gender-specific reference ranges, and we fail to report the imprecision implicit in any measurement.

Computer-aided diagnostic support has existed for more than 40 years but only recently has it reached the stage where it might make a substantial difference. Early studies showed that decision-support could perform as well as experts; now it can do better, particularly when analysing huge data sets. Information technology helps with diagnosis, research, and audit, but its potential for diagnostic imaging will be harnessed only if clinicians interact with informatics engineers to ensure that the tools that we need are developed (Table 1).

For diagnosis, automated systems can analyse faster and more accurately than visual assessment,<sup>102</sup> and machine learning can reveal diagnostic patterns in complex data, including imaging data collected during exercise, that would not otherwise be discernible.<sup>103-105</sup> Computer-reporting tools can help novices to gain experience rapidly,<sup>106</sup> shortening the learning curve and reducing net risk to patients. Using 'clinical knowledge support' improves patient safety, reduces complications, and shortens length of stay; adverse outcomes were halved in hospitals with the highest rates of use.<sup>107</sup> Diagnostic reporting systems could prompt the interpreter to consider alternative diagnoses or to reanalyse data that are inconsistent, and given relevant details, they could calculate pre- and post-test probabilities for the individual patient. Giving this information to the clinician rather than a simple yes/no answer would be a more honest way of reporting a diagnostic study.

More objective methods of analysing images have important applications in research. When open access was allowed to trial data, reanalysis by independent investigators led to different conclusions in 35% of studies.<sup>108</sup> Examples from diagnostic imaging included a computer-assisted study of regional wall motion abnormalities, instead of visual interpretation, and an MR study of myocardial infarction that applied a new measurement; each review concluded that one treatment was effective, after the original analysis had found no difference. Open access to imaging databases such as the Biobank study in the UK and the Cardiac Atlas project in the USA may yield

substantial dividends. The optimal size of a training set to develop automated methods for detecting heart failure has been estimated to be 4000 patients.<sup>109</sup>

A 'big data' audit study of 3.6 million hospitalizations demonstrated that patients who had echocardiography during their admission for heart failure, had 20% lower mortality,<sup>110</sup> implying that accurate diagnosis may have guided more effective treatment.

## Regulatory governance of diagnostic imaging

Diagnostic imaging machines are medical devices. Until recently, their market access in Europe depended on the manufacturer demonstrating that its device was safe (with a positive ratio of net benefit relative to risk) and that it performed its designated tasks satisfactorily.<sup>111</sup> Most diagnostic imaging systems are intermediate-risk (Class II) devices, so their authorization has not depended on the manufacturer submitting supporting evidence from clinical trials.

In 2016, a revised guide to the clinical evaluation of medical devices was published by the European Commission.<sup>112</sup> For the first time, this gives examples of performance data for diagnostic imaging that manufacturers can submit for approval, including sensitivity, specificity and reproducibility, comparisons of iterations of diagnostic software with previous versions, and normal values by age and gender for all groups in which the diagnostic system may be used. Adherence to the advice is voluntary, but manufacturers might collect these data if clinicians considered their availability as a factor when choosing which system to purchase.

The European Union process for approving medical devices is changing. New Regulations were published in May 2017 that will be fully implemented from 2020.<sup>113,114</sup> Requirements for clinical evidence will be strengthened especially for high-risk (Class III) devices. There is provision for designation of expert laboratories; one undertaking preclinical evaluation of diagnostic imaging systems would be an excellent choice. The Regulation on *in vitro* diagnostic devices (Annex 1, Chapter II, Paragraph 9.1.b) states that a manufacturer shall demonstrate the scientific validity of its method by submitting results from clinical performance studies including 'diagnostic sensitivity, diagnostic specificity, positive predictive value, negative predictive value, likelihood ratio, (and) expected values in normal and affected populations'.<sup>114</sup> Similar standards could be developed for imaging tests but they would need to be included in advisory documents. Here there will be an important role for experts to advise regulators on what is needed, from a clinical perspective.<sup>115</sup>

It can be argued that manufacturers have not integrated clinical decision support, because their machines have been approved without it. Smarter diagnostic systems would reduce diagnostic error and provide better reports, so their development should be a responsibility for manufacturers, shared with researchers. It is hoped that the new regulations will promote more evidence-based imaging (see Table 2) and encourage research without stifling innovation.

## Conclusions

The cardiologist Sir Thomas Lewis invented the term 'clinical science'. He wrote in 1929 that 'the lack of progress on the clinical side of

Medicine (and Surgery) is due chiefly, not to inherent difficulties presented by the subject, but to what has become a traditionally low standard of work and thought from a scientific standpoint'.<sup>116</sup> It would be unfair and misleading to imply that a huge amount of excellent research has not already been done in cardiovascular imaging, but we can also agree with Lewis that accomplished thinkers could achieve much more.

New diagnostic tools are expensive, and diagnostic tests account for an increasing proportion of the burgeoning costs of health care. There are major geographical and inter-hospital variations in the use of diagnostic imaging which cannot be explained by differences in the prevalence of disease. Many of the principles that we should adopt when developing collaborative multimodality imaging have been described,<sup>117,118</sup> but they remain insufficiently implemented. Our subspecialty is much more than 'imaging'; our challenge is to make it better clinical science.

## Acknowledgement

I would like to thank my first mentors in echocardiographic research, Professors George Sutherland and Jos Roelandt, for their vital encouragement, and my colleagues and fellows for their continuing support and interesting questions.

**Conflict of interest:** None declared.

## References

- Fraser AG. Inge Edler and the origins of clinical echocardiography. *Eur J Echocardiogr* 2001;**2**:3–5.
- World Health Organization. *International Statistical Classification of Diseases and Related Health Problems 10th Revision (2010)/ICD 10 5th Edition/Version 2016*. <http://apps.who.int/classifications/icd10/browse/2016/en> (1 September 2017, date last accessed).
- Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf* 2013; **22**(Suppl. 2):ii21–7.
- Balogh EP, Miller BT, Ball JR, eds. Committee on Diagnostic Error in Health Care, National Academies of Sciences, Engineering, and Medicine. *Improving Diagnosis in Health Care*. Washington, DC: The National Academies Press; 2015. ISBN 9780309377690.
- Benavidez OJ, Gauvreau K, Jenkins KJ, Geva T. Diagnostic errors in pediatric echocardiography: development of taxonomy and identification of risk factors. *Circulation* 2008;**117**:2995–3001.
- Benavidez OJ, Gauvreau K, Geva T. Diagnostic errors in congenital echocardiography: importance of study conditions. *J Am Soc Echocardiogr* 2014;**27**:616–23.
- Ward CJ, Purdie J. Diagnostic accuracy of paediatric echocardiograms interpreted by individuals other than paediatric cardiologists. *J Paediatr Child Health* 2001;**37**:331–6.
- Faletta F, Constantin C, De Chiara F, Masciocco G, Santambrogio G, Moreo A. Incorrect echocardiographic diagnosis in patients with mechanical prosthetic valve dysfunction: correlation with surgical findings. *Am J Med* 2000;**108**:531–7.
- Fox KF, Porter A, Unsworth B, Collier T, Leech G, Mayet J. Report on a national echocardiography quality-control exercise. *Eur J Echocardiogr* 2008;**10**:314–8.
- Bomma C, Rutberg J, Tandri H, Nasir K, Roguin A, Tichnell C et al. Misdiagnosis of arrhythmogenic right ventricular dysplasia/cardiomyopathy. *J Cardiovasc Electrophysiol*. 2004;**15**:300–6.
- Yan RT, Miller JM, Rochitte CE, Dewey M, Niinuma H, Clouse ME et al. Predictors of inaccurate coronary arterial stenosis assessment by CT angiography. *JACC Cardiovasc Imaging* 2013;**6**:963–72.
- Thurnheer R, Hoess C, Doenecke C, Moll C, Muntwyler J, Krause M. Diagnostic performance in a primary referral hospital assessed by autopsy: evolution over a ten-year period. *Eur J Intern Med* 2009;**20**:784–7.
- Kahneman D. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux; 2011. ISBN: 978-0374275631.
- Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974;**185**:1124–31.
- Croskerry P. Clinical cognition and diagnostic error: applications of a dual process model of reasoning. *Adv Health Sci Educ* 2009;**14**:27–35.
- Croskerry P. From mindless to mindful practice—cognitive bias and clinical decision making. *N Engl J Med* 2013;**368**:2445–8.



17. Saleeb SF, Margossian R, Spencer CT, Alexander ME, Smoot LB, Dorfman AL et al. Reproducibility of echocardiographic diagnosis of left ventricular noncompaction. *J Am Soc Echocardiogr* 2012;**25**:194–202.
18. Kohli SK, Pantazis AA, Shah JS, Adeyemi B, Jackson G, McKenna WJ et al. Diagnosis of left-ventricular non-compaction in patients with left-ventricular systolic dysfunction: time for a reappraisal of diagnostic criteria? *Eur Heart J* 2008;**29**:89–95.
19. Petersen SE, Selvanayagam JB, Wiesmann F, Robson MD, Francis JM, Anderson RH et al. Left ventricular non-compaction: insights from cardiovascular magnetic resonance imaging. *J Am Coll Cardiol* 2005;**46**:101–5.
20. Greutmann M, Mah ML, Silversides CK, Klaassen S, Attenhofer Jost CH, Jenni R et al. Predictors of adverse outcome in adolescents and adults with isolated left ventricular noncompaction. *Am J Cardiol* 2012;**109**:276–81.
21. Rooms I, Dujardin K, De Sutter J. Non-compaction cardiomyopathy: a genetically and clinically heterogeneous disorder. *Acta Cardiol* 2015;**70**:625–31.
22. Kohn LT, Corrigan JM, Donaldson MS, eds; Institute of Medicine (US) Committee on Quality of Health Care in America. *To Err is Human: Building a Safer Health System*. Washington (DC): National Academies Press (US); 2000. ISBN 0-309-06837-1.
23. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to improve physicians' use of diagnostic tests. A new conceptual framework. *JAMA* 1998;**280**:2020–7.
24. McDonald KM, Matesic B, Contopoulos-Ioannidis DG, Lohhart J, Schmidt E, Pineda N et al. Patient safety strategies targeted at diagnostic errors. *Ann Intern Med* 2013;**158**:381–9.
25. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ Qual Saf* 2013;**22**(Suppl. 2):ii58–64.
26. Croskerry P, Singhal G, Mamede S. Cognitive debiasing 2: impediments to and strategies for change. *BMJ Qual Saf* 2013; **22**(Suppl. 2):ii65–72.
27. Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. *Acad Med* 2011;**86**:307–13.
28. Bertrand PB, Levine RA, Isselbacher EM, Vandervoort PM. Fact or artifact in two-dimensional echocardiography: avoiding misdiagnosis and missed diagnosis. *J Am Soc Echocardiogr* 2016;**29**:381–91.
29. Faculty of Clinical Radiology. *Standards for Learning from Discrepancies Meetings*. London: The Royal College of Radiologists; 2014. [https://www.rcr.ac.uk/sites/default/files/docs/radiology/pdf/BFCR%2814%2911\\_LDMs.pdf](https://www.rcr.ac.uk/sites/default/files/docs/radiology/pdf/BFCR%2814%2911_LDMs.pdf) (1 September 2017, date last accessed).
30. Popescu BA, Stefanidis A, Nihoyannopoulos P, Fox KF, Ray S, Cardim N et al. Updated standards and processes for accreditation of echocardiographic laboratories from The European Association of Cardiovascular Imaging: an executive summary. *Eur Heart J Cardiovasc Imaging* 2014;**15**:1188–93.
31. Levine RA, Stathogiannis E, Newell JB, Harrigan P, Weyman AE. Reconsideration of echocardiographic standards for mitral valve prolapse: lack of association between leaflet displacement isolated to the apical four chamber view and independent echocardiographic evidence of abnormality. *J Am Coll Cardiol* 1988;**11**:1010–9.
32. Warth DC, King ME, Cohen JM, Tesoriero VL, Marcus E, Weyman AE. Prevalence of mitral valve prolapse in normal children. *J Am Coll Cardiol* 1985;**5**:1173–7.
33. Baltabaeva A, Marciniak M, Bijmens B, Moggridge J, He FJ, Antonios TF et al. Regional left ventricular deformation and geometry analysis provides insights in myocardial remodelling in mild to moderate hypertension. *Eur J Echocardiogr* 2008;**9**:501–8.
34. Chirinos JA, Segers P, Duprez DA, Brumback L, Bluemke DA, Zamani P et al. Late systolic central hypertension as a predictor of incident heart failure: the Multi-ethnic Study of Atherosclerosis. *J Am Heart Assoc* 2015;**4**:e001335.
35. Diaz T, Pencina MJ, Benjamin EJ, Aragam J, Fuller DL, Pencina KM et al. Prevalence, clinical correlates, and prognosis of discrete upper septal thickening on echocardiography: the Framingham Heart Study. *Echocardiography* 2009;**26**:247–53.
36. Peters SA, den Ruijter HM, Bots ML, Moons KG. Improvements in risk stratification for the occurrence of cardiovascular disease by imaging subclinical atherosclerosis: a systematic review. *Heart* 2012;**98**:177–84.
37. Elias-Smale SE, Kavousi M, Verwoert GC, Koller MT, Steyerberg EW, Mattace-Raso FU et al. Common carotid intima-media thickness in cardiovascular risk stratification of older people: the Rotterdam Study. *Eur J Prev Cardiol* 2012;**19**:698–705.
38. Fraser AG, Kyaw Y, Kozáková M, Palombo C. Ultrasonic imaging of the carotid arteries, from intima-media thickness to histological markers for plaque vulnerability: what do we know? *Dialog Cardiovasc Med* 2013;**18**:87–98. <https://www.dialogues-cvm.org/wp-content/documents/DCVM68.pdf> (1 September 2017, date last accessed).
39. Lorenz MW, Polak JF, Kavousi M, Mathiesen EB, Völzke H, Tuomainen TP et al. PROG-IMT Study Group. Carotid intima-media thickness progression to predict cardiovascular events in the general population (the PROG-IMT collaborative project): a meta-analysis of individual participant data. *Lancet* 2012;**379**:2053–62.
40. Simon A, Chironi G, Levenson J. Comparative performance of subclinical atherosclerosis tests in predicting coronary heart disease in asymptomatic individuals. *Eur Heart J* 2007;**28**:2967–71.
41. Inaba Y, Chen JA, Bergmann SR. Carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: a meta-analysis. *Atherosclerosis* 2012;**220**:128–33.
42. Kozáková M, Palombo C, Paterni M, Anderwald CH, Konrad T, Colgan MP et al.; Relationship between Insulin Sensitivity Cardiovascular risk Investigators. Body composition and common carotid artery remodeling in a healthy population. *J Clin Endocrinol Metab* 2008;**93**:3325–32.
43. Ng AC, Sitges M, Pham PN, Tran D. T, Delgado V, Bertini M et al. Incremental value of 2-dimensional speckle tracking strain imaging to wall motion analysis for detection of coronary artery disease in patients undergoing dobutamine stress echocardiography. *Am Heart J* 2009;**158**:836–44.
44. Weideman F, Jamal F, Kowalski M, Kukulski T, D'hooge J, Bijmens B et al. Can strain rate and strain quantify changes in regional systolic function during dobutamine infusion, B-blockade, and atrial pacing—implications for quantitative stress echocardiography. *J Am Soc Echocardiogr* 2002;**15**:416–24.
45. Urheim S, Edvardsen T, Torp H, Angelsen B, Smiseth OA. Myocardial strain by Doppler echocardiography. Validation of a new method to quantify regional myocardial function. *Circulation* 2000;**102**:1158–64.
46. Davidavicius G, Kowalski M, Williams RI, D'Hooge J, Di Salvo G, Pierre-Justin G et al. Can regional strain and strain rate measurement be performed during both dobutamine and exercise echocardiography, and do regional deformation responses differ with different forms of stress testing? *J Am Soc Echocardiogr* 2003;**16**:299–308.
47. Greenberg NL, Firstenberg MS, Castro PL, Main M, Travaglini A, Odabashian JA et al. Doppler-derived myocardial systolic strain rate is a strong index of left ventricular contractility. *Circulation* 2002;**105**:99–105.
48. Uusitalo V, Luotolahti M, Pietilä M, Wendelin-Saarenhovi M, Hartiala J, Saraste M et al. Two-dimensional speckle-tracking during dobutamine stress echocardiography in the detection of myocardial ischemia in patients with suspected coronary artery disease. *J Am Soc Echocardiogr* 2016;**29**:470–9.
49. Barbier P, Solomon SB, Schiller NB, Glantz SA. Left atrial relaxation and left ventricular systolic function determine left atrial reservoir function. *Circulation* 1999;**100**:427–36.
50. Margulescu AD, Rees E, Coulson RM, Rees AD, Vinereanu D, Fraser AG. Do left atrial strain and strain rate reflect intrinsic atrial function or are they determined by left ventricular function? *Kardiol Pol* 2015;**73**:539–48.
51. Sanchis L, Andrea R, Falces C, Lopez-Sobrinho T, Montserrat S, Perez-Villa F et al. Prognostic value of left atrial strain in outpatients with de novo heart failure. *J Am Soc Echocardiogr* 2016;**29**:1035–42.
52. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;**271**:703–7.
53. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE et al.; GRADE Working Group. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *Br Med J* 2008;**336**:1106–10.
54. Padmanabhan S, Melander O, Hastie C, Menni C, Delles C, Connell JM et al. Hypertension and genome-wide association studies: combining high fidelity phenotyping and hypercontrols. *J Hypertens* 2008;**26**:1275–81.
55. Ng D, Johnston JJ, Teer JK, Singh LN, Peller LC, Wynter JS et al.; NIH Intramural Sequencing Center (NISC) Comparative Sequencing Program. Interpreting secondary cardiac disease variants in an exome cohort. *Circ Cardiovasc Genet* 2013;**6**:337–46.
56. Vasani RS, Glazer NL, Felix JF, Lieb W, Wild PS, Felix SB et al. Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA* 2009;**302**:168–78.
57. Vinereanu D, Mädlar CF, Gherghinescu C, Ciobanu AO, Fraser AG. Cumulative impact of cardiovascular risk factors on regional left ventricular function and reserve: progressive long-axis dysfunction with compensatory radial changes. *Echocardiography* 2011;**28**:813–20.
58. Dalen H, Thorstensen A, Aase SA, Ingul CB, Torp H, Vatten LJ et al. Segmental and global longitudinal strain and strain rate based on echocardiography of 1266 healthy individuals: the HUNT study in Norway. *Eur J Echocardiogr* 2010;**11**:176–83.
59. Selvaraj S, Aguilar FG, Martinez EE, Beussink L, Kim KY, Peng J et al. Association of comorbidity burden with abnormal cardiac mechanics: findings from the HyperGEN study. *J Am Heart Assoc* 2014;**3**:e000631.
60. Echocardiographic Normal Ranges Meta-Analysis of the Left Heart Collaboration. Ethnic-specific normative reference values for echocardiographic LA and LV size, LV mass, and systolic function: the EchoNormal Study. *JACC Cardiovasc Imaging* 2015;**8**:656–65.
61. Fischer M, Baessler A, Hense HW, Hengstenberg C, Muscholl M, Holmer S et al. Prevalence of left ventricular diastolic dysfunction in the community.

- Results from a Doppler echocardiographic-based survey of a population sample. *Eur Heart J* 2003;**24**:320–8.
62. Staessen JA, Fagard RH, Lijnen PJ, Thijs L, Van Hoof R, Amery AK. Mean and range of the ambulatory pressure in normotensive subjects from a meta-analysis of 23 studies. *Am J Cardiol* 1991;**67**:723–7.
  63. Kikuya M, Hansen TW, Thijs L, Björklund-Bodegård K, Kuznetsova T, Ohkubo T; et al. International Database on Ambulatory blood pressure monitoring in relation to Cardiovascular Outcomes Investigators. Diagnostic thresholds for ambulatory blood pressure monitoring based on 10-year cardiovascular risk. *Circulation* 2007;**115**:2145–52.
  64. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002;**360**:1903–13.
  65. Toth G, Hamilos M, Pyxaras S, Mangiacapra F, Nelis O, De Vroey F et al. Evolving concepts of angiogram: fractional flow reserve discordances in 4000 coronary stenoses. *Eur Heart J* 2014;**35**:2831–8.
  66. Wiener RS, Schwartz LM, Woloshin S. Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. *Arch Intern Med* 2011;**171**:831–7.
  67. Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL et al. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;**169**:1881–7.
  68. Orme NM, Fletcher JG, Siddiki HA, Harmsen WS, O'byrne MM, Port JD et al. Incidental findings in imaging research: evaluating incidence, benefit, and burden. *Arch Intern Med* 2010;**170**:1525–32.
  69. D'Hooge J, Barbosa D, Gao H, Claus P, Prater D, Hamilton J et al.; EACVI/ASE/Industry Task Force to Standardize Deformation Imaging. Two-dimensional speckle tracking echocardiography: standardization efforts based on synthetic ultrasound data. *Eur Heart J Cardiovasc Imaging* 2016;**17**:693–701.
  70. Ann SH, Kim JH, Ha ND, Choi SH, Garg S, Singh GB et al. Reproducibility of coronary artery calcium measurements using 0.8-mm-thickness 256-slice coronary CT. *Jpn J Radiol* 2014;**32**:677–84.
  71. Kitkungvan D, Johnson NP, Roby AE, Patel MB, Kirkeeide R, Gould KL. Routine clinical quantitative rest stress myocardial perfusion for managing coronary artery disease: clinical relevance of test-retest variability. *JACC Cardiovasc Imaging* 2017;**10**:565–77.
  72. Margulescu AD, Thomas DE, Ingram TE, Vintila VD, Egan MA, Vinereanu D et al. Can isovolumic acceleration be used in clinical practice to estimate ventricular contractile function? Reproducibility and regional variation of a new noninvasive index. *J Am Soc Echocardiogr* 2010;**23**:423–31.
  73. Nagata Y, Takeuchi M, Mizukoshi K, Wu VC, Lin FC, Negishi K et al. Intervendor variability of two-dimensional strain using vendor-specific and vendor-independent software. *J Am Soc Echocardiogr* 2015;**28**:630–41.
  74. Farsalinos KE, Daraban AM, Ünlü S, Thomas JD, Badano LP, Voigt JU. Head-to-head comparison of global longitudinal strain measurements among nine different vendors: The EACVI/ASE Inter-Vendor Comparison Study. *J Am Soc Echocardiogr* 2015;**28**:1171–81.
  75. Mirea O, Pagourelis ED, Duchenne J, Bogaert J, Thomas JD, Badano LP et al.; EACVI-ASE-Industry Standardization Task Force. Variability and reproducibility of segmental longitudinal strain measurement: a report from the EACVI-ASE Strain Standardization Task Force. *JACC Cardiovascular Imaging* 2017; doi: 10.1016/j.jcmg.2017.01.027.
  76. Willemlink MJ, Takx RA, de Jong PA, Budde RP, Bleys RL, Das M et al. The impact of CT radiation dose reduction and iterative reconstruction algorithms from four different vendors on coronary calcium scoring. *Eur Radiol* 2014;**24**:2201–12.
  77. Symons R, Morris JZ, Wu CO, Pourmorteza A, Ahlman MA, Lima JA et al. Coronary CT angiography: variability of CT scanners and readers in measurement of plaque volume. *Radiology* 2016;**281**:737–48.
  78. Iraha R, Tsuchiya N, Yamashiro T, Iwasawa T, Murayama S. Reproducibility of pulmonary blood flow measurements by phase-contrast MRI using different 1.5 T MR scanners at two institutions. *Acta Radiol Open* 2017;**6**:2058460116684370.
  79. Voigt JU, Pedrizzetti G, Lysyansky P, Marwick TH, Houle H, Baumann R et al. Definitions for a common standard for 2D speckle tracking echocardiography: consensus document of the EACVI/ASE/Industry Task Force to standardize deformation imaging. *Eur Heart J Cardiovasc Imaging* 2015;**16**:1–11.
  80. Yang H, Marwick TH, Fukuda N, Oe H, Saito M, Thomas JD et al. Improvement in strain concordance between two major vendors after the strain standardization initiative. *J Am Soc Echocardiogr* 2015;**28**:642–8.
  81. Shiino K, Yamada A, Ischenko M, Khandheria BK, Hudaverdi M, Speranza V et al. Intervendor consistency and reproducibility of left ventricular 2D global and regional strain with two different high-end ultrasound systems. *Eur Heart J Cardiovasc Imaging* 2017;**18**:707–16.
  82. Harris EK, Yasaka T. On the calculation of a 'reference change' for comparing two consecutive measurements. *Clin Chem* 1983;**29**:25–30.
  83. Bugdayci G, Oğuzman H, Arattan HY, Sasmaz G. The use of reference change values in clinical laboratories. *Clin Lab* 2015;**61**:251–7.
  84. Sevrukov AB, Bland JM, Kondos GT. Serial electron beam CT measurements of coronary artery calcium: has your patient's calcium score actually changed? *Am J Roentgenol* 2005;**185**:1546–53.
  85. Chung H, McClelland RL, Katz R, Carr JJ, Budoff MJ. Repeatability limits for measurement of coronary artery calcified plaque with cardiac CT in the Multi-Ethnic Study of Atherosclerosis. *AJR Am J Roentgenol* 2008;**190**:W87–92.
  86. Bayes M, Price M. An essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr Price, in a Letter to John Canton. December 23, 1763. *Philos Trans R Soc Lond* 1763;**53**:370.
  87. Fagan TJ. Letter: nomogram for Bayes theorem. *N Engl J Med* 1975;**293**:257.
  88. Bangalore S, Gopinath D, Yao SS, Chaudhry FA. Risk stratification using stress echocardiography: incremental prognostic value over historic, clinical, and stress electrocardiographic variables across a wide spectrum of Bayesian pre-test probabilities for coronary artery disease. *J Am Soc Echocardiogr* 2007;**20**:244–52.
  89. Mårtensson M, Olsson M, Segall B, Fraser AG, Winter R, Brodin LA. High incidence of defective ultrasound transducers in use in routine clinical practice. *Eur J Echocardiogr* 2009;**10**:389–94.
  90. Fraser AG, Dunstan FD. On the impossibility of being expert. *Br Med J* 2010;**341**:c6815.
  91. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;**11**:88–94.
  92. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;**6**:e012799.
  93. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA* 2009;**301**:831–41.
  94. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF 3rd, Dokainish H, Edvardsen T et al. Recommendations for the evaluation of left ventricular diastolic function by echocardiography: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2016;**17**:1321–60.
  95. Sharifov OF, Schiros CG, Aban I, Denney TS, Gupta H. Diagnostic accuracy of tissue doppler index e/e' for evaluating left ventricular filling pressure and diastolic dysfunction/heart failure with preserved ejection fraction: a systematic review and meta-analysis. *J Am Heart Assoc* 2016;**5**:e002530.
  96. Huttin O, Fraser AG, Coiro S, Bozec E, Selton-Suty C, Lamiral Z et al. Impact of changes in consensus diagnostic recommendations on the echocardiographic prevalence of diastolic dysfunction. *J Am Coll Cardiol* 2017;**69**:3119–21.
  97. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ et al.; Authors/Task Force Members. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). *Eur Heart J* 2016;**37**:2129–200.
  98. Dunlay SM, Roger VL, Weston SA, Jiang R, Redfield MM. Longitudinal changes in ejection fraction in heart failure patients with preserved and reduced ejection fraction. *Circ Heart Fail* 2012;**5**:720–6.
  99. Tsuji K, Sakata Y, Nochioka K, Miura M, Yamauchi T, Onose T et al.; CHART-2 Investigators. Characterization of heart failure patients with mid-range left ventricular ejection fraction—a report from the CHART-2 Study. *Eur J Heart Fail* 2017;**19**:1258–69.
  100. Patel MR, Spertus JA, Brindis RG, Hendel RC, Douglas PS, Peterson ED et al.; American College of Cardiology Foundation. ACCF proposed method for evaluating the appropriateness of cardiovascular imaging. *J Am Coll Cardiol* 2005;**46**:1606–13.
  101. Fonseca R, Negishi K, Otahal P, Marwick TH. Temporal changes in appropriateness of cardiac imaging. *J Am Coll Cardiol* 2015;**65**:763–73.
  102. Knackstedt C, Bekkers SC, Schummers G, Schreckenber M, Muraru D, Badano LP et al. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs Multicenter Study. *J Am Coll Cardiol* 2015;**66**:1456–66.
  103. Sanchez-Martinez S, Duchateau N, Erdei T, Fraser AG, Bijmens BH, Piella G. Characterization of myocardial motion patterns by unsupervised multiple kernel learning. *Med Image Anal* 2017;**35**:70–82.
  104. Omar AM, Narula S, Abdel Rahman MA, Pedrizzetti G, Raslan H, Rifaie O et al. Precision phenotyping in heart failure and pattern clustering of ultrasound data for the assessment of diastolic dysfunction. *JACC Cardiovasc Imaging* 2017; doi: 10.1016/j.jcmg.2016.10.012.
  105. Katz DH, Deo RC, Aguilar FG, Selvaraj S, Martinez EE, Beussink-Nelson L et al. Phenomapping for the identification of hypertensive patients with the myocardial substrate for heart failure with preserved ejection fraction. *J Cardiovasc Transl Res* 2017;**10**:275–84.
  106. Ovrehus KA, Munkholm H, Böttcher M, Bøtker HE, Nørgaard BL. Coronary computed tomographic angiography in patients suspected of coronary artery

- disease: impact of observer experience on diagnostic performance and interobserver reproducibility. *J Cardiovasc Comput Tomogr* 2010;**4**:186–94.
107. Bonis PA, Pickens GT, Rind DM, Foster DA. Association of a clinical knowledge support system with improved patient safety, reduced complications and shorter length of stay among Medicare beneficiaries in acute care hospitals in the United States. *Int J Med Inform* 2008;**77**:745–53.
  108. Ebrahim S, Sohani ZN, Montoya L, Agarwal A, Thorlund K, Mills EJ et al. Reanalyses of randomized clinical trial data. *JAMA* 2014;**312**:1024–32.
  109. Ng K, Steinhubl SR, deFilippi C, Dey S, Stewart WF. Early detection of heart failure using electronic health records: practical implications for time before diagnosis, data diversity, data quantity, and data density. *Circ Cardiovasc Qual Outcomes* 2016;**9**:649–58.
  110. Papolos A, Narula J, Bavishi C, Chaudhry FA, Sengupta PP. U.S. Hospital use of echocardiography: insights from the Nationwide Inpatient Sample. *J Am Coll Cardiol* 2016;**67**:502–11.
  111. Fraser AG, Daubert JC, Van de Werf F, Estes NA 3rd, Smith SC Jr, Krucoff MW, participants et al. Clinical evaluation of cardiovascular devices: principles, problems, and proposals for European regulatory reform. Report of a policy conference of the European Society of Cardiology. *Eur Heart J* 2011;**32**:1673–86.
  112. European Commission (June 2016). Clinical evaluation: a guide for manufacturers and notified bodies under Directives 93/42/EEC and 90/385/EEC. MEDDEV 2.7/1 revision 4 (page 47). <http://ec.europa.eu/DocsRoom/documents/17522/attachments/1/translations/> (1 September 2017, date last accessed).
  113. European Union. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. [http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2017.117.01.0001.01.ENG&toc=OJ:L:2017:117:TOC](http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2017.117.01.0001.01.ENG&toc=OJ:L:2017:117:TOC) (1 September 2017, date last accessed).
  114. European Union. Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU. [http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L\\_.2017.117.01.0176.01.ENG&toc=OJ:L:2017:117:TOC](http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2017.117.01.0176.01.ENG&toc=OJ:L:2017:117:TOC) (1 September 2017, date last accessed).
  115. Fraser AG, Krucoff MW, Brindis RG, Komajda M, Smith CS Jr. Commentary: International collaboration needed on device clinical standards. *Br Med J* 2011;**342**:d2952.
  116. Hollmann A. *Sir Thomas Lewis—Pioneer Cardiologist and Clinical Scientist*. London: Springer; 1997. p138. ISBN 978-1-4471-0927-3.
  117. Fraser AG, Buser PT, Bax JJ, Dassen WR, Nihoyannopoulos P, Schwitter J. The future of cardiovascular imaging and non-invasive diagnosis: a joint statement from the European Association of Echocardiography, the Working Groups on Cardiovascular Magnetic Resonance, Computers in Cardiology, and Nuclear Cardiology, of the European Society of Cardiology, the European Association of Nuclear Medicine, and the Association for European Paediatric Cardiology. *Eur Heart J* 2006;**27**:1750–3.
  118. Douglas PS, Cerqueira MD, Berman DS, Chinnaiyan K, Cohen MS, Lundbye JB et al; ACC Cardiovascular Imaging Council. The future of cardiac imaging: report of a think tank convened by the American College of Cardiology. *JACC Cardiovasc Imaging* 2016;**9**:1211–23.