

Cardiovascular imaging "guidelines" – High time for a paradigm shift

Alan Fraser, Nicolas Girerd

▶ **To cite this version:**

Alan Fraser, Nicolas Girerd. Cardiovascular imaging "guidelines" – High time for a paradigm shift. International Journal of Cardiology, 2020, S0167-5273 (20), pp.34278-9. 10.1016/j.ijcard.2020.12.030 . hal-03135625

HAL Id: hal-03135625

https://hal.univ-lorraine.fr/hal-03135625

Submitted on 9 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cardiovascular imaging "guidelines" – high time for a paradigm shift

Alan G Fraser [1], Nicolas Girerd [2]

[1] Department of Cardiology, University Hospital of Wales, Cardiff, U.K.

[2] Université de Lorraine, Inserm Centre d'Investigations Cliniques-1433, and Inserm DCAC, CHRU Nancy, F-CRIN INI-CRCT, Nancy, France.

Address for correspondence:

Professor Alan G. Fraser,

University Hospital of Wales,

Heath Park,

Cardiff, CF14 4XW,

Wales, U.K.

fraserag@cf.ac.uk

We live in an era of unprecedented information overload in medicine. Even within a tightly defined field of research or clinical practice, it is impossible for an individual to keep abreast of the expanding medical literature. We have therefore seen the inexorable growth of clinical practice guidelines as a collective mechanism for summarising the evidence that should inform our clinical decisions – which gives the authors of guidelines a special responsibility to ensure the thoroughness and transparency of their search strategies as well as the accuracy and utility of their conclusions.

But we have a particular problem with diagnostic "guidelines" because insufficient high-quality evidence is available. In 2009 Tricoci et al reviewed guidelines from the American College of Cardiology and the American Heart Association. Diagnostic recommendations were defined as class I (meaning that the intervention is beneficial) in 43% and class II (possibly beneficial) in 51%, but the level of evidence behind those recommendations was high (A, meaning derived from multiple randomised controls trials and/or meta-analysis) in only 2.4%.[1] Forty four percent were based on consensus. In 2019 a review of current European Society of Cardiology (ESC) guidelines showed that there has been disappointingly little progress, since only 8.7% of 503 class I diagnostic recommendations are supported by level of evidence A.[2] In a similar analysis, Esposito et al calculated that 56% of European diagnostic recommendations were based on expert opinions.[3]

Accurate diagnosis is the bedrock of evidence-based clinical practice, but we know that diagnostic error is common. Ideally, the designation of a document as a "Guideline" should be reserved for an authoritative summary of all available evidence, that has been informed by systematic reviews and meta-analyses. Clinical practice guidelines could already be interactive living documents that can be searched on-line, updated collectively perhaps using wikis, and curated by professional scientific associations. If that process was transparent, then one would

expect different expert evaluators overviewing the same studies to verify their relevance and reach similar conclusions. Unfortunately, most diagnostic consensus statements fall short of those standards.

In this issue of the International Journal of Cardiology, Yeung et al present an interesting analysis of the consequences of changing diagnostic guidance.[4] They refer specifically to recommendations for the echocardiographic assessment of left ventricular (LV) diastolic function from the European Association of Cardiovascular Imaging and the American Society of Echocardiography, that were updated and simplified in 2016.[5] A particular strength of their study is that it is based on reports from >50,000 unique patients, which were coded by a computer algorithm that they share as supplementary material. Yeung et al found significant but modest differences in the proportions of subjects allocated to each diagnostic class using the 2016 criteria compared with the previous 2009 criteria. There was considerable variability at an individual level and poor overall agreement (kappa statistic 0.32) with up to 48% of studies receiving a different interpretation.[4] Notably, the updated recommendations leave many subjects unclassifiable, with 36% in an indeterminate category.[4] Yeung et al confirmed other recent reports that the revised criteria reclassify significant numbers of subjects with preserved ejection fraction and/or mild diastolic dysfunction to normal (Table).

Substantial changes would be justifiable – indeed, even desirable – if they were supported by evidence, but the new recommendations were published before they had been validated. One year later, the Eurofilling study reported that their sensitivity and specificity for identifying subjects with a high LV end-diastolic pressure were 75% and 74% respectively.[6] There is sound advice available from the Guidelines International Network that diagnostic criteria should not be changed without knowledge of the repeatability, reproducibility, and accuracy of any new definition, or without evidence of its impact on the expected prevalence and incidence

of the disease.[7] Ideally a new definition should also have greater prognostic ability to predict clinically meaningful outcomes.[7] It appears that these criteria were not met before the new diastolic recommendations were published, which reinforces the questions raised by Yeung et al.

Another prominent recent example concerns the diagnosis of heart failure with preserved ejection fraction, for which the assessment of diastolic function is a core criterion. The Heart Failure Association of the ESC proposed a new diagnostic score in 2019 before it had been tested[8] and despite the fact that an alternative and validated score had already been published.[9] A fundamental problem with recommendations for both diastolic function and heart failure is that they have tried to simplify diagnosis into broad categories that may encompass patients with different pathophysiological mechanisms of disease. Also, and surprisingly, they do not include criteria that assess diastolic function during exercise although that is when patients experience symptoms, or propose cut-offs that are adjusted for age and gender. These limitations may partly explain why clinical trials that have recruited patients using current diagnostic labels have failed to discover effective treatments.

Developing diagnostic recommendations by consensus and without prior evidence of their impact has become a disservice to our patients that we should no longer accept. The availability of imaging biobanks, big echocardiographic databases, advances in informatics and computing technology, and developments in artificial intelligence, mean that we already know how we could do things better. As just one example, a recent analysis of 44,886 patients in the Swedish Heart Failure Registry showed that LV ejection fraction did not predict mortality at 1 year whereas an unsupervised random forest analysis identified four cohorts with a differential mortality hazard ratio of 4.5.[10]

In future, phenomapping derived from interpretable machine learning is likely to identify

new diagnostic criteria that we can use to predict responses to specific treatments. We should be asking our patients for permission to include their data pseudonymously within large collaborative imaging biobanks, to be used for developing and independently validating criteria that correlate with outcomes. We should only amend diagnostic recommendations when we have substantial evidence that they are more accurate or useful. This change is overdue but evolution has been slow, so it is high time to do things differently. Guidelines incorporating indisputable evidence-based algorithms would attract great support and probably be universally applied.

References

1. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC. Scientific evidence underlying the ACC/AHA Clinical Practice Guidelines. JAMA. 2009;301:831–41.

2. van Dijk WB, Grobbee DE, de Vries MC, Groenwold RHH, van der Graaf R, Schuit E. A systematic breakdown of the levels of evidence supporting the European Society of Cardiology guidelines. Eur J Prev Cardiol. 2019;26:1944–52.

3. Esposito A, Gallone G, Palmisano A, Marchitelli L, Catapano F, Francone M. The current landscape of imaging recommendations in cardiovascular clinical guidelines: toward an imaging-guided precision medicine. Radiol Med. 2020;125:1013–23.

4. Yeung DF, Jiang R, Behnami D, Jue J, Sharma R, Turaga M, Luong CL, Tsang MY, Gin KG, Girgis H, Lee PK, Nair P, Abolmaesumi P, Tsang TS. Impact of the updated diastolic function guidelines in the real world. Int J Cardiol. 2020 Oct 31;S0167-5273(20)34059-6. doi: 10.1016/j.ijcard.2020.10.068. Online ahead of print.

5. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF 3rd, Dokainish H, Edvardsen T, Flachskampf FA, Gillebert TC, Klein AL, Lancellotti P, Marino P, Oh JK, Popescu BA, Waggoner AD. 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.

6. Lancellotti P, Galderisi M, Edvardsen T, Donal E, Goliasch G, Cardim N, Magne J, Laginha S, Hagendorff A, Haland TF, Aaberge L, Martinez C, Rapacciuolo A, Santoro C, Ilardi F, Postolache A, Dulgheru R, Mateescu AD, Beladan CC, Deleanu D, Marchetta S, Auffret V, Schwammenthal E, Habib G, Popescu BA. Echo-Doppler estimation of left ventricular filling pressure: results of the multicentre EACVI Euro-Filling study. Eur Heart J Cardiovasc Imaging. 2017;18:961–8.

7. Doust J, Vandvik PO, Qaseem A, Mustafa RA, Horvath AR, Frances A, Al-Ansary L, Bossuyt P, Ward RL, Kopp I, Gollogly L, Schunemann H, Glasziou P; Guidelines International Network (G-I-N) Preventing Overdiagnosis Working Group. Guidance for modifying the definition of diseases: a checklist. JAMA Intern Med. 2017;177:1020–5.

8. Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, Edelmann F, Fu M, Guazzi M, Lam CSP, Lancellotti P, Melenovsky V, Morris DA, Nagel E, Pieske-Kraigher E, Ponikowski P, Solomon SD, Vasan RS, Rutten FH, Voors AA, Ruschitzka F, Paulus WJ, Seferovic P, Filippatos G. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm. A consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). Eur Heart J. 2019;40:3297–317.

9. Reddy YNV, Carter RE, Obokata M, Redfield MM, Borlaug BA. A simple, evidence-based approach to help guide diagnosis of heart failure with preserved ejection fraction. Circulation 2018;138:861–70.

10. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, Dahlström U, O'Connor CM, Felker GM, Desai NR. Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. J Am Heart Assoc. 2018;7(8):e008081.

Table: Published comparisons of the 2009 and 2016 ASE/EACVI diagnostic recommendations for diastolic function

Publication	Number of subjects	Mean age (years)	Methodology	Main findings	Impact of change in 2016 recommendations, comments
Huttin O et al J Am Coll Cardiol. 2017; 69: 3119–21	1,485	47	Comparison of 4 methods of quantifying diastolic function (including the 2009 and 2016 recommendations).	The proportion of individuals with diastolic dysfunction in this population study varied greatly according to the algorithm used; and the variations were influenced by age.	Diastolic dysfunction almost disappeared in individuals without overt cardiovascular disease. This suggests that the threshold may be too high for it to be used as a screening tool in populations.
Balaney B et al J Am Soc Echocardiogr. 2018; 31: 79–88	90	61	Echocardiography before left heart catheterisation in patients without heart failure.	The 2016 algorithm had a sensitivity of 0.69 and specificity of 0.81, for predicting a pre-a LV diastolic pressure >12 mmHg. The 2009 algorithm had a sensitivity of 0.79 and specificity of 0.70.	Higher specificity, but less sensitivity at <0.70 which is rather low for a screening test.
Sanchis L et al J Am Soc Echocardiogr. 2018; 31: 1203–8	157	73	Patients referred to ambulatory clinic; 48% had heart failure.	By the 2009 recommendations, 68% had grade I diastolic dysfunction; by 2016, 49% were reclassified as normal diastolic function. Only the 2016 classification was associated with prognosis (p=0.003).	The new diagnostic criteria appear more stringent but they are more related to outcome. A quarter of patients were in the undetermined class.

Gopalakrishnan P et al Echocardiography. 2020; 37: 546–53	20,843	67	Studies from a single high-volume centre analysed before and twice after revision of the guidance.	Using the 2016 recommendations decreased diastolic dysfunction by 57%. Heterogeneity for reporting diastolic function improved for grades 2 and 3 but worsened for grade 1.	Failure/inability to report diastolic function increased from 5% to 20%, which the authors suggest may imply a decrease of confidence in their use.
Bursi F et al Int J Cardiol. 2020; 305: 18–24	380	66	Independent blinded investigators graded echocardiographic digital images from patients with acute coronary syndromes.	Proportions in each diastolic grade were similar but concordance was weak (kappa = 0.21). Both definitions predicted heart failure and death (for 2016, hazard ratio 2.15, p <0.001, with AUC 0.79; for 2009, hazard ratio 1.82, p<0.001, with AUC 0.66) (p=0.002).	Agreement between the two recommendations was poor. The 2016 definition had more prognostic value.
Lin TT et al Eur Radiol. 2020; 30: 630–9	451	65	Prognostic power of the 2009 and 2016 recommendations compared in patients with heart failure with preserved ejection fraction.	Using the 2016 classification, grades II and III predicted mortality, and grade III hospitalisations for heart failure. With the 2009 guidance, only grade III was associated with mortality. There was no difference between algorithms for predicting heart failure hospitalisations.	The net reclassification index for mortality increased using the 2016 algorithm (by 10.6%, p < 0.001). Its ability to predict mortality, but not episodes related to heart failure, suggests that the 2016 classification does not have better specificity.
Yeung DF et al Int J Cardiol. 2020[4]	52,083	59	Echocardiographic reports reanalysed retrospectively by computer algorithm.	Increase in the proportion of subjects graded as normal, from 23% to 32%. Agreement between 2009 and 2016 definitions was poor (kappa = 0.32).	2016 recommendations decreased the proportion with undetermined diastolic function, but patients with preserved EF and no evidence of myocardial disease were reclassified

					from mild diastolic dysfunction to normal diastolic function.

ASE American Society of Echocardiography, EACVI European Association of Cardiovascular Imaging.
AUC area under the curve of a receiver-operating characteristic. EF ejection fraction.