Guest Editor: Javier Sanz

Maximal wall thickness measurement in hypertrophic cardiomyopathy–Biomarker variability and its impact on clinical care

Brief title: Maximal wall thickness in HCM

Gabriella Captur MD MRCP PhD MSc[1,2,3], Charlotte H Manisty MBBS PhD[3,4], Betty Raman MBBS PhD[5], Alberto Marchi MD[6], Timothy C Wong PhD[7,8,9], Rina Ariga MBBS PhD[5], Anish Bhuva MBBS PhD[3,4], Elizabeth Ormondroyd MSc[10], Ilaria Lobascio MD PhD[4], Claudia Camaioni MD[4], Savvas Loizos MD[4], Jenade Bonsu-Ofori MSc[11], Aslan Turer PhD[12], Vlad G Zaha MD PhD[12], João B Augutsto MD[3,4], Rhodri H Davies MBBS PhD[3,4], Andrew J Taylor PhD[13,14,15], Arthur Nasis MD[16], Mouaz H Al-Mallah MD[17], Sinitsyn Valentin MD[18], Diego Perez de Arenaza MD[19], Vimal Patel MD PhD[3], Mark Westwood MBBS[4], Steffen E Petersen MBBS PhD FRCP[4,20], Chunming Li PhD[21,22], Lijun Tang MD[23], Shiro Nakamori PhD[24], Reza Nezafat PhD[24], Raymond Y Kwong PhD[25], Carolyn Y Ho PhD[25], Alan G Fraser PhD[26], Hugh Watkins PhD[10], Perry M Elliott MBBS, MD, FRCP[3,4], Stefan Neubauer FMedSci FRCP[5], Guy Lloyd FRS DPhil[3,4], Iacopo Olivotto MD PhD[6], Petros Nihoyannopoulos MD FRCP FACC FESC[27], James C Moon MBBS MD FRCP[3,4]

1. UCL MRC Unit for Lifelong Health and Ageing, University College London, Fitzrovia, London WC1E 7HB, UK.
2. The Royal Free Hospital, Centre for Inherited Heart Muscle Conditions, Cardiology Department, Pond Street, Hampstead, London NW3 2QG, UK
3. UCL Institute of Cardiovascular Science, University College London, Gower Street, London WC1E 6BT, UK.
4. Barts Heart Center. St Bartholomew's Hospital, West Smithfield, London EC1A 7BE, UK.
5. University of Oxford Centre for Clinical Magnetic Resonance Research (OCMR), Division of Cardiovascular Medicine, Radcliffe Department of Medicine, Oxford NIHR Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK.
6. Cardiomyopathy Unit and Genetic Unit, Careggi University Hospital, Florence, Italy.
7. Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA.
8. UPMC Cardiovascular Magnetic Resonance Center, Heart and Vascular Institute, Pittsburgh, Pennsylvania, USA.
9. Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.
10. Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.
11. UCL Medical School, Bloomsbury Campus, University College London, Gower Street, London WC1E 6BT, UK.
12. Cardiology Division, Department of Internal Medicine and the University of Texas Southwestern Medical Center, Dallas, Texas, USA.
13. Department of Cardiovascular Medicine, Alfred Hospital, Melbourne Australia.
14. Baker Heart and Diabetes Institute, Melbourne Australia.
15. Department of Medicine, Monash University, Melbourne Australia.
16. Monash Cardiovascular Research Centre, MonashHEART, Monash University, Clayton, Australia.
17. King Abdulaziz Cardiac Center (KACC) (Riyadh), National Guard Health Affairs, Kingdom of Saudi Arabia.

18. Lomonosov Moscow State University, Department of Multidisciplinary Clinical Studies, Federal Center of Treatment and Rehabilitation Moscow, Russia.
19. Cardiology Department, Hospital Italiano de Buenos Aires, Argentina.
20. William Harvey Research Institute, NIHR Barts Biomedical Research Centre, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK.
21. Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA.
22. School of Electronic Engineering, University of Electronic Science and Technology of China (UESTC), Chengdu P.R. China.
23. The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Rd, Gulou, Nanjing, Jiangsu, China, 210029.
24. Department of Medicine (Cardiovascular Division) Beth Israel Deaconess Medical Center, Harvard Medical School, Cardiology East Campus, Room E/SH455, 330 Brookline Ave, Boston, MA, 02215, USA
25. Cardiovascular Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.
26. School of Medicine, Cardiff University, Cardiff CF14 4YS, UK
27. Imperial College London, NHLI, National Heart & Lung Institute, The Hammersmith Hospital, Du Cane Road, London W12 0NN, UK.

CORRESPONDENCE:
James C Moon MBBS, MD, FRCP
Professor and Consultant Cardiologist
UCL and Barts Heart Center London
University College London, Gower Street
London WC1E 6BT, UK
Email: j.moon@ucl.ac.uk
Phone No: +44 2039316477

SOURCES OF FUNDING: This program was funded by the Barts Charity grant #1107/2356/MRC0140 to G.C. G.C. is supported by British Heart Foundation Special Programme Grant MyoFit46 (SP/20/2/34841), the National Institute for Health Research Rare Diseases Translational Research Collaboration (NIHR RD-TRC) and by the NIHR UCL Hospitals Biomedical Research Center. S.E.P. acknowledges support from the Barts Biomedical Research Centre funded by the NIHR. J.C.M. is directly and indirectly supported by the UCL Hospitals NIHR BRC and Biomedical Research Unit at Barts Hospital respectively.

FINANCIAL DISCLOSURES & CONFLICTS OF INTEREST: NA.

ACKNOWLEDGEMENTS: NA.

ABSTRACT

Objectives: To define the variability of maximal wall thickness (MWT) measurements across modalities and predict its impact on care in patients with hypertrophic cardiomyopathy (HCM).
Background: Left ventricular MWT measured by echocardiography or cardiovascular magnetic resonance (CMR) contributes to the diagnosis of HCM, stratifies risk and guides key decisions including whether or not to implant a cardioverter defibrillator (ICD).
Methods: A 20-center global network provided paired echocardiographic and CMR datasets from patients with HCM from which 17 paired datasets of the highest quality were selected. These were presented as 7 randomly ordered pairs (at 6 cardiac conferences) to experienced readers who report HCM imaging in their daily practice and their MWT caliper measurements were captured. The impact of the measurement variability on ICD implant decisions was estimated in 769 separately recruited multi-center HCM patients using the European Society of Cardiology (ESC) 5-year risk of sudden cardiac death algorithm.
Results: MWT analysis was completed by 70 readers (from 6 continents; 91% with >5years experience). 79% and 68% scored echocardiography and CMR image quality as excellent. For both modalities (echocardiography and then CMR results) intra-modality inter-reader MWT percent variability was large (range | standard deviation: –59 to 117% | ±20%; –61 to 52% | ±11%). Agreement between modalities was low (standard error of measurement 4.8mm [95% confidence intervals 4.3–5.2mm], modest correlation r=0.56). In the multi-center HCM cohort, this estimated echocardiographic MWT percent variability (±20%) applied to the ESC algorithm, reclassified risk in 19.5% which would have led to inappropriate ICD decision-making in one in seven HCM patients (8.7% would have had ICD recommended despite potential low risk; 6.8% would not have had ICD recommended despite intermediate or high risk).
Conclusion: Using best available images and experienced readers, MWT as a biomarker in HCM has a high degree of inter-reader variability and should be applied with caution as part of decision-making for ICD implantation. Better standardization efforts in HCM recommendations by current governing societies are needed to improve our clinical decision-making in HCM patients.

KEY WORDS: Hypertrophic cardiomyopathy, wall thickness, cardiovascular magnetic resonance, echocardiography.

ABBREVIATIONS:
CMR, cardiovascular magnetic resonance
ESC, European Society of Cardiology
HCM, hypertrophic cardiomyopathy
ICD, implantable cardioverter defibrillator
LVH, left ventricular hypertrophy
MWT, maximal wall thickness
SAX, short axis
SCD, sudden cardiac death
SD, standard deviation
SEM, standard error of measurement

INTRODUCTION

The left ventricular (LV) wall thickens in hypertrophic cardiomyopathy (HCM) measured by cardiac imaging determines diagnosis, risk stratification and therapeutic decisions. The precision of its measurement influences estimates of prognosis, so it is key to clinical confidence and optimal decision-making.

Measurement importance is most codified for HCM. Both the European Society of Cardiology(1–3) (ESC) and American College of Cardiology/American Heart Association(4,5), place LV maximal wall thickness (MWT) as central to clinical care including diagnosis, sudden cardiac death (SCD) risk stratification, and to guide interventions such as myectomy. It is used either as a continuous variable in primary prevention risk algorithms or as a dichotomous cut-point (MWT >30mm) to indicate heightened risk, and to determine the clinical appropriateness of implantable cardioverter defibrillators (ICD): ESC–class IIa-b, level of evidence B(3) and American College of Cardiology/American Heart Association class IIa, level of evidence C respectively(5).

Consequently, the LV MWT biomarker is a key part of echocardiographic and cardiovascular magnetic resonance (CMR) clinical work(2,6). However, measurement reproducibility is affected by variations in scanning protocols, image quality, and readers' training, reliance, and understanding of available guidelines.

We sought to define “best-case” global variability in HCM MWT measurement and assess its potential impact on clinical decision-making using a multi-institutional cohort of HCM patients. Accordingly, we measured “best-case” human performance (excellent images, experienced clinicians) and assessed the predicted impact of this measurement variation on SCD risk stratification and consequent ICD implant recommendations.

METHODS

Imaging dataset. Twenty international HCM centers with both echocardiography and CMR imaging facilities, were invited to share at least one pair of high-quality, temporally matched HCM echocardiography and CMR datasets in a single patient. Inclusion criteria were echocardiography and CMR within 6 months of each other with cines devoid of contrast agent; any morphological HCM subtype; ≥15mm of LV hypertrophy (LVH) and not individuals with familial HCM criteria and lesser degrees of hypertrophy. A total of 78 paired echocardiography/CMR HCM imaging datasets were received (78 patients). From these 17 pairs (34 datasets) with the best image quality were chosen by a 10-person expert panel (**Supplementary Table 1**). Selected datasets were from 14 centers (8 countries, 5 in the USA; **Supplementary Table 2**), and time between echocardiography and CMR was +/–25 days. All patients had provided written informed consent conforming to the Declaration of Helsinki (fifth revision, 2000) and contributing centers had the approval of their national research ethics services and institutional review boards. Image views were standard – echocardiography: parasternal long axis, apical 4-chamber, and the 3 usual parasternal short axis (SAX) views (base, mid-cavity, apex) – CMR (same patient): 3-chamber, 4-chamber and the entire LV SAX cine stack refined to 3 SAX cines matching the echocardiography slices using point-to-curve mean error methodology, **Supplementary Table 3**.

Datasets were organized into modules consisting of 14 (out of the 34) randomly ordered datasets (**Figure 1**) comprised of 5 paired echocardiographic/CMR datasets, 1 duplicate echocardiographic pair and 1 duplicate CMR pair. Each dataset was organized as follows: parasternal long axis (or 3-chamber for CMR), 4-chamber, basal, mid and apical SAX, with a full cardiac cycle ready for diastolic frame selection and caliper measurements. HCM morphological subtype was defined as previously described(7). The septomarginal trabeculum was defined as previously described(8).

Interview procedure. We planned for face-to-face measurements to take a maximum of 30 minutes to ensure reader acceptability. Randomly sorted datasets were presented using OsiriX MD on 15-inch laptops with retina display. Measurements took place at 6 cardiac conferences (**Supplementary Table 4**) via scheduled face-to-face meetings targeting experienced clinicians and academics who report HCM images by echocardiography and CMR in their daily practice. Recruitment was via posters (**Supplementary Figure 1**) or direct invitation to senior readers/guideline contributors (email/telephone). An instruction leaflet (**Supplementary Figure 2**) explained the analysis exercise. There were four steps: firstly, readers were shown a brief movie of the images (e.g. **Supplementary Movies 1** and **2**) and invited to grade the dataset image quality; secondly, readers completed a web-based questionnaire to capture professional experience and reporting practices (Research Electronic Data Capture(9,10), REDCap); thirdly, readers performed the analysis without prior instruction on the choice of imaging plane (readers could look at the clip of a whole cardiac cycle before selecting which phase to caliper); finally, readers were invited to re-grade dataset image quality.

Image post-processing. Calipers and MWT measurements were saved as regions of interest. Image post-processing used in-house scripts in Matlab (R2012b).

Risk simulation HCM cohort. To assess the impact of measurement variability, an international multi-center cohort of 769 patients with HCM (adult unexplained hypertrophy ≥15mm or ≥13mm in first degree relatives) was used. Patients were recruited from dedicated cardiomyopathy clinics at 4 hospitals: John Radcliffe Hospital, Oxford, UK (n=101, local research ethics committee # 09/H0604/110), The Heart Hospital, University College London Hospital, London, UK (n=110, local research ethics committee # 04/0035 and 11/LO/0913),

University of Pittsburgh, Pennsylvania, USA (n=72, local institutional review board approval # PRO009010051) and Careggi University Hospital, Italy (n=486, local institutional review board approval). Each participant provided written informed consent conforming to the Declaration of Helsinki (fifth revision, 2000). Estimates for 5-year risk of SCD were calculated using the ESC online clinical tool(2,6). Patients were categorized as having low (<4%), intermediate (≥4–<6%) or high (≥6%) 5-year risk based upon their original clinically-available data that included MWT by echocardiography. By re-running the ESC algorithm on each patient, we estimated the impact of increasing each individual patient's measurement by the mean variability of all echocardiographic MWT measurements, and then of decreasing all measurements by the same percentage, on the numbers of subjects then allocated within each risk category.

Statistics. Statistical analysis was performed in R (version 3.0.1)(11). Distribution of data were assessed on histograms and using the Shapiro-Wilk test. Continuous variables are expressed as mean ± 1 standard deviation (SD); categorical variables, as counts and percent. $\chi 2$ or Fisher's exact test was used for the comparison of noncontinuous variables between readers of different levels of expertise. Correlation between inter-modality MWT and SD and mean MWT was assessed with Pearson correlation coefficient. Agreement of MWT measurements (intra-modality inter-reader; inter-modality intra-reader; intra-modality intra-reader) was calculated as standard error of measurement (SEM) including 95% confidence intervals (CI) using a two-way analysis of variance approach and random effects model as previously described(12). Intra-reader SEMs expressed the random error by a typical reader, while inter-reader SEMs expressed a random error in measurements performed by different readers. The SEM is expressed in the same units as the measurements and has comparable meaning to a SD. The derived SEMs were then used to compare variabilities across readers

using paired or unpaired *t*-tests for intra- and inter-reader SEMs respectively. Agreement between modalities of severe hypertrophy rulings (MWT ≥30mm) was calculated using Cohen's Kappa. The Bland-Altman method was used to estimate the mean difference (bias) and 95% limits of agreement (±1.96 SD) between paired echocardiographic and CMR datasets. The percent variability(13) of MWT measurements per modality was calculated as the ratio of the difference in the measured MWT and the mean MWT for that dataset, to the mean MWT for that dataset (**Equation 1**). Percent variability was measured separately for echocardiographic and CMR datasets and their SD subsequently calculated.

Equation 1

$$Percent\ variability\ (\%) = \frac{Measured\ MWT - Mean\ MWT\ for\ that\ dataset}{Mean\ MWT\ for\ that\ dataset} \times 100$$

RESULTS

MWT measurements and practices. Echocardiographic and CMR SAX slice-matching was good (point-to-curve errors all <1mm; range 0.13–0.98, **Supplementary Table 3**). Morphological subtypes were: reverse curvature septum *n*=11; sigmoid septum *n*=3; neutral septum *n*=2; apical *n*=1. Seventy readers from 6 continents completed the analysis (~40 hours collective reading time); 50% were heads of department, service leads or international guideline committee members. Prior to caliper placement, readers rated 79% of echocardiography and 68% of CMR datasets in the 14-dataset analysis module assigned to them, as above or well-above average image quality. Following measurement, these were profoundly downgraded to 10% and 41% (**Table 1**). Overall, CMR measured the MWT thicker than echocardiography by 3.7 mm (and see Bland-Altman plots, **Supplementary Figure 3**).

MWT variability. –*View.* The majority of readers applied calipers to a combination of 'parasternal long axis/3-chamber + 4-chamber + SAX' views by echocardiography or CMR (57% and 64% respectively; **Supplementary Table 5**, **Figure 2** and **Supplementary Figure 4**). The least popular sole views calipered were echocardiography 4-chamber (1%) and CMR 3-chamber (3%).

–Phase selection. Echocardiography readers were split between those defining end-diastole at mitral valve closure and using the electrocardiographic R wave. CMR readers consistently took the time of mitral valve closure to indicate end-diastole. There were no cases of calipers deployed in the wrong (e.g. systolic) phase.

–Intra-modality inter-reader variability. The scatter of HCM MWT measurements between readers was broad for both modalities (**Figure 3**, mean SEMs [±95% CI] considering all imaging datasets by echocardiography and CMR: 4.63mm [2.22–7.05]; CMR 2.98mm [1.81–4.16]) with no significant difference between less and more experienced readers (echocardiography P=0.699; CMR P=0.239). The percent variability of echocardiographic MWT measurements ranged from –59% to 117% (SD ±20%) and for CMR MWT it ranged from –61% to 52% (SD ±11%).

–Inter-modality intra-reader variability *(paired echocardiography/CMR datasets).* There was great variability (high SEM) between MWT measurements made by the same reader by echocardiography and CMR (SEM 4.8mm [4.3–5.2mm]) and only modest correlation between modalities (r=0.56). Less and more experienced readers had the same variability statistically (SEMs respectively 4.6mm [4.0–5.3mm] *vs.* 4.4mm [3.6–5.2mm], P=0.100).

–Intra-modality intra-reader variability *(duplicate datasets, all readers).* For datasets presented twice to the same reader assessing self-disagreement, SEMs were similar for echocardiography and CMR: 2.1mm [1.7–2.6mm] *vs.* 2.2mm [1.7–2.7mm], *P*=0.821). Less experienced and senior readers had similar variability: echocardiography SEMs 2.0mm [1.4–2.7mm] *vs.* 2.1mm [1.3–2.8mm], *P*=0.699; CMR SEMs 2.7mm [1.8–3.6mm] *vs.* 2.0mm [1.2–2.8mm] *P*=0.239, **Supplementary Figure 5**).

–***Septomarginal trabeculum.*** The septomarginal trabeculum was included in MWT less often by echocardiography than CMR (19% *vs.* 43%, *P*=0.002; **Supplementary Table 5**). On echocardiography experienced readers included the septomarginal trabeculum more often than less experienced readers (24% *vs.* 13%, *P*=0.045). Septomarginal trabeculum inclusion on CMR was reader experience independent (40% *vs.* 50% respectively, *P*=0.155).

–Severe hypertrophy. Inter-reader scatter of MWT measurements increased visually with severe LVH (e.g. datasets 15–17 in **Figure 3**). Statistically, with increasing MWT the SD rose (correlation: echocardiography *r*=0.89, *P*<0.0001; CMR *r*=0.69; *P*=0.004; **Supplementary Figure 6**). This meant that the MWT measurement variability (1SD) of datasets with <24mm LVH (echocardiography *n*=8 and CMR *n*=7), was 3.2mm/2.2mm respectively, compared to 6.1mm/5.1mm for ≥24mm LVH (*n*=7 and *n*=8; *P*<0.0001 and *P*=0.006 respectively). There was only moderate intra-reader agreement for detecting severe LVH rulings (MWT ≥30mm) between modalities (Cohen's Kappa 0.5, 95% CI 0.16–0.85).

Impact of echocardiographic MWT variability on patient risk stratification. The risk simulation 4-center tertiary care HCM sample comprised 769 patients (500 male, 55±16 years) of whom 589 were low-risk, 81 intermediate-risk and 99 high-risk (**Table 2**). The

estimated ±1SD percent variability of echocardiographic MWT measurements (±20%) when applied to this risk simulation cohort using the ESC algorithm, would lead to risk re-stratification of 150 (19.5%) HCM patients: 102 patients (13%) convert to or out of, a low-risk category and 48 patients (6%) convert to or out of, a high-risk category. This consists of a +1SD change re-classifying 50 patients from low- to intermediate-risk, 17 from intermediate- to high-risk, and counterintuitively (MWT and risk have an inverse U-shaped relationship at some ages) downgrading risk in 1 (**Supplementary Figure 7**). A further 4 patients could no longer be risk scored by the ESC algorithm as the +1SD change inflates MWT to >35mm which the tool rejects. A –1SD change reclassifies 31 patients from high- to intermediate-risk, and 52 from intermediate- to low-risk. There were no category jumps. In summary therefore and within this cohort, a +1SD MWT change is potentially equivalent to 8.7% of patients being inappropriately recommended an ICD (risk upgraded to intermediate or high), while a –1SD MWT change is potentially equivalent to 6.8% of patients being left potentially unprotected by and ICD (risk downgraded to low).

DISCUSSION

We have shown that the measurement of MWT as an imaging biomarker in HCM, both by echocardiography and by CMR, is challenging because:

1) Measurement is unstandardized
2) There is high measurement variability within a reader, between readers and between modalities
3) Measurement is less reliable as hypertrophy increases
4) Measurement variability for echocardiography meant that, when these (best case) results were applied to a multi-center clinical cohort, inappropriate clinical decision making related to ICDs may have occurred in one in seven HCM patients

The datasets here were high quality (**Figure 3** and **Supplementary Movies**) and readers were experienced suggesting measurement error is likely higher in usual global practice. Agreement between echocardiography and CMR MWT measurement was poor (SEM ~5mm), as previously described(14–18). Phelan et al. also showed significant WT measurement differences between echocardiography and CMR, but counter to our findings, CMR measured thinner and there was no association between WT variability and degree of LVH(19). This may be due to differences in our populations and measurement approaches: Phelan et al. studied HCM patients that were undergoing surgical myectomy, only calipered the septum in 3- and 4-chamber views and ignored LVH in the lateral wall or apex.

MWT variability and practices uncovered by this study

We found considerable variability between modalities in terms of how readers used calipers to measure MWT, particularly in the views that were used whether the septomarginal trabeculum was included in the measurement or excluded. Although the inter-reader variability by CMR was less than echocardiography (1SD percent variability of ±11% *vs.* ±20%, respectively), CMR variability may be higher in clinical practice when a full stack of images rather than 3 LV SAX slices are present. The absence of an electrocardiographic trace alongside the CMR images potentially constrained CMR readers to caliper at end-diastole more consistently than by echocardiography. The fact that a significant number of readers substantially downgraded their image quality scores after analysis, could reflect how challenging they found it to caliper the HCM heart for MWT. As we have used an estimate of MWT variability derived from a small HCM cohort and applied it to a larger unmatched clinical cohort of patients, future larger systematic studies of this kind are needed to validate our findings. Yet. the MWT variability uncovered here raises concerns about the reliability of MWT cut-offs used for inclusion of patients in HCM clinical trials.

Challenges and solutions for MWT measurement

Many approaches for HCM MWT measurement have been published (**Supplementary Table 6**), but these are inconsistent and hard to follow. In spite of its limitations, MWT does reflect structural burden of disease in HCM and predict risk but these data suggest that there is work to do. The problem appears to be the lack of measurement standardization and education and not the modality. Possible solutions include the provision of more detailed modality-specific, authoritative, illustrated guidance on MWT measurement; demonstration of site measurement quality recommended by governing bodies; acquisition of a gapless 3-dimensional CMR dataset; and use of machine-based MWT measurement approaches. A machine learning-based MWT algorithm would have some advantages: first, it would have zero re-read variability on the same images; second, it could be trained to consistently exclude the septomarginal trabeculum and other confounding paraseptal structures; and third, it could avoid the human pitfall of trying to measure MWT off imaginary radial spokes emanating from an unreliable LV centrum. Although automated algorithms are objective, they lack higher-level executive function to consider other factors–humans typically do not just measure, they attempt to influence decisions with measurement.

Why MWT matters clinically

Of all the ESC algorithm risk factors(2), MWT was the only nonlinear predictor with an inverse-U shaped relationship to SCD risk. The interpretation was that SCD risk decreases in patients with extreme hypertrophy (≥35mm), although the same algorithm rightly cautions in this domain. As a consequence of this inverse-U shaped relationship, the publicly available ESC HCM risk stratification tool incorporates both the linear and quadratic terms for MWT to improve the prediction. However, it should be noted that linear and quadratic terms of

MWT will be correlated, which may potentially lead to wide standard errors for both coefficients, magnifying the impact of measurement error. Error increases with hypertrophy–a concern as it is typically these patients where ICD decisions are most needed–at the nadir of the inverse-U shaped curve(2), or near the 30mm cut-off(4,5). In one patient we observed the quadratic model delivering 'counterintuitive' downgrading of SCD risk with increasing MWT (but still below 35mm: going from 28mm to 34mm; **Supplementary Figure 7**). Another large retrospective study comparing MWT by 2-dimensional or M-mode echocardiography and CMR, found significant inter-modality discrepancy overall (median difference of 3mm) but especially for patients with ≥30mm of LVH(14). The American College of Cardiology/American Heart Association strategy for HCM SCD risk stratification(4) acknowledges uncertainty and measurement error relaxing the high-risk cut-off to additionally absorb any patient at the 28–29mm interface, as if that heart actually measured ≥30mm. Our findings are likely most relevant at the borderline (MWT between 25–35mm) so future imaging-modality specific HCM guidelines should focus on these more challenging phenotypes.

In the simulations applied to the multi-center clinical cohort, a relatively conservative echocardiographic MWT percent variability of 20% caused significant ICD recommendation changes in 15% of cases. Multi-disciplinary meeting consensus may help, as will serial measurement, but there is also the possibility of a "trapdoor function": measure enough times, and a patient may end up, through chance, with an ICD–an irreversible decision. On account of its high degree of variability, MWT should be considered with caution when facing irreversible decisions in clinical practice. There is the potential for harm with both deferring or delivering too early or inappropriately, an ICD(19).

Collectively these findings suggest that major room for improvement exists, that more robust imaging biomarkers of risk are needed, and that any one biomarker or measure cannot

completely encapsulate disease–this holds true for MWT. Although developing a systematic approach to foster more standardized measures is certainly important, we also have to recognize that MWT will just be one piece of the story. Given the challenges and variability of MWT measurement shown here, it would appear that reliance on any one feature for HCM risk stratification is to be discouraged and that we have to look beyond just MWT to diagnose and stratify disease. Furthermore, the issue of caliper-based measurement variability uncovered for the MWT exercise here, can probably be at least partly generalized to any LV wall thickness measurement (maximal or not) highlighting the scale of the problem.

Limitations

This study used "best-case" images. Real-world echocardiographic and CMR images may have higher variability. We only displayed 3 SAX cine slices for CMR. Intra-modality test-retest variability (i.e. with repeated acquisition of new images) was not captured by this study and overall intra-reader variability was ascertained for all readers on a limited number of paired datasets. Only a maximum of 30 minutes of readers' time could be expected so the number of paired datasets presented to readers had to be limited. The 70 readers at conferences might have over-stated their current active measurement skills and we did not limit recruitment to readers who made clinical measurements of HCM on both modalities equally. Echocardiography experts (51%) measured CMR and vice versa (13%) although performance via SEMs was the same for preferred *vs.* non-preferred modalities (respectively: 5.0mm [4.5–5.6mm] *vs.* 4.4mm [3.5–5.4mm], P=0.165). In simulations that employed the ESC algorithm, left atrial size measurement variability was not considered. No attempt was made to match the ESC SCD risk profiles of patients in the imaging dataset (n=17) to those in the risk simulation cohort (n=769), and although the 17-patient imaging dataset was not designed to be representative of the risk simulation cohort, the two cohorts were similar on

the basis of their average echocardiographic MWTs (21.9±2mm *vs* 19.3±3mm) and overall SCD risk profiles (imaging dataset: 3.4±2.9% [excluding 3 fully anonymized datasets without available clinical metadata] *vs*. clinical cohort: 3.3±2.8%). CMR data was not used for SCD risk estimation in simulations, as CMR data was not part of the original risk tool (CMR measured 3.7mm thicker than echocardiography so substitution without adjustment into the ESC algorithm is not acceptable). Due to time restrains only a few images could be shown to readers, which may have contributed to larger variability than expected.

CONCLUSION

The maximum wall thickness–a standard imaging biomarker acquired in almost all cardiac imaging tests–shows a high degree of inter-reader variability even under optimal conditions. The downstream consequences are quantifiable for echocardiography using the ESC 5-year risk of sudden cardiac death algorithm, where it may be leading to the misallocation of primary prevention ICDs in one in seven HCM patients. Better standardization of MWT measurement in HCM guidelines by current governing societies is needed to improve our clinical decision-making in patients with HCM.

CLINICAL PERSPECTIVES

Competency in patient care and procedural skills

The measurement of maximal wall thickness as an imaging biomarker in hypertrophic cardiomyopathy (HCM), both by echocardiography and cardiovascular magnetic resonance, is challenging because of unstandardized measurement approaches, high measurement variability and worsening reliability as hypertrophy increases.

The measurement variability for echocardiography estimated from this study, means that inappropriate clinical decision-making related to implantable cardioverter defibrillators may be occurring in one in seven HCM patients.

Translational outlook

In patients with HCM, the maximal wall thickness biomarker should be applied with caution as part of decision-making related to the implantation of cardioverter defibrillators.
Better measurement standardization efforts in recommendations by current governing societies are needed to improve our clinical decision-making in patients with HCM.

REFERENCES

1. Elliott P, Andersson B, Arbustini E, et al. Classification of the cardiomyopathies: A position statement from the European Society of Cardiology Working Group on Myocardial and Pericardial Diseases. *Eur Heart J* . 2008;29:270–6.
2. O'Mahony C, Jichi F, Pavlou M, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *Eur Heart J*. 2014;35:2010–20.
3. Elliott PM, Anastasakis A, Borger MA, et al. 2014 ESC Guidelines on diagnosis and management of hypertrophic cardiomyopathy. *Eur Heart J*. 2014;35:2733–79.
4. Maron MS, Rowin EJ, Wessler BS, et al. Enhanced American College of Cardiology/American Heart Association strategy for prevention of sudden cardiac death in high-risk patients with hypertrophic cardiomyopathy. *JAMA Cardiol*. 2019;4:644–657.
5. Gersh BJ, Maron BJ, Bonow RO, et al. 2011 ACCF/AHA guideline for the diagnosis and treatment of hypertrophic cardiomyopathy: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: developed in collaboration with the American American Association for Thoracic Surgery, American Society of Echocardiography, American Society of Nuclear Cardiology, Heart Failure Society of America, Heart Rhythm Society, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2011;58:e212-60.
6. O'Mahony C, Jichi F, Ommen SR, et al. International external validation study of the 2014 European Society of Cardiology guidelines on sudden cardiac death prevention in hypertrophic cardiomyopathy (EVIDENCE-HCM). *Circulation*. 2018;137:1015–1023.

7. Syed IS, Ommen SR, Breen JF, Tajik AJ. Hypertrophic cardiomyopathy: identification of morphological subtypes by echocardiography and cardiac magnetic resonance imaging. *JACC Cardiovasc Imaging*. 2008;1:377–379.

8. Maron MS. Clinical utility of cardiovascular magnetic resonance in hypertrophic cardiomyopathy. *J Cardiovasc Mag Res*. 2012;14:12–15.

9. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009; 2009;42:377–81.

10. Captur G, Stables RH, Kehoe D, Deanfield J, Moon JC. Why democratize bioinformatics? *BMJ Innov*. 2016; 2:166–71.

11. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

12. Eliasziw M, Young SL, Woodbury MG, Fryday-Field K. Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. *Phys Ther*. 1994;74:777–788.

13. McNitt-Gray MF, Kim GH, Zhao B, et al. Determining the variability of lesion size measurements from CT patient data sets acquired under “no change” conditions. *Transl Oncol*. 2015;8:55–64.

15. Hindieh W, Weissler-Snir A, Hammer H, Adler A, Rakowski H, Chan RH. Discrepant measurements of maximal left ventricular wall thickness between cardiac magnetic resonance imaging and echocardiography in patients with hypertrophic cardiomyopathy. *Circ Cardiovasc Imaging*. 2017;10:1–9.

14. Bois JP, Geske JB, Foley TA, Ommen SR, Pellikka PA. Comparison of maximal wall thickness in hypertrophic cardiomyopathy differs between magnetic resonance

imaging and transthoracic echocardiography. *Am J Cardiol*. 2017;119:643–650.

15. Webb J, Villa A, Bekri I, et al. Usefulness of cardiac magnetic resonance imaging to measure left ventricular wall thickness for determining risk scores for sudden cardiac death in patients with hypertrophic cardiomyopathy. *Am J Cardiol*. 2017;119:1450–1455.
16. Devlin AM, Moore NR, Ostman-Smith I. A comparison of MRI and echocardiography in hypertrophic cardiomyopathy. *Br J Radiol*. 1999;72:258–264.
17. Posma JL, Blanksma PK, van der Wall EE, Hamer HP, Mooyaart EL, Lie KI. Assessment of quantitative hypertrophy scores in hypertrophic cardiomyopathy: magnetic resonance imaging versus echocardiography. *Am Heart J*. 1996;132:1020–1027.
18. Phelan D, Sperry BW, Thavendiranathan P, et al. Comparison of ventricular septal measurements in hypertrophic cardiomyopathy patients who underwent surgical myectomy using multimodality imaging and implications for diagnosis and management. *Am J Cardiol*. 2017;119:1656–1662.
19. Olde Nordkamp LRA, Postema PG, Knops RE, et al. Implantable cardioverter-defibrillator harm in young patients with inherited arrhythmia syndromes: A systematic review and meta-analysis of inappropriate shocks and complications. *Hear Rhythm*. 2016;13:443–454.

FIGURE LEGENDS

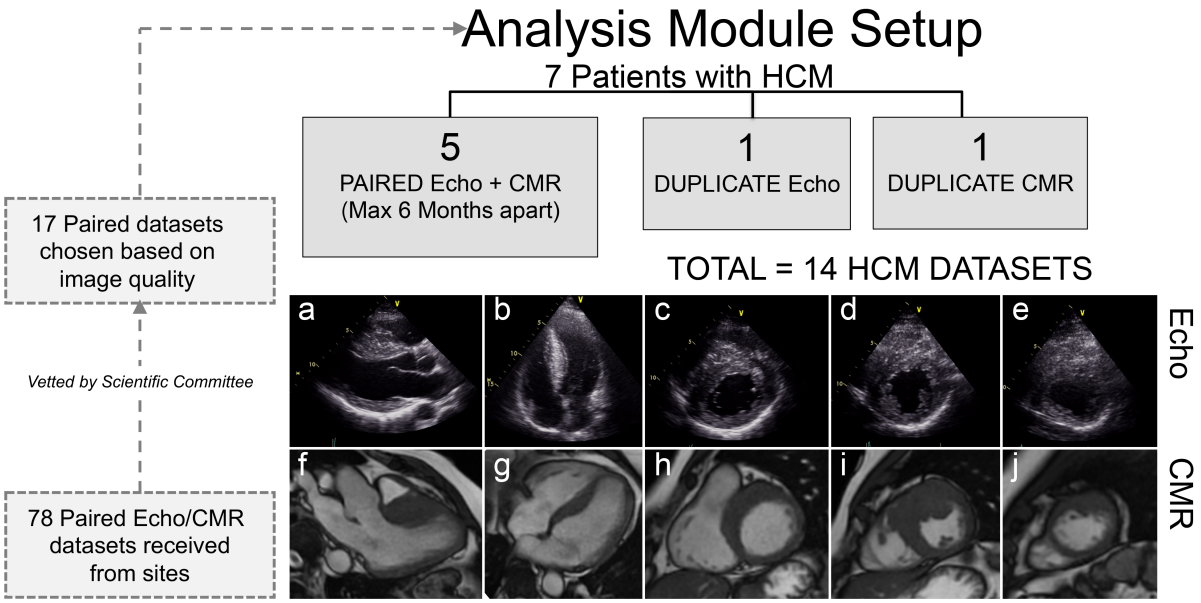


Figure 1. Organization of the analysis modules presented to readers. From the 78 submitted paired datasets, 17 paired datasets (from 17 patients, i.e. 34 datasets) were selected on the basis of image quality by the scientific committee. Out of these 34 datasets, each reader was presented with a total of 14 HCM datasets (50% echocardiography | 50% CMR). Of the 14 HCM datasets, 5 pairs were temporally matched echocardiography/CMR datasets from 5 different patients, 1 pair was a CMR duplicate from a different patient, and 1 pair was an echocardiography duplicate from another patient (7 patients in total per analysis module). Each echocardiography dataset was made up of 5 views (**a**, PLAX; **b**, A4C; **c**, bSAX; **d**, mSAX; **e**, aSAX) as was each CMR dataset (**f**, LVOT view; **g**, 4C; **h**, basal SAX; **i**, mid SAX; **j**, apical SAX). In this example echocardiography (**a**–**e**) and CMR images (**f**–**j**) belong to 2 different HCM patients. See also **Supplementary Movies 1** and **2**.
A4C, apical 4-chamber; bSAX, short-axis view basal ventricular level; CMR, cardiovascular magnetic resonance; Echo, echocardiography; HCM, hypertrophic cardiomyopathy; LVOT, left ventricular outflow tract; Max, maximum; PLAX, parasternal long axis; bSAX, short-axis view basal ventricular level; mSAX, short-axis view mid-ventricular level.

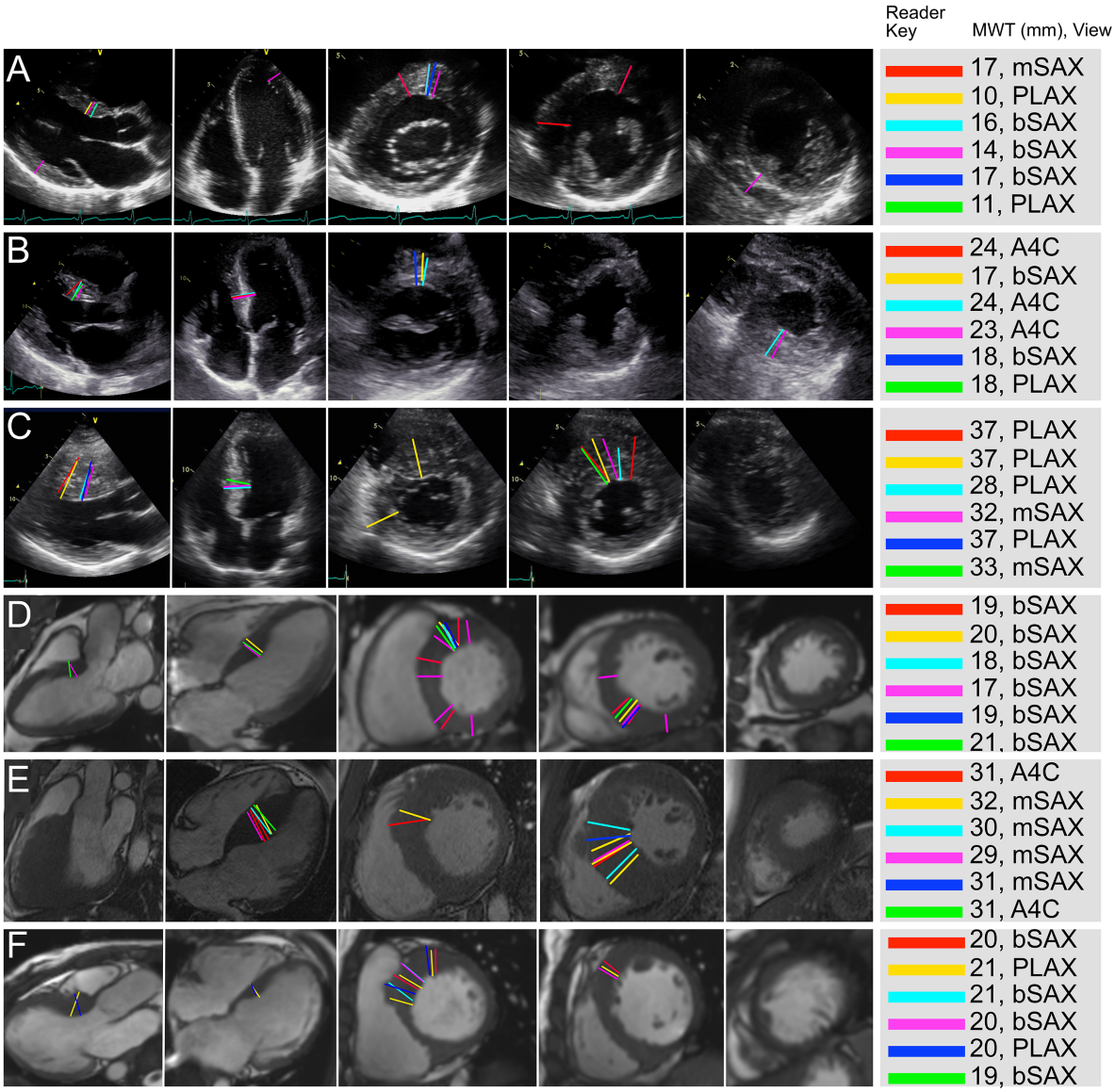


Figure 2. Superimposed examples of caliper measurements applied by a different set of randomly sampled readers to each of 6 randomly selected analysis modules (A–C echocardiography; D–F CMR). Caliper line colors indicate the different readers per module. Some readers applied a single MWT caliper whilst others applied multiple calipers per module (all calipers are shown here). Resultant MWTs for each reader and the view from which these were derived, are reported on the right.
MWT, maximal wall thickness. Other abbreviations as in ***Figure 1****.*

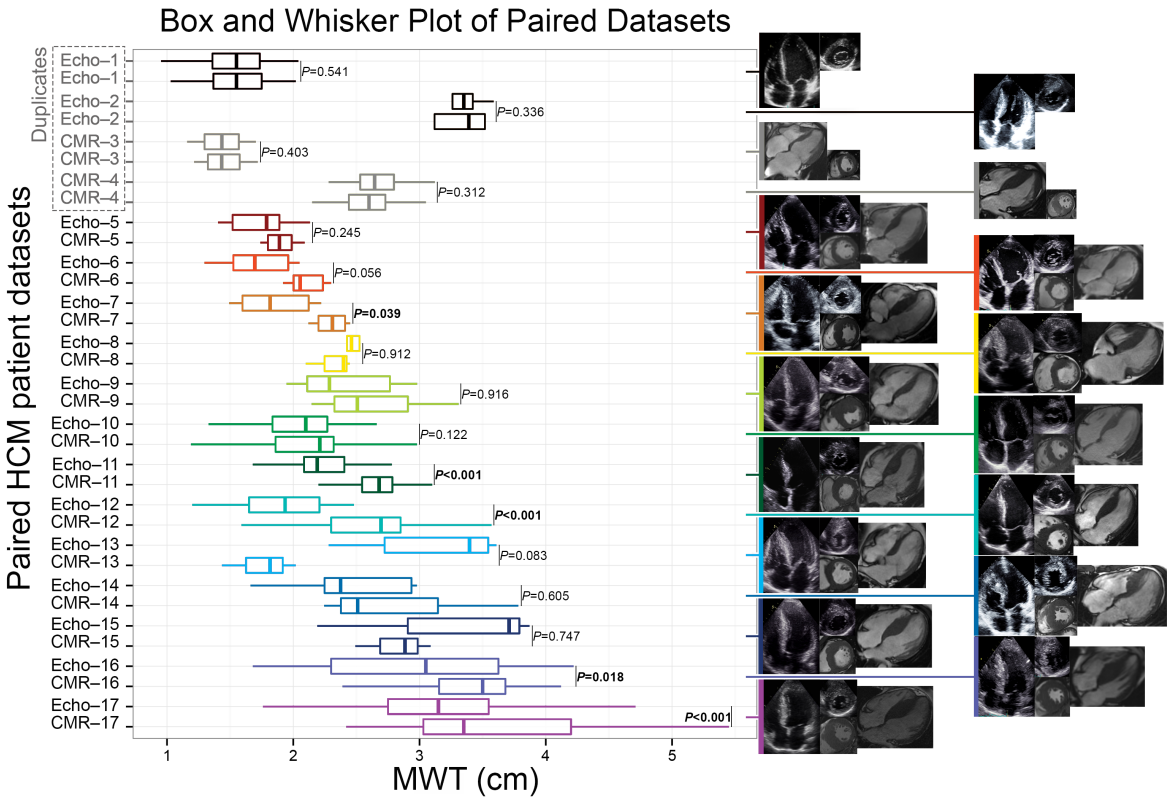
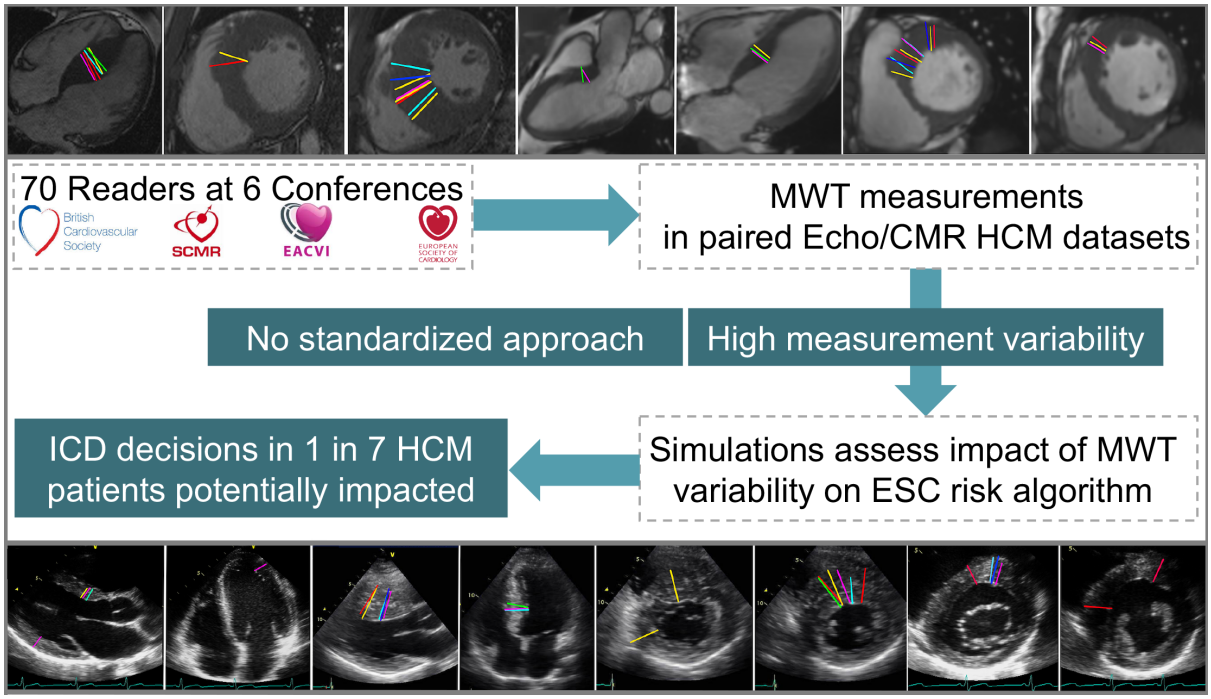


Figure 3. Box and whisker plots comparing measured MWT between pairs of HCM datasets from the same patient (echocardiography + CMR, same color) arranged from top to bottom, in ascending order of average measured MWT. The 8 datasets at the top represent the duplicate scans (both echocardiography [black], or both CMR [grey]) and provide a measure of intra-observer variability. Paired echocardiography/CMR 4C and SAX views (right panels) illustrate the HCM morphology of each patient. *P* values for measurement differences are shown for each dataset pair. MWT scatter noticeably broadens from top to bottom, that is as average measured MWT increases. Image quality is subjective so all images are shown here. Centerline = median (50th percentile); box = 25th/75th percentiles (IQR); whiskers = smallest/largest values within 1.5 times IQR below or above the 25th/75th percentiles.
Ap, apical; IQR, interquartile range. Other abbreviations as in ***Figures 1*** *and* ***2****.*



Central illustration. High variability of the maximal wall thickness biomarker in hypertrophic cardiomyopathy impacts on clinical care. Using best available images and experienced readers, the maximal wall thickness as a biomarker in hypertrophic cardiomyopathy has a high degree of inter-reader variability. The variability by echocardiography estimated from this study means that inappropriate clinical decision-making involving implantable cardioverter defibrillators, may occur in one in seven patients. Better measurement standardization efforts in recommendations by current governing societies are needed to improve our clinical decision-making in patients with hypertrophic cardiomyopathy.

Video Legends:

Video 1. Example-1 of analysis module presented to readers. Example movie showcasing the quality of echocardiographic and cardiovascular magnetic resonance datasets used in the analysis modules presented to readers.

Video 2. Example-2 of analysis module presented to readers. Example movie showcasing the quality of echocardiographic and cardiovascular magnetic resonance datasets used in the analysis modules presented to readers.

TABLES

Table 1. Professional profiles of participating readers (n = 70) and their image quality ratings before and after analysis.

Readers taking HCM measurements that matter for clinical care as part of routine work (%)			
Yes	100	No	0

Clinical measurements of HCM taken primarily on (%)	
Echo	51
CMR	13
Both equally	36

Sub-specialty* (%)	
Cardiomyopathy	39
Echocardiography	70
CMR	36
Other†	14

Professional role* (%)	
Guideline committee member	11
Head of department/Professor of Cardiology	24
Service lead	13
Consultant cardiologist	41
Cardiology senior specialist trainee‡	30

Years in main role (years)	
1–5 years	36
5–10 years	20
>10 years	44

Place of work (%)	
Europe	73
North America	6
South America	4
Australia	7
Asia	7
Africa	3

Confidence in measuring wall thickness in HCM based on clinical experience (%)		
	Echo	**CMR**
Below average	0	0
Average	47	41
Above average	34	33
Well above average	19	26

Echo image quality (%)	**Pre-analysis**	**Post-analysis**
Below average	4	20
Average	17	66
Above average	56	10
Well above average	23	0

CMR image quality (%)	**Pre-analysis**	**Post-analysis**
Below average	0	11
Average	32	48
Above average	39	41
Well above average	29	0

***** Sub-specialty and roles not mutually exclusive (readers selected all that applied).

† General cardiology, interventional cardiology, congenital heart disease, heart failure and imaging.

‡ Within 2 years of expected completion of cardiology specialist training.

CMR, cardiovascular magnetic resonance; HCM, hypertrophic cardiomyopathy.

Table 2. Multi-center pooled HCM patient data forming the basis for the simulations.

	Oxford, UK	London, UK	Pennsylvania, USA	Florence, Italy
Original Clinical Cohort (*n* = 769)				
Total *n*	101	110	72	486
Age, yrs	57 ± 13	50 ± 15	52 ± 16	57 ± 17
Male (%)	81 (80)	78 (71)	43 (60)	298 (61)
Echo MWT, mm	19 ± 3	18 ± 4	21 ± 5	19 ± 4
Echo LAd, mm	39 ± 7	45 ± 8	42 ± 7	45 ± 7
Echo LVOT Gradient, mmHg	14 ± 23	32 ± 40	61 ± 50	31 ± 35
NSVT (%)	24 (24)	30 (27)	18 (25)	104 (21)
Syncope (%)	13 (13)	3 (3)	8 (11)	103 (21)
FH SCD (%)	6 (6)	35 (32)	14 (19)	117 (24)
MWT ≥28mm (%)	2 (2)	3 (3)	6 (8)	24 (5)
SCD Risk Score, %	2.5 ± 1.9	3.6 ± 3.0	3.9 ± 3.8	3.3 ± 2.4
SCD Risk Categories* *Low* (%)	87 (86)	79 (72)	50 (69)	373 (77)
Intermediate (%)	6 (6)	12 (11)	14 (19)	49 (10)
High (%)	8 (8)	19 (17)	8 (11)	64 (13)
Simulated Cohort Data				
Total Simulated Risk Re-classifications† (%)	11 (11)	23 (21)	14 (19)	102 (21)
Simulated ↑ Risk Re-classifications†				
Total all groups	6	7	6	48
Low→Intermediate	6	6	6	32
Low→High	0	0	0	0
Intermediate→High	0	1	0	16
Magnitude of SCD Risk Score Change	0.23 [0.17–0.37, +1.19]	0.36 [0.24–0.58, +2.37]	0.37 [0.22–0.57, +1.47]	0.39 [0.25–0.57, +3.03]
Simulated ↓ Risk Re-classifications†				
Total all groups	5	16	8	54
High→Intermediate	2	9	1	19
High→Low	0	0	0	0
Intermediate→Low	3‡	7	7	35
Magnitude of SCD Risk Score Change	0.30 [0.25–0.53, –1.60]	0.48 [0.34–0.73, –2.65]	0.47 [0.34–0.67, –3.35]	0.46 [0.32–0.71, –3.63]

Data presented as counts (%), mean ± standard deviation or median [Q1–Q3 interquartile range | maximum], as appropriate.

*Categories are based on the 5-year risk of SCD estimated by the European Society of Cardiology HCM SCD risk algorithm and imputation of the original clinically available data that included echocardiographic MWT measurements.

†Net upward (↑) and downward (↓**)** patient re-classifications (reported as counts) of SCD risk categories, resulting from the simulation.

HCM, hypertrophic cardiomyopathy; PI, principal investigator, SCD, sudden cardiac death. Other abbreviation as in **Table 2**.

‡ Excludes one case of 'counterintuitive' downgrading of risk in spite of increasing MWT.