European Heart Journal - Cardiovascular Imaging Advance Access published August 18, 2015

European Heart Journal – Cardiovascular Imaging
doi:10.1093/ehjci/jev197

Two-dimensional speckle tracking echocardiography: standardization efforts based on synthetic ultrasound data

Jan D'hooge[1]*, Daniel Barbosa[1], Hang Gao[1], Piet Claus[1], David Prater[2], Jamie Hamilton[3], Peter Lysyansky[4], Yasuhiko Abe[5], Yasuhiro Ito[6], Helene Houle[7], Stefano Pedri[8], Rolf Baumann[9], James Thomas[10], and Luigi P. Badano[11], on behalf of the EACVI/ASE/Industry Task Force to Standardize Deformation Imaging

[1]Department of Cardiovascular Sciences, University of Leuven, Herestraat 49, Leuven 3000, Belgium; [2]Philips Medical Systems, Andover, MA, USA; [3]Epsilon Imaging, Ann Arbor, MI, USA; [4]GE Healthcare, Tel Aviv, Israel; [5]Toshiba Medical Systems Corporation, Otawara-shi, Tochigi, Japan; [6]Hitachi-Aloka Medical, Tokyo, Japan; [7]Siemens Ultrasound, Mountain View, CA, USA; [8]Esaote, Genoa, Italy; [9]TomTec Imaging Systems, Unterschleissheim, Germany; [10]Bluhm Cardiovascular Institute, Northwestern University, Chicago, IL, USA; and [11]University of Padova, Padova, Italy

Received 25 April 2015; accepted after revision 16 July 2015

Aims Speckle tracking echocardiography has already demonstrated its clinical potential. However, its use in routine practice is jeopardized by recent reports on high inter-vendor variability of the measurements. As such, the European Association of CardioVascular Imaging (EACVI) and the American Society of Echocardiography (ASE) set up a standardization task force, which was joined by all manufacturers of echocardiographic equipment as well as by companies offering software solutions only, with the ambition to tackle this problem by standardization and quality assurance (QA).

Methods and results In this study, a first step towards QA of all commercially available tracking solutions based on computer-generated ultrasound images is presented. The accuracy of the products was acceptable with relative errors below 10% and intra-vendor reproducibility within 5%.

Conclusion Whether these results can be extrapolated to the clinical setting is the topic of an ongoing study of the EACVI/ASE/Industry Task Force to standardize deformation imaging. This study was an important first step in the development of generally accepted tools for QA of speckle tracking echocardiography.

Keywords speckle tracking echocardiography • variability • standardization • cardiac function

Introduction

Speckle tracking echocardiography (STE) relies on the fact that image patterns arising from ultrasound interference with the myocardium remain rather stable under small tissue displacements/deformations.[1] As such, frame to frame tracking the motion of the speckles throughout the cardiac cycle allows recovering the motion of the underlying myocardium. Several approaches to estimate the motion of the speckles exist, but block-matching algorithms have become particularly popular.[2]

Over the past decade, the potential of STE to support clinical decision-making has been demonstrated in a variety of scenarios.[1] A limitation to widespread clinical application of this technique has been the reported suboptimal inter-vendor reproducibility of strain measurements obtained with STE.[3,4]

Some form of standardization thus seems to be mandatory to make STE a tool used in daily practice for clinical decision-making. As such, the European Association of CardioVascular Imaging (EACVI) and the American Society of Echocardiography (ASE) set up a standardization task force that was joined by all manufacturers of echocardiographic equipment as well as by companies offering software solutions only.[5,6] Although an over-regularization of any technology under development may limit or even completely stop further evolution, the lack of standards has clear negative consequences as well. A subtle balance thus needs to be found between 'regulations' to avoid unnecessary variability between products on

* Corresponding author. Tel: +32 16 349012; Fax: +32 16 344240, *E-mail address*: jan.dhooge@uzleuven.be

the one hand and 'R&D freedom' to enable further development of the technology on the other. Accordingly, the strategy proposed by the Task Force was one of the *quality assurance* (*QA*) that offers vendors/developers the required freedom of developing their algorithms in any possible way as long as they fulfil certain commonly accepted criteria on reliability and accuracy of the proposed methodology.

QA can be done at multiple levels where a trade-off needs to be made between the realism of the ultrasound data and the reliability of the reference measure (*Figure 1*). Biomechanical models in combination with software tools that mimic the ultrasound image formation process allow generating synthetic ultrasound images of the heart (i.e. obtained purely by computer simulation). In such an environment, the motion/deformation values to be measured are exactly known but, despite the quality of these models, the image appearance remains over-idealized and is not very realistic. At the other end of the spectrum, echocardiographic data can be obtained from patients; but as any imaging modality/measurement technique has its limitations and uncertainties, the reference values to be used as a benchmark are relatively unreliable. Finally, mock model setups and open-chest (large) animals preparations fall in between these two extreme cases (*Figure 1*).

From the above, it is clear that a solid QA framework requires testing at different levels, i.e. in computer-generated synthetic ultrasound images, in mock models, in (large) animal experiments, and/or the human setting. The present document reports on the first steps (i.e. computer-generated synthetic ultrasound images) taken towards QA by the EACVI/ASE/Industry Task Force to Standardize Deformation Imaging components. An initial in-vivo QA study has very recently been reported by the same task force.[7]

Methods

Simulation approach to generate synthetic ultrasound images

The left ventricle (LV) was geometrically modelled as a truncated ellipsoid and deformed using a previously described kinematic model.[8] As input to this model, experimentally recorded LV volume and torsion curves were used. Scattering sites were distributed at random positions within the LV wall at a density of 5 scatterers/mm^3. This scatterer density was used in order to be above the Rayleigh limit required to obtain fully developed speckle.[9] In addition, scattering sites were positioned at random positions over the whole volume of interest in order to mimic background noise. The amplitude of the scatterers mimicking noise could be varied with respect to the myocardial scattering sites in order to obtain data sets with different signal-to-noise ratios.

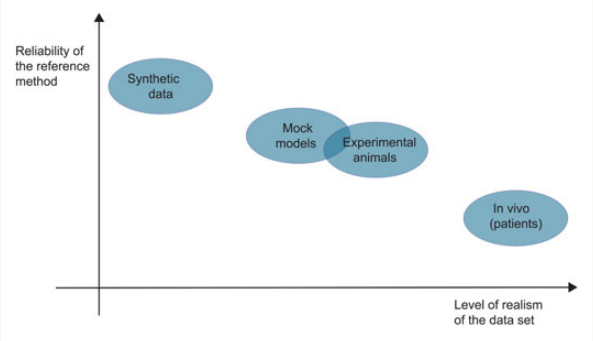

Figure 1 Schematic overview of the different setups available for QA purposes with their relative trade-offs between level of realism of the data and the reliability of the reference methodology available.

Ultrasound data were subsequently generated by convolving the ultrasound transmit pulse with a Dirac comb generated using a weighted projection of the individual scatterers on the image line as previously described.[10] By repeating the above process for all lines in an image, the radio-frequency (RF) data could be simulated. For each frame in the image sequence, the position of the scatterers was updated according to their motion profiles dictated by the kinematic model.

The above simulation approach was implemented in Matlab (MathWorks, Natick, MA) and could be run as a stand-alone application using a Graphical User Interface (*Figure 2*). In order to keep vendor-specific information confidential, parameters like line density, bandwidth of the transmit pulse, and others (cf. *Table 1*) could be entered upon runtime of the simulation program. As such, a common simulation platform could be used amongst all vendors without potentially disclosing confidential information while still allowing for vendor-specific modifications to the imaging system modelled by the simulator. The latter is particularly important as system-specific characteristics can have a direct influence on the performance of a given tracking technology.

The output of the simulator consisted of the RF data of an entire cineloop (in a Matlab format). Vendors were encouraged to reconstruct the B mode and associated DICOM images themselves in order to keep the image reconstruction processes (e.g. envelope detection, scan conversion, filtering) as close as possible to what is done on their ultrasound system. However, as this was not feasible for all vendors, the simulation program also provided DICOM sequences as output.

Data sets generated

Upon running the simulator, four different kinematic models were automatically created in order to mimic different clinical settings as summarized in *Table 2*. Moreover, in order to take the statistical nature of speckle into account, each model was generated three times. Indeed, as scattering sites are positioned at random positions inside the LV model, multiple realizations of this scattering model resulted in different scattering distributions and therefore distinct ultrasound interference patterns (i.e. speckle). Finally, in order to test the sensitivity of the different tracking solutions to noise, all models were generated at three different noise levels (i.e. 20, 40, and 60% in relative amplitude). In total, 36 (i.e. $4 \times 3 \times 3$) data sets were thus generated for every run of the program. All images simulated a 2D apical echocardiographic view.

Processing of the synthetic data

Application experts from the companies involved in this standardization effort processed all images according to the recommendations provided to their customers to obtain optimal tracking results. Typically, this involved manual/semi-automatic contouring of the endocardial and—for some products—epicardial border(s). All data sets were analysed three times in order to account for measurement variability. End-systolic global longitudinal strain values measured at endocardium were provided by all vendors in a spreadsheet format (Microsoft Excel). When the specific software packages allowed extracting mid-myocardial strain values, vendors were encouraged to provide these values as well.

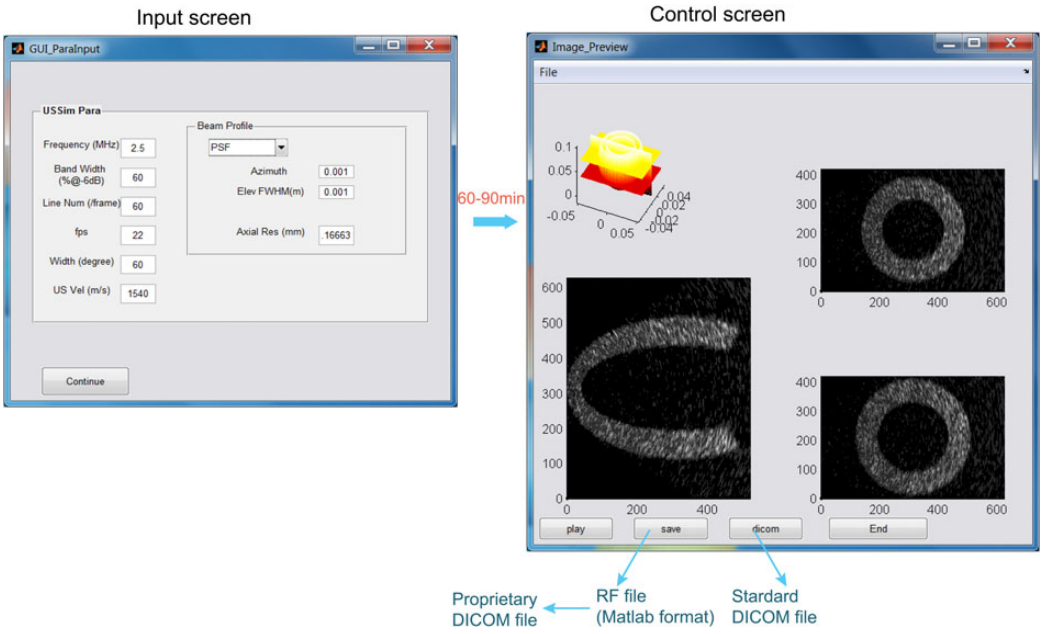

Figure 2 Screenshot of the graphic user interface used to simulate synthetic ultrasound data with an indication of the data output options.

Table 1 Overview of the system-specific parameters to be used in the ultrasound simulator

Transmit pulse	B-mode image	Physical constants
Center frequency (MHz)	Number of lines	Speed of sound (m/s)
−6 dB relative bandwidth (%)	Frame rate (Hz)	
	Sector angle (degrees)	

Table 2 Overview of the parameters used in the kinematic cardiac model in order to generate synthetic ultrasound data sets

Clinical model	Wall thickness (mm)	Ejection fraction (%)	Heart rate (bpm)
Normal	10	55	70
Dilatation	7	30	80
Hypertrophy	15	55	70
Exercise	10	65	150

Statistics

Both the average (over three realizations of each model and over the three analyses per realization) absolute (E_{abs}) and relative (E_{rel}) errors were defined per vendor and per clinical model for all noise levels or more explicitly:

$$E_{abs}\ (\%) = \frac{\sum_{i=1}^{3}\sum_{j=1}^{3}|\varepsilon_{ij} - \text{ref}|}{9}$$

$$E_{rel}\ (\%) = \frac{E_{abs}}{\text{ref}} \times 100$$

with ε_{ij} the jth measurement of the global longitudinal strain for realization i of the model and ref the ground truth reference value for that model (i.e. the one defined by the kinematic model).

The intra-vendor reproducibility of each measurement was assessed by calculating the coefficient of variation of the above-mentioned repeated measures with respect to the ground truth:

$$\text{Repr}\ (\%) = \frac{\sqrt{1/9\sum_{i=1}^{3}\sum_{j=1}^{3}(\varepsilon_{ij} - \mu)^2}}{\text{ref}} \times 100$$

with μ the mean measurement, i.e. $\sum_{i=1}^{3}\sum_{j=1}^{3}\varepsilon_{ij}/9$. Please note that this measure of reproducibility integrates variability due to repeated analyses of the same image (i.e. commonly referred to as intra-observer variability) as well as variability caused by different image characteristics in a new recording of the same subject (i.e. inter-study variability also referred to as test–retest variability).

In addition, as an upper limit of variability between (intra-vendor) observations, the maximal intra-exam difference (D_{max}) between the three analyses for the three realizations of a given model was determined and expressed relative to the ground truth reference strain as a percentage:

$$D_{max}\ (\%) = \frac{\max_i(D^i_{max})}{\text{ref}} \times 100$$

with D^i_{max} the maximal difference between two observations made for realization i of the model, i.e. $D^i_{max} = \max\{|\varepsilon_{i1} - \varepsilon_{i2}|, |\varepsilon_{i1} - \varepsilon_{i3}|, |\varepsilon_{i2} - \varepsilon_{i3}|\}$.

Finally, a test–retest analysis (i.e. inter-study variability measurement) was done by calculating the coefficient of variation of the three global strain values obtained for the three realizations of each model:

$$\mathrm{TRT}\ (\%) = \frac{\sqrt{1/3\sum_{i=1}^{3}(\mu_i - \mu)^2}}{\mathrm{ref}} \times 100$$

with μ_i the average measurement over three analyses of realization i of the model, i.e. $\mu_i = \sum_{j=1}^{3} \varepsilon_{ij}/3$. Important to note is that inter-vendor reproducibility and repeatability tests were not part of the current study protocol as each vendor performed the measurements only on its own-generated synthetic data sets.

Results

Table 3 lists the vendors that participated in this first QA test as well as the software packages used to measure global longitudinal strain. An example of a synthetic image of a normal and a hypertrophic LV model at high noise level (i.e. 60% relative amplitude) is given in *Figure 3*. The characteristic speckle pattern is clearly observed within the myocardium, but typical imaging artefacts and surrounding structures are absent. A movie of the cineloop is provided in the online appendix (Movie 1).

The mean absolute and relative errors for all ventricular models, image noise levels, and software products are given in *Figure 4*. Vendors 'A' and 'F' underestimated the global longitudinal strain for all models and noise levels. Others showed a more variable behaviour with either under- or overestimation depending on the morphology or noise level (e.g. vendors 'C', 'D', 'E', and 'G'), while yet others systematically overestimated (e.g. vendors 'B', 'H', and 'I' when not considering the exercise model of the highest noise level). Please note that 'over-' and 'underestimate' was used here with respect to the absolute magnitude of the strain value implying that a more negative value than the one expected is considered an 'overestimation'. Consistently amongst all products was the fact that global longitudinal strain was underestimated for the exercise model with the exception of vendors 'B' and 'G' at low and intermediate noise levels.

Table 3 Overview of the participating companies including the type of data they processed and the software version used for analysis

Vendor	Identifier	Transducer simulated	Frame rate simulated (Hz)	Software package and version number used	Proprietary (DICOM) file generated from the RF data	Default DICOM file generated by the simulator
Hitachi-Aloka	A	Default	60	2D Tissue Tracking Analysis v3.0	X	
TomTec	B	Default	Default	TomTec- Arena 1		X
Philips	C	S5-1	55	QLab 9.0		X
Esaote	D	SP2430	73	Sw Rel 4.xx		X[a]
General Electrics	E	M5S-D	62 rest 84 stress	EchoPac v113.0.0	X	
Toshiba	F	Default	55	ACP-V3.0	X[b]	
Siemens	G	8V3c	80	VVI 3.5		
Siemens	H	4V1c	80	VVI 3.5		X
Epsilon Imaging	I	Default	90	EchoInsight v2.1		X

[a]DICOM tags required by this software package (i.e. RWave information) were inserted manually by this vendor.
[b]No DICOM file but rather a proprietary raw data file was generated from the RF data by this vendor.

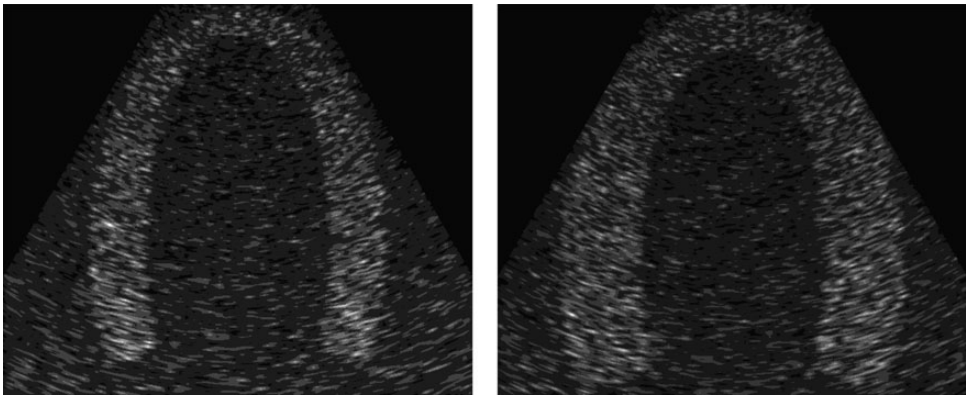

Figure 3 Example synthetic data sets of a normal physiology (left) and a hypertrophic model (right) at high (i.e. 60%) relative noise level.

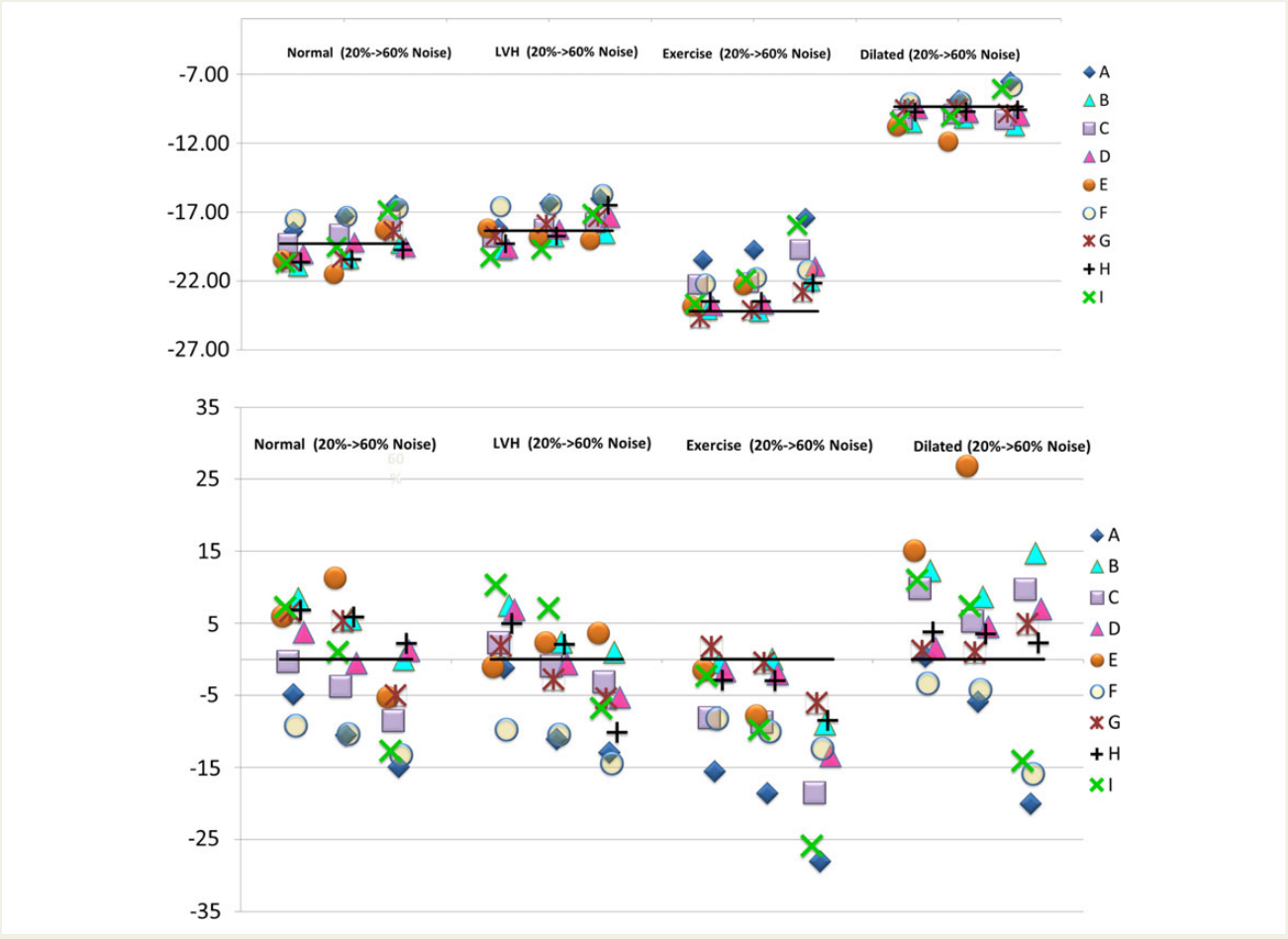

Figure 4 Absolute (top) and relative (bottom) errors in percentage of the averaged measurement of global longitudinal strain for different cardiac models and for different noise levels. The horizontal lines represent the reference ground truth strain value.

Similarly, strain was consistently overestimated in the dilated model for all vendors except vendors 'A' and 'F'. Finally, more noise was consistently associated with a reduction in the measured value. As such, (some) noise could bring the measured value closer to the reference value for the tools that had a tendency to overestimate strain.

Despite this variable bias in the strain estimates, the relative error of any of the software packages remained within a range of about 10% for the normal and hypertrophied models when the noise level was low to intermediate (i.e. 20–40%). The same relative accuracy was obtained for the exercise model with the exception of vendor 'A' and the highest noise level where relative errors up to 28% occurred. For the dilated model, the relative error increased to about 15% for most cases, although larger deviations were observed at higher noise levels and—for vendor 'E'—at the intermediate noise level.

The intra-vendor reproducibility of the measurements of global longitudinal strain for each model, noise level, and vendor is shown in *Figure 5*. Although there are clear differences in variability amongst vendors, at the lowest noise level, all measurements were done with a coefficient of variation below 6% with the exception of vendor 'D' that showed slightly higher coefficient of variation for the exercise and dilated models. In fact, most of the measurements stayed below 5%. Overall, vendor 'D' had higher variation than most of the other vendors with coefficients of variation typically between 10 and 15%. Measurement variation increased with increasing noise level for most vendors. However, the impact of noise on intra-vendor reproducibility of the measurement seemed to be very variable between vendors. Some vendors had very little influence from noise (e.g. vendors 'A' and 'F'), while others showed significantly higher variability in noisy datasets (e.g. vendors 'E' and 'I'). Although most vendors could reproduce their measurements within 5% for the dilated model, one vendor (i.e. vendor 'H') clearly showed larger variability in this particular model (i.e. coefficients of variation 5.5 and 7.4% at the low and intermediate noise levels, respectively). Finally, *Figure 6* summarizes the largest difference measured between analyses of a given data set and reconfirms previous intra-vendor reproducibility findings, as do the test–retest values presented in *Figure 7*.

Discussion

Given the inter-vendor variability in strain measurements reported in some clinical studies, the EACVI and ASE set up a task force to work towards a standardization of speckle tracking echocardiography.

The results of the present study on synthetic ultrasound images confirmed that the accuracy of the different software packages in

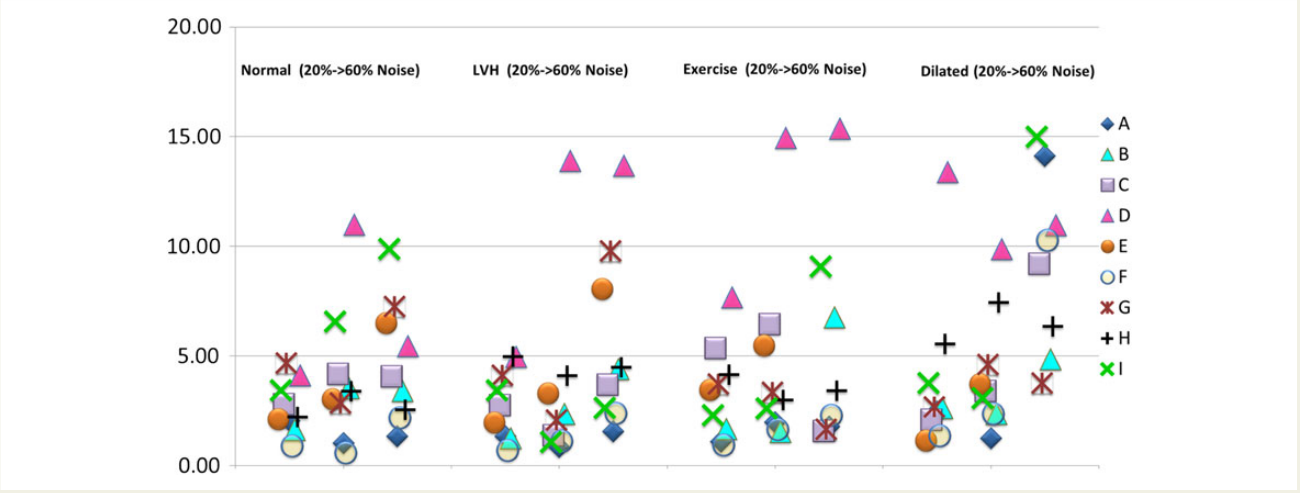

Figure 5 Coefficient of variation of the nine measurements (three analyses on three realizations of the model) in percentage for each model and noise level.

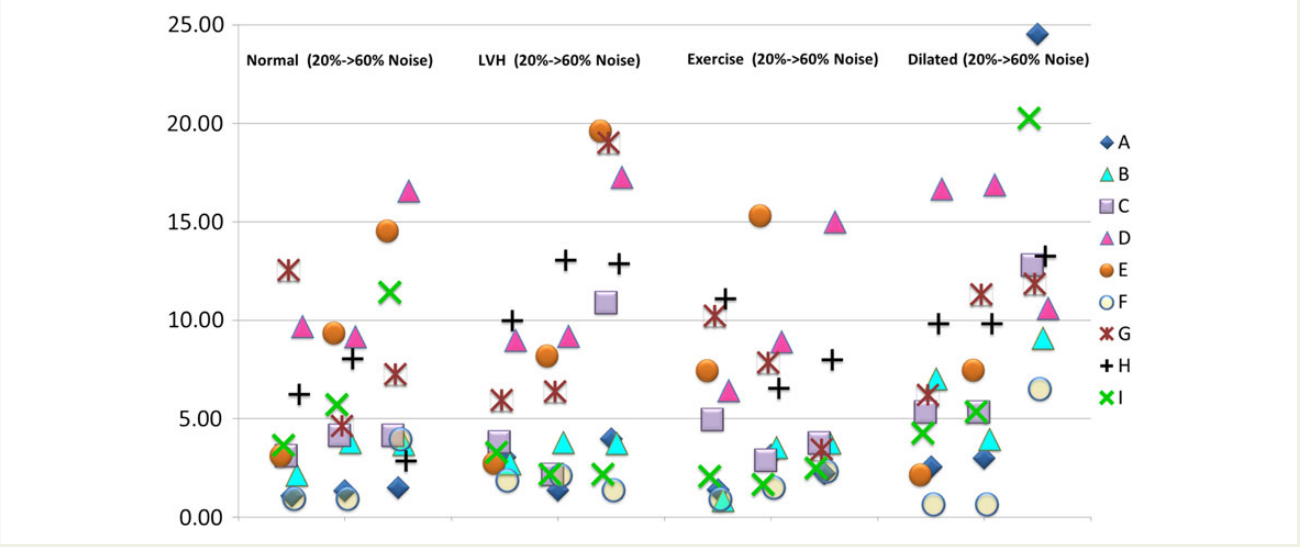

Figure 6 Maximal difference in percentage between repeated measures on a given model related to the true value to be measured.

measuring global longitudinal strain varied amongst vendors. However, importantly, this study showed that the relative accuracy for all vendors remained within 10%. This means that the accuracy of commercially available software packages to measure global longitudinal strain was in line with other standard measurements currently used in clinical echocardiography (e.g. left ventricular ejection fraction by 2D echocardiography).

The differences amongst vendors may at first appear unnatural as all products are developed to measure the same physical quantity. However, although all vendors use the speckle tracking principle (i.e. the fact that grey patterns within the image do not change—at least to some extent—from one frame to the next), many design choices have to be made in order to make such an algorithm working in the clinical arena. For example, how large will the tracking kernel be taken? Will the image be pre-processed/filtered prior to tracking? Which metric is used to quantify correspondence between patterns? In addition, 'strain' can be defined and calculated in several ways, and different vendors/software packages may have chosen different definitions. For example, will endo-, mid-, epicardial, or transmural strain values be reported? How will end-diastole—as the beginning of the strain integration period—be defined? Are end-systolic, peak systolic, or peak values reported?

The latter sources of possible differences amongst software packages (i.e. the ones related to strain definitions) were investigated by the Task Force and led to a consensus document.[11,12] As such, this consensus report could help in reducing the variability of the results presented in the current manuscript. The investigation of the effect of the standard definitions included in the consensus

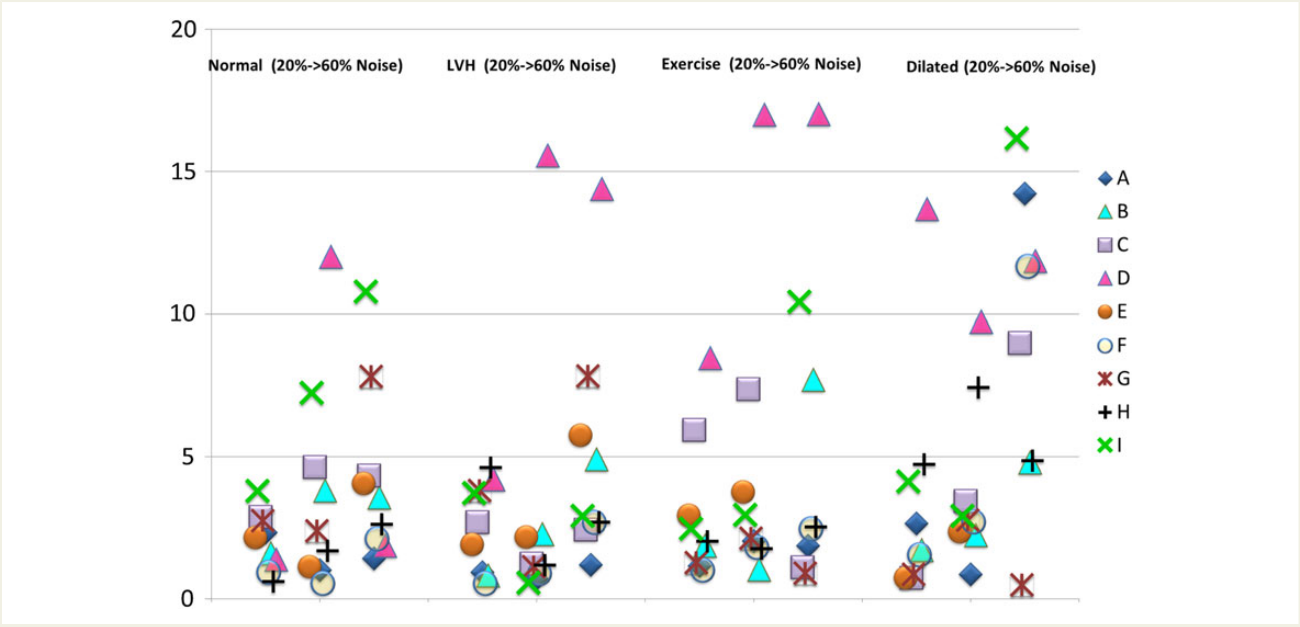

Figure 7 Test–retest coefficient of variation in percentage for all models at all noise levels for each of the vendors.

document on the inter-vendor variability, as assessed through synthetic data, is the topic of ongoing activities of the Task Force.

This study was designed to perform a part of the QA tests on the available commercial software packages and to build a snapshot of the current inter-vendor reproducibility of measurements. As stated above, this study showed that the relative accuracy for all vendors remains within 10%. Direct extrapolation of these results to 2D images obtained in real patients would imply that, for a given patient, the absolute value of the global longitudinal strain measured by two different software packages could be as high as ~4–5% (i.e. 20% of a normal strain value of approximately −21%). In other words, in this patient, the value measured by one vendor could be −22%, while another vendor may measure −17%. Of course, this example is a worst-case scenario but yet should be considered.

Although it is difficult to put an upper limit on the maximal variability allowed, the consensus within the Task Force was that a relative error below 10% would have been acceptable. The relative errors reported here were obtained as the average over nine measurements (i.e. three measurements on three different realizations of the models). Individual measurements (as typically done clinically) may thus deviate more as shown by the maximal difference between observations (*Figure 6*) that represents the maximal difference in relative error between observations.

Noise had a clear impact on the accuracy of global longitudinal strain for all vendors. As expected, accuracy decreased with increasing noise levels since image quality remains of paramount importance in measuring myocardial strain with speckle tracking echocardiography. The better the image quality, the more accurate the subsequent (automated) measurements will be. However, we should acknowledge that the highest noise level included in this study (i.e. 60% in amplitude) is likely unrealistically high and would normally not be encountered in clinical recordings. Therefore, despite the fact that this noise level was included as an upper bound, the findings at this noise level should be interpreted cautiously and should not be extrapolated to the clinical setting. In this context, the threshold of 10% relative error set out by the Task Force seems justified as a noise-free setting using synthetic data might otherwise call for a more restrictive error margin.

The impact of noise on the measurement accuracy of global longitudinal strain varied significantly amongst vendors with some algorithms being more sensitive to noise (e.g. vendors 'E' and 'I') than others (e.g. 'A' and 'F'). This might be related to the way the initial speckle tracking results are being filtered in space and/or time, which could impact space-time resolution of the derived strain information and thereby the ability to detect local dysfunctional areas. However, as only global longitudinal strain was assessed in this study, this potential trade-off between noise sensitivity and spatial resolution could not be tested. This will need to be addressed in follow-up studies.

Consistently amongst all vendors was the fact that high heart rate (i.e. 150 bpm) at a conventional frame rate of ~70 Hz resulted in underestimation of the global longitudinal strain. Again, this is not surprising, as higher heart rate at a fixed frame rate will result in more speckle de-correlation between frames making accurate tracking more challenging. These findings are consistent with the clinical experience that accuracy of STE is only moderate during exercise or dobutamine stress echocardiography. It is likely that better results could be obtained if the frame rate could be appropriately adjusted (i.e. in order to keep the heart rate to frame rate ratio constant), although this remains to be proven as higher frame rates typically come at the expense of lower spatial resolution and/or signal-to-noise ratio of the image. It is important to acknowledge that some software products are not released for the stress echocardiography setting.

Similarly, most vendors underestimated the true strain for the dilated, thin-walled model, although the relative error remained below 10%. The exact reason for the systematic underestimation of global longitudinal strain in this model remains unclear, but it could be related to the thickness of the wall making the tracking solution potentially more sensitive to the endocardial border identification (i.e. the segmentation process). Interestingly, some software packages overestimated longitudinal strain of the dilated, thin-walled model, while the same algorithms underestimated the longitudinal strain of the exercise model. This behaviour of the software package may suggest that they use a given prior value that somehow biases the motion estimation towards a normal value.

Taken all together, in the particular setting of synthetic data tested in our study, vendors 'D', 'G', and 'H' showed better accuracy with relative errors below 5% for almost all noise levels and models.

Next to measurement accuracy, measurement reproducibility is important. Overall, intra-vendor reproducibility was fairly good with coefficients of variation below 5% for most vendors. This would imply that a normal global longitudinal strain value of −21% could be reproduced within a range from about −20 to −22%. However, variability in the longitudinal strain measurements was significantly different amongst vendors with some of them being very robust to noise while others were very noise sensitive. Interestingly, the two vendors that systematically underestimated global strain values were very reproducible and very robust to noise. Likely, these vendors use a strong prior on 'normal deformation' that is integrated in their measurements. Similarly, the vendors that showed the best accuracy on average (i.e. 'D', 'G', and 'H') showed a higher coefficient of variation amongst measurements, which might suggest that these algorithms have a strong emphasis on the measured data and regularize less strongly. Overall, a subtle balance thus seems to exist between the amount of regularization imposed by the algorithm and the accuracy and the reproducibility of the resulting measurement.

Finally, the results reported in this study refer to measurements of the global longitudinal strain taken at the endocardium. The reason why the Task Force choose to work with endocardial longitudinal strain measurements is purely pragmatic since only this measurement could be provided by all vendors and could therefore easily be used to compare the various software packages. However, some vendors can also provide mid-myocardial strain values. As there is currently no evidence about the relative clinical value of endo- and mid-myocardial strain values, the Task Force currently proposes to have either measured and asked vendors to clearly report which value will be measured. For completeness, the absolute and relative accuracies of the mid-myocardial values obtained by three vendors are also reported in *Figure 8*.

Study limitations

The findings of this study should be interpreted taking into account the limitations of our methodology. Obviously, the synthetic data sets used in this study do not represent the true clinical scenario. Indeed, the kinematic model used to mimic ventricular motion/deformation during the cardiac cycle is relatively simple and does neither represent the true physiologic contraction nor the anatomical structures like the mitral valve or papillary muscles. Similarly, the ultrasound simulator does model the acoustic interferences leading to speckle patterns in the image but does not mimic typical artefacts such as acoustic shadowing, attenuation, dropouts, and reverberations that are commonly encountered during clinical scanning. Despite these disadvantages, synthetic data remain useful, since the exact reference values to be measured are available.

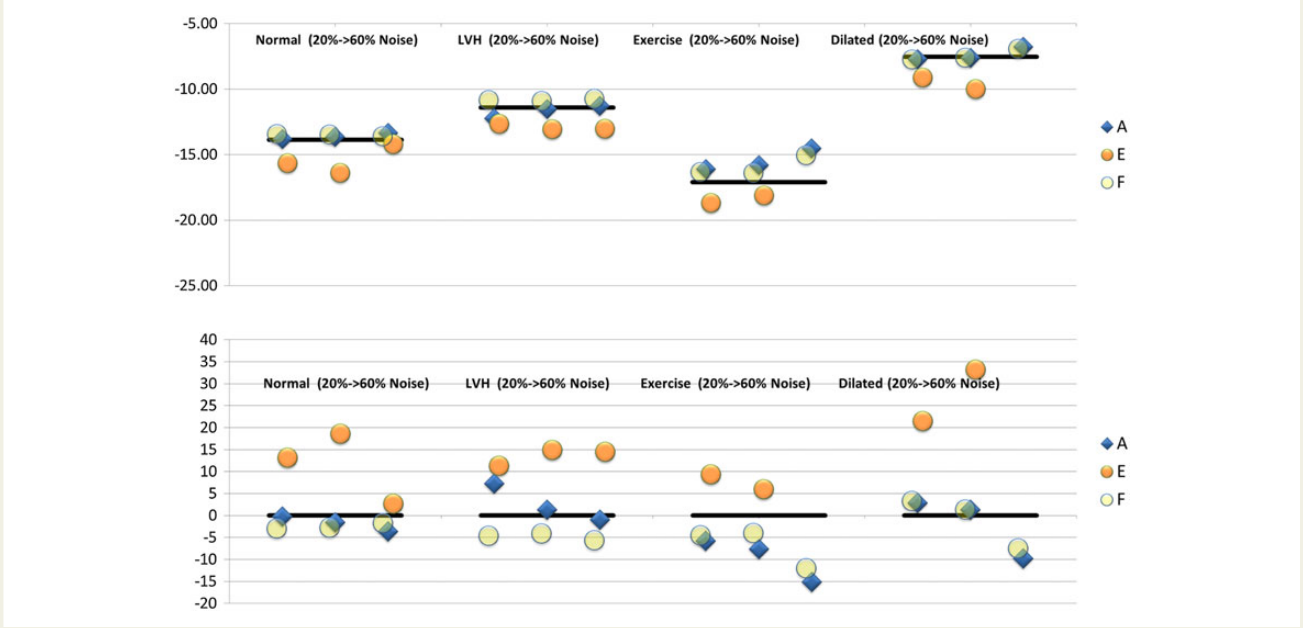

Figure 8 As for *Figure 4* but reporting the mid-myocardial strain values instead of the endocardial ones. Only three software products provided these values and are therefore reported here.

Whether or not these findings can be extrapolated to the clinical setting remains the topic of further investigations and separate studies. It is worth to note that the extrapolation of these results to the clinical scenario can be vendor dependent (i.e. some products may behave similar on synthetic data and clinical data, while others may perform quite differently).

Further developments of our algorithm have recently been made that allow generating more realistic synthetic ultrasound data sets.[13,14] This new methodology will be used in the next months to repeat the present study to further elucidate on the quality of different STE software packages under more clinical (but still synthetic) imaging conditions including regional strain parameters, timing parameters, and circumferential measurements. Similarly, (some of) the speckle tracking solutions tested have—in the meanwhile—further been developed. Therefore, these initial results may not reflect the actual status. However, the improvements in tracking algorithms towards better accuracy and robustness reflect the positive outcome of this study and overall the positive role of the Standardization Task Force activity. This was recently more explicitly demonstrated by comparing strain values of two different vendors over time, i.e. software versions.[15,16]

Additionally, although intra-vendor variability measures were made and reported, a direct inter-vendor variability study was not part of the protocol of the current study.

Finally, analyses using any of the software products require user interaction, e.g. to identify the region of interest or a decision to re-track when the initial tracking solution does not seem adequate. The fact that different persons analysed the data generated for different software products could have induced bias as not all operators may have the same level of experience. However, as the analyses were done by application specialists of the different vendors, it is likely that both tracking and analyses were performed in the most ideal conditions and the effect of the inter-operator variability was minimal.

Conclusions

This is the first study in which all currently commercially available 2D speckle tracking software packages have been tested on synthetic ultrasound data. It was an important first step in the development of generally accepted tools for QA of speckle tracking echocardiography. The accuracy of the current available products was acceptable with relative errors below 10% and intra-vendor reproducibility within 5%. Whether these results can be extrapolated to the clinical setting is the topic of an ongoing study of the EACVI/ASE/Industry Task Force to standardize deformation imaging.

Conflict of interest: None declared.

Funding

This work was funded in part by the American Society on Echocardiography and the European Association on CardioVascular Imaging.

References

1. Marwick TH, Yu CM, Sun JP. Myocardial Imaging: Tissue Doppler and Speckle Tracking. Oxford, UK: Wiley-Blackwell; 2008.
2. Jasaityte R, Heyde B, D'hooge J. Current state of three-dimensional myocardial strain estimation using echocardiography. *J Am Soc Echocardiogr* 2013; **26**:15–28.
3. Koopman LP, Slorach C, Hui W, Manlhiot C, McCrindle BW, Friedberg MK *et al.* Comparison between different speckle tracking and color tissue Doppler techniques to measure global and regional myocardial deformation in children. *J Am Soc Echocardiogr* 2010;**23**:919–28.
4. Takigiku K, Takeuchi M, Izumi C, Yuda S, Sakata K, Ohte N *et al.* Normal range of left ventricular 2-dimensional strain: Japanese ultrasound speckle tracking of the left ventricle (JUSTICE) study. *Circ J* 2012;**76**:2623–32.
5. Thomas JD, Badano LP. EACVI-ASE-industry initiative to standardize deformation imaging. A brief update from the co-chairs. *Eur Heart J Cardiovasc Imaging* 2013;**14**: 1039–40.
6. Thomas JD, Badano LP. EACVI-ASE-industry initiative to standardize deformation imaging. A brief update from the co-chairs. *J Am Soc Echocardiogr* 2013;**26**:21A–2.
7. Farsalinos KE, Daraban AM, Unlu S, Thomas SD, Badano LP, Voigt JU. Head-to-head comparison of global longitudinal strain measurements among nine different vendors. *J Am Soc Echocardiogr* 2015; doi: 10.1016/j.echo.2015.06.011.
8. Arts T, Hunter WC, Douglas A, Muijtjens AM, Reneman RS. Description of the deformation of the left ventricle by a kinematic model. *J. Biomech* 1992;**25**:1119–27.
9. Oosterveld BJ, Thijssen JM, Verhoef WA. Texture of B-mode echograms: 3-D simulations and experiments of the effects of diffraction and scatterer density. *Ultrason Imaging* 1985;**7**:142–60.
10. Gao H, Choi HF, Claus P, Boonen S, Jaecques S, van Lenthe GH *et al.* A fast convolution-based methodology to simulate 2-D/3-D cardiac ultrasound images. *IEEE Transactions on Ultrasonics, Ferro-Electrics and Frequency Control* 2009; **56**:404–9.
11. 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.
12. 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. *J Am Soc Echocardiogr* 2015;**28**:183–93.
13. De Craene M, Alessandrini M, Allain P, Marchesseau S, Wächter-Stehle I, Weese J *et al.* Generation of ultra-realistic synthetic echocardiographic sequences. In: *Conference: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* 2014, pp. 73–6.
14. Alessandrini M, De Craene M, Bernard O, Giffard-Roisin S, Allain P, Waechter-Stehle I. A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: methodology and open-access database. *IEEE Trans Med Imaging* 2015;**34**:1436–51
15. 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.
16. Thomas JD. Strain imaging in echocardiography: converging on congruence? *J Am Soc Echocardiogr* 2015;**28**:649–51.