Use of an automatic application for wall motion classification based on longitudinal strain: is it affected by operator expertise in echocardiography? A multicentre study by the Israeli Echocardiography Research Group

David S. Blondheim[1]*, Zvi Friedman[2], Peter Lysyansky[2], Rafael Kuperstein[3], Ilan Hay[3], Micha S. Feinberg[3], Ronen Beeri[4], Mordehay Vaturi[5], Alik Sagie[5], Sarah Shimoni[6], Wolfgang Fehske[7], Lisa Deutsch[2], Marina Leitman[8], Dan Gilon[4], Yoram Agmon[9], Yossi Tsadok[10], David Rosenmann[11], and Noah Liel-Cohen[12]

[1]Non-invasive Cardiology Unit, Hillel-Yaffe Medical Center, Hadera, Israel; [2]General Electric, Healthcare, Haifa, Israel; [3]Non-invasive Cardiology Unit, Sheba Medical Center, Tel Aviv, Israel; [4]Echocardiography Unit, Heart Institute, Hadassah Hebrew University Medical Center, Jerusalem, Israel; [5]Echocardiography Unit and Valvular Clinic, Department of Cardiology, Rabin Medical Center, Beilinson Hospital, Petah Tikva, Israel; [6]Non-invasive Cardiology, Kaplan Medical Center, Rehovot, Israel; [7]Department of Cardiology, St Vinzenz Hospital, Cologne, Germany; [8]Department of Cardiology, Assaf Harofeh Medical Center, Zerifin and Sackler School of Medicine, Israel; [9]Non-invasive Cardiology Unit, Rambam Medical Center, Haifa, Israel; [10]Ben-Gurion University, Beer Sheva, Israel; [11]Echocardiography Unit, Shaare Zedek Medical Center, Jerusalem, Israel; and [12]Echocardiography Unit, The Cardiology Department, Soroka University Medical Center, Beer Sheva, Israel

Aims	Assessing the quality of wall motion (WM) on echocardiograms remains a challenge. Previously, we validated an automated application used by experienced echocardiographers for WM classification based on longitudinal two-dimensional (2D) strain. The aim of this study was to show that the use of this automatic application was independent of the user's experience.
Methods and results	We compared the WM classifications obtained by the application when used by 12 highly experienced readers (Exp-R) vs. 11 inexperienced readers (InExp-R). Both classifications were compared with expert consensus classifications using the standard visual method. Digitized clips of cardiac cycles from three apical views in 105 patients were used for these analyses. Reproducibility of both groups was high (overall intra-class correlation coefficient: InExp-R = 0.89, Exp-R = 0.83); the lowest was noted for hypokinetic segments (InExp-R = 0.79, Exp-R = 0.72). InExp-R scores were concordant with Exp-R mode scores in 88.8% of segments; they were overestimated in 5.8% and underestimated in 3.2%. The sensitivity, specificity, and accuracy of InExp-R vs. Exp-R for classifying segments as normal/abnormal were identical (87, 85, and 86%, respectively).
Conclusion	Classification of WM from apical views with an automatic application based on longitudinal 2D strain by InExp-R vs. Exp-R was similar to visual classification by Exp-R. This application may be useful for inexperienced echocardiographers/technicians and may serve as an automated 'second opinion' for experienced echocardiographers.
Keywords	Automated scoring • Echocardiography • Segmental strain • Wall motion analysis • Speckle tracking

Introduction

Evaluation of segmental left ventricular (LV) wall motion (WM), one of the most difficult faucets of echocardiography, is still done only qualitatively, and has recently been shown to have considerable variability even between experienced readers (Exp-R).[1,2] The visual evaluation is subjective, extensive training is mandatory,[3] and, because of its reliance on expertise, is not always immediately

---

* Corresponding author. Tel: +972 46304488; fax: +972 46341369. Email: dsblond@gmail.com

available when needed for urgent decision-making. Automated evaluation is desirable in many areas of medical imaging and is lacking in evaluations of echocardiograms. Automatic evaluation was discussed long ago by Spiegelhalter.[4] He emphasized the crucial importance of the reproducibility of automated evaluation and stated its advantages. In cardiology, automatic quantification has been shown to be partially feasible and has been applied in some areas such as electrocardiographic (ECG) analysis and shown to be feasible in the analysis of single-photon emission computed tomography, magnetic resonance imaging, and computed tomographic imaging.[5–9] Unfortunately, applications for automatic interpretation of LV function in echocardiography are currently unavailable to clinicians, although some have been proposed, especially for automatic endocardial boundary detection,[10,11] but to the best of our knowledge, none have disseminated into routine clinical practice. The main reason for their failure to gain widespread acceptance is probably due to the lack of large, scientifically sound validation studies.

Recently, a method capable of automatic determinations of longitudinal strain derived from speckle tracking on two-dimensional (2D) echo loops from apical views was shown to classify WM similar to the regular visual method that is routinely used by Exp-R.[12–14] A software application based on this method was developed using segmental peak systolic longitudinal 2D strain calibrated by expert readers' visual classification of WM on the same loops. Liel-Cohen *et al.*[14] presented data on the reproducibility of WM classifications using this application compared with the classical method and demonstrated that in the hands of Exp-R, it was slightly higher than their visual assessments. However, the real need is for an automatic application that will assist inexperienced readers (InExp-R) in detection and classification of WM abnormalities.

Therefore, in the present study, we tested the hypothesis that the use of this automatic application for classification of WM by longitudinal 2D strain would yield similar results if used by either totally inexperienced or highly experienced echocardiographers.

Methods

Patients

The patient population has been previously described in detail.[1,14] In short, 105 patients took part in this study, mean age 59.3 ± 14.4, 67% males. The group included 62 patients with an acute myocardial infarction, 28 subjects with chest pain but who had ischaemic heart disease excluded by history, lack of ischaemic changes on ECG, no rise in enzymatic markers of infarction, and by a negative coronary angiogram (the latter not performed in 3 patients), and 15 patients who had a confirmed diagnosis of dilated cardiomyopathy.

Readers

Two groups used the application in this study: a group of 12 echocardiographers reading echos in a full job capacity, Exp-R, and a second group of 11 students with neither prior knowledge of cardiology in general nor experience in viewing or analysing echocardiograms in particular, InExp-R. Most of the students were from various university faculties who responded to an advertisement. Each student received a

15 min 'hands on' tutorial in the use of the application for analysis of echo loops.

Data analysis

Readers from both groups were provided with CDs on which 2D clips of a single cardiac cycle from the apical long-axis, two-chamber, and four-chamber views from each of 105 patients were obtained, using a Vivid 7 digital ultrasound scanner (General Electric, Horten, Norway) at a frame rate of 50–80 fps. The 2D strain data were analysed on EchoPac workstations (EchoPac-based 2D Strain research tool, GE software package). For this analysis, each reader was required to mark three 'anchor points' (both sides of the mitral annulus and the LV apex) in each of the three views. The following were then automatically generated: the end-systolic frame, endocardial border, division of the LV myocardium into 18 segments (6 basal, 6 mid-, and 6 apical segments per patient), determination of the mean segmental peak systolic longitudinal strain (only peak systolic strain values were used in this study), quality of speckle tracking (segments with poor tracking quality were not analysed), and recording results of the application's analysis in an Excel file. The students group used a newer version of the software that automatically adjusted the 'anchor points'. Aside from this feature, the applications were identical. An example of the analysis of the four-chamber view in one of the patients is presented in *Figure 1*.

The 'application'

The method used to create the application was described in detail[14] and can be summarized as follows: to create an automatic application for WM analysis based on longitudinal strain that would correlate with analyses by the human eyeballing method, it was necessary to categorize peak longitudinal systolic strain (continuous data) into ranges that correspond to normal, hypokinetic, and akinetic segments by the eyeballing method. Twelve highly Exp-R classified each segment of the LV myocardium from 105 patients. Only segments with a classification that had a majority of at least three readers above the next most prevalent classification were included in the 'gold standard' (1536 segments). Cut-off points for segmental peak systolic strain which best discriminated between normal, hypokinetic, and akinetic segments in this 'gold standard' were determined separately for base, mid-, and apical LV levels by receiver-operating characteristic (ROC) analysis (cut-off values were derived from 630 segments at each level). This analysis yielded the following cut-off points: basal segments: normal, less than −10.9%; hypokinetic, −10.9 to −6.2%; akinetic, greater than −6.2%; mid LV segments: normal, less than −12.6%; hypokinetic, −12.6 to −7.9%; akinetic, greater than −7.9%; apical segments: normal, less than −14.1%; hypokinetic, −14.1 to −9.1%; akinetic, greater than −9.1%. These ranges are used herein as reference values.

The segmental classifications obtained by the Exp-R and InExp-R using the application were compared with the mode of the segmental classifications made visually by the Exp-R who were blinded to the application's classification.

Statistical analysis

Data were analysed using Matlab (The MathWorks Inc., Natick, MA, USA) and SAS® version 9.2 (SAS Institute, Cary, NC, USA) software. Continuous data are presented as mean ± standard deviation (SD) (or residual SD from LSMeans). Categorical data are presented as a count and percentage. Statistical significance was defined as a *P*-value of <0.05.

The relationship between peak systolic strain analysis of the InExp-R and Exp-R groups was assessed by comparing their mean values using a
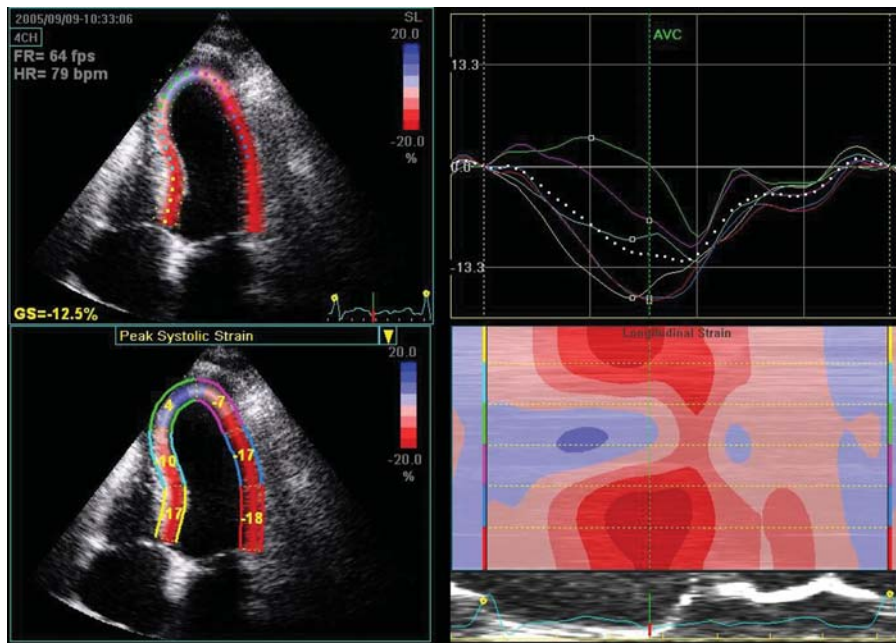
Figure 1 An example of a longitudinal strain distribution in the four-chamber view with segmental strain values indicated.

repeated-measures ANOVA model and visually on distribution plots (*Figure 2*). ROC curves are presented (*Figure 3*), showing the power of the average segmental peak systolic strain of both Exp-R and InExp-R to discriminate normal from abnormal (i.e. normal from hypokinetic or akinetic segments). The ROC curves are summarized by the area under the curve (AUC).

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were calculated for each reader in the Exp-R and InExp-R groups using their respective classifications of WM contraction vs. the visual mode Exp-R classification. The mean values per group (Exp-R and InExp-R) are presented.

Inter-observer reproducibility was assessed by the intra-class correlation coefficient (ICC) which was calculated with the SAS PROC MIXED procedure, by fitting two random-effects models where the three-level ordinal score (normal = 1, hypokinetic = 2, akinetic or worse = 3) was modelled with the physician and segment (within subject) entered as random effects. The ICCs were then calculated from each of the models' variance components as the ratio of the between-reader error variance and the total variance.

Results

Each of the 12 Exp-R and the 11 InExp-R analysed the 1890 segments available from the 105 patients. The number of segments that could not be assigned a classification by the Exp-R ranged from 17 (0.9%) to 217 (14.5%) with a mean of 77.8 (4.1%), whereas the range for the InExp-R was 20 (1%)–110 (5.8%) with a mean of 50.4 (2.7%). These segments could not be assigned a classification because of poor tracking quality determined automatically by the application. The distribution of these segments

within the LV levels was similar in both groups (in decreasing order: basal, apical, and middle).

The overall ICC for segmental classification was high for both groups but slightly better for the InExp-R (ICC 0.89) vs. the Exp-R group (ICC 0.83) (an ICC of >0.8 is considered excellent). ICCs for overall, normal, hypokinetic, and akinetic segments for each group of readers are presented in *Table 1*.

The model-estimated overall mean (LSMeans) peak systolic strain of all segments from all patients was similar for the Exp-R and InExp-R groups (−15.5 ± 0.27 vs. −15.2 ± 0.28%, respectively, P = NS). Differences in peak systolic strain between groups were negligible for patient subgroups as well: ischaemic heart disease 0.22 ± 0.32%, dilated cardiomyopathy −0.08 ± 0.33%, and normal 0.58 ± 0.32%, P = NS for all. *Figure 2* demonstrates the similarity of the distribution of peak systolic strain by both groups of analysers over the whole range of the peak systolic strain.

Segmental classifications of each InExp-R and each Exp-R were compared with the mode classifications determined visually by the Exp-R group. On average, 90.8% of the InExp-R segmental scores were concordant with the Exp-R visual mode scores. There were 1218 segments considered 'normal' by Exp-R and also by InExp-R (65.9% of all readable segments, concordance of 94%), 184 segments considered 'hypokinetic' (10% of readable segments, concordance of 72%), and 275 segments considered 'akinetic' (14.9% of readable segments, concordance of 90%). Thus, the largest difference in classification was in segments classified as 'hypokinetic' by Exp-R, and of these, half were over-classified and half under-classified by the InExp-R group. Overall, in a head-to-head comparison of both groups, InExp-R under-classified
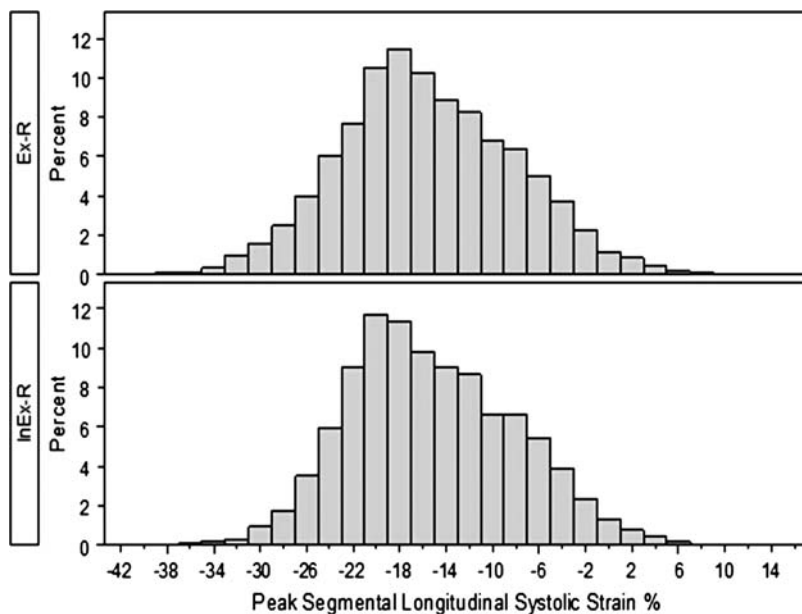
Figure 2 Histograms of segmental peak systolic strain values of experienced and inexperienced readers.
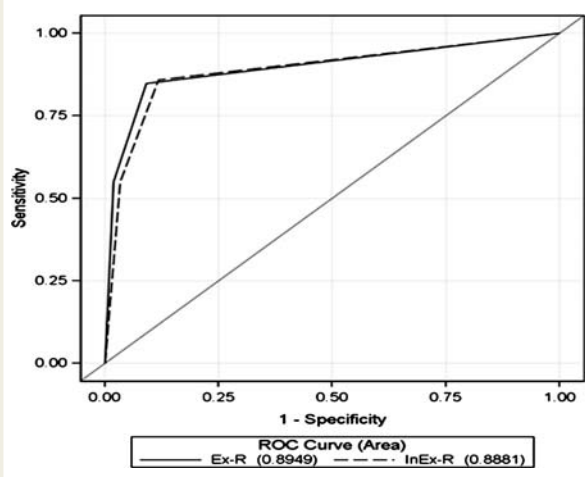

Figure 3 Receiver-operating characteristic curves for the ability of experienced and inexperienced reader groups to classify segments according to the experienced reader visual mode classification.

Table 1 Intra-class correlation coefficient statistics using score 3 and 2 levels for normal, hypokinetic, and akinetic segmental classifications by both groups of readers

	Overall	Normal vs. others	Hypokinetic vs. others	Akinetic vs. others
Exp-R	0.83	0.77	0.47	0.75
InExp-R	0.89	0.84	0.65	0.84

and abnormal segments as determined by the visual mode classification of the Exp-R. The mean AUC for the InExp-R group was 0.888, and for the Exp-R, it was 0.895.

Discussion

The ability to evaluate LV WM abnormalities is essential for clinical decision-making, often in urgent situations. Yet, so far, no automatic application gained acceptance and it is still necessary to get an expert-on-call to determine segmental WM. Our group has recently shown that WM assessments made by Exp-R applying the automatic longitudinal 2D strain application we developed were similar to their own assessments using the classical eyeballing method.[14] However, before implementing the new automatic application for classification of WM in the clinical setting, it is essential to show that it can be used with confidence, regardless of the degree of training the operator has in the field of

3.2% and over-classified 5.8% of segments compared with the mode classification of the Exp-R group.

Compared with the visual mode classification of the Exp-R, the mean sensitivity, specificity, PPV, NPV, and accuracy of the Exp-R mode-segmental score for dichotomizing segments into normal or abnormal contraction were 87, 85, 94, 70, and 86%, respectively. These accuracy measures are identical to those of the InExp-R.

Figure 3 presents the ROC curve of median Exp-R and median InExp-R classification capability to distinguish between normal

echocardiography. Thus, this study was designed to test the ability of the application to assess LV WM based on longitudinal 2D strain, independently of the operator's training in echocardiography.

We compared the classification of segmental contraction derived entirely from an automated application based on longitudinal 2D strain, operated by totally InExp-R vs. highly Exp-R, without any added insight or findings apparent by visual inspection of the echocardiograms, in a large group of patients. The results of both groups of readers were compared with the mode classifications made by the Exp-R group using the visual method as a reference. The results show that InExp-R could utilize the automatic application to derive classifications of segmental WM with results comparable to those determined by a group of highly experienced echocardiographers using this application.

Despite the lack of any significant effect of expertise in echocardiography on the segmental classification by the application when operated by InExp-R and Exp-R, this application cannot replace the comprehensive evaluation of LV WM done by experienced echocardiographers including assessment of wall thickening and excursion from multiple sites and repeated beats, and is therefore limited. However, this application may serve for preliminary evaluation, for teaching purposes, and even for an unbiased 'second opinion'.

Although the average peak systolic strain of both groups were almost identical, the sensitivity and specificity of the InExp-R classifications compared with those of the Exp-R were not perfect. Some degree of non-concordance was found between the results of both groups especially regarding segments considered by the Exp-R group to be hypokinetic. We and others have shown that even among Exp-R using the classical visual method, there is considerable variability regarding hypokinetic segments,[1,2] but had hoped that the lesser room for subjective judgement using the automated application would produce less variability between the groups (regardless of which group was closer to the true quality of contraction).

Paradoxically, the InExp-R group had a marginally better reproducibility than the Exp-R group, although the difference was far from being statistically significant. Perhaps, the Exp-R outsmarted themselves in choosing the three 'anchor points', whereas among the InExp-R, these points were automatically corrected leaving very little room for their personal judgement to affect the outcome of the automatic calculations. Similarly, the Exp-R's option to 'correct' the automatically generated endocardial borders and the timing of end-systole (aortic valve closure) probably decreased their reproducibility.

Liel-Cohen et al.[14] have shown that when used by Exp-R, this application has a lower PPV than a NPV. Thus, it is important to keep in mind that a positive detection of WM abnormalities is probably less reliable than a negative result and false positives may occur. Thus, if this application was to be used in the emergency room setting, where it is important not to miss a patient with an acute coronary syndrome, this set-up would be beneficial because of its high specificity which enables safe discharge of patients with negative findings at the cost of some unnecessary admissions due to the application's tendency to overestimate WM abnormalities. However, in the setting of stress echo testing, false-positive results could be a problem that may be eliminated by manually viewing the whole test in conjunction with other clinical variables, thus avoiding unnecessary further investigation. A different set of cut-off points may be necessary depending on the purpose of the test.

It should be stressed that this study was not designed to validate the automatic application we used. This was done and published elsewhere.[14] The purpose of the current study was to compare results obtained with the application when used by Exp-R vs. InExp-R in order to verify its objective nature and its reproducibility. Although we have previously shown that this application's classifications are similar to those of Exp-R when limited to the analysis of only apical views, the use of this application by novice readers cannot replace Exp-R. Novice readers cannot evaluate numerous factors that Exp-R incorporate into their assessments in order to make them clinically useful, such as the technical quality of the clip, adequacy of speckle tracking, additional information obtained from parasternal views, etc.

A recent consensus statement of the ASE/EAE[15] concluded that there are still unresolved issues regarding WM analysis with strain and tissue velocity imaging because not all clinical studies have unanimously confirmed their accuracy compared with expert readers' visual interpretation in ischaemic heart disease. In a previous study,[14] we have shown a good correlation between classification of WM by strain analysis and by visual assessment as did others.[16] The current study goes one step further and highlights the fact that the application for automated strain analysis does not necessitate an expert to run it.

The main limitation of the study is that the application used by the two groups was not identical. The students used a newer version, which had only one additional feature: automatic correction of the three manually marked 'anchor points'. Although we could not compare the three points determined by the Exp-R manually vs. the three points automatically generated by the application for the InExp-R, we expect the automatic correction to have improved only the InExp-R results, whereas for an experienced echocardiographer, the determination of these points is a trivial task. Also, PPV and NPV are highly dependent on the prevalence of the finding in the given population. Thus, our data derived mostly from a population presenting to emergency rooms with a suspected myocardial infarction may not be applicable to all patient populations. Other potential limitations are the use of a single cardiac cycle from each view, the use of only apical views, and longitudinal strain for segmental WM classification. Future studies may suggest that more accurate classifications can be obtained by averaging a number of cycles from each view and with the possible incorporation of circumferential and radial strain.

In conclusion, we have shown that automatic assessment of segmental LV WM, based on 2D longitudinal strain, can be done by inexperienced personnel with similar results to its use by experienced echocardiographers. This application may be useful in the hands of InExp-R for preliminary WM assessments as well as a 'second opinion' in the hands of Exp-R to assist in the complicated task of quantifying segmental WM. However, interpretation of results and medical decisions based on them must be applied only after careful analysis of comprehensive echo studies and consideration of the specific clinical context.

Conflict of interest: Two of the authors (P.L. and Z.F.) work for GE Healthcare and are involved in developing echo software for automatic detection and quantification of LV wall motion.

References

1. Blondheim DS, Beeri R, Feinberg MS, Vaturi M, Shimoni S, Fehske W *et al.* Reliability of visual assessment of global and segmental left ventricular function: a multicenter study by the Israeli Echocardiography Research Group. *J Am Soc Echocardiogr* 2010;**23**:258–64.
2. Hoffmann R, Lethen H, Marwick T, Arnese M, Fioretti P, Pingitore A *et al.* Analysis of interinstitutional observer agreement in interpretation of dobutamine stress echocardiograms. *J Am Coll Cardiol* 1996;**27**:330–6.
3. Picano E, Lattanzi F, Orlandini A, Maorini C, L'abbate A. Stress echocardiography and the human factor: the importance of being expert. *J Am Coll Cardiol* 1991; **17**:666–9.
4. Spiegelhalter DJ. Computer aided decision making in medicine. *BMJ* 1984;**289**: 567–8.
5. Zizzo C, Hassani A, Turner D. Automatic detection and imaging of ischemic changes during electrocardiogram monitoring. *IEEE Trans Biomed Eng* 2008; **55**:1243–7.
6. Porenta G, Dorffner G, Kundrat S, Petta P, Duit-Schedimayer J, Sochor H. Automated interpretation of planar thallium-201-dipyridamole stress-redistribution scintigrams using artificial neural networks. *J Nucl Med* 1994; **35**:2041–7.
7. Mazonakis M, Grinias E, Pagonidis K, Tziritas G, Damilakis J. Development and evaluation of a semiautomatic segmentation method for the estimation of LV parameters on cine MR images. *Phys Med Biol* 2010;**55**:1127–40.
8. Theisen D, Sandner TA, Bauner K, Hayes C, Rist C, Reiser MF *et al.* Unsupervised fully automated inline analysis of global left ventricular function in CINE MR imaging. *Invest Radiol* 2009;**44**:476–82.
9. Plumhans C, Keil S, Ocklenburg C, Mühlenbruch G, Behrendt FF, Günther RW *et al.* Comparison of manual, semi- and fully automated heart segmentation for assessing global left ventricular function in multidetector computed tomography. *Invest Radiol* 2009;**44**:476–82.
10. Bosch JG, Mitchell SC, Lelieveldt BPF, Nijland F, Kamp O, Sonka M *et al.* Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Trans Med Imaging* 2002;**21**:1374–83.
11. Comaniciu D, Zhou XS, Krishnan S. Robust real-time myocardial border tracking for echocardiography: an information fusion approach. *IEEE trans Med Imaging* 2004;**23**:849–60.
12. Perk G, Tunick PA, Kronzon I. Non-Doppler two-dimensional strain imaging by echocardiography—from technical considerations to clinical applications. *J Am Soc Echocardiogr* 2007;**20**:234–43.
13. Leitman M, Lysyansky P, Gurevich J, Friedman Z, Sucher E, Rosenblatt S *et al.* Real-time quantitative automatic assessment of left ventricular ejection fraction and regional wall motion by speckle imaging. *Isr Med Assoc J* 2007;**9**:281–5.
14. Liel-Cohen N, Tsadok Y, Beeri R, Lysyansky P, Agmon Y, Feinberg MS *et al.* A new tool for automatic assessment of segmental wall motion based on longitudinal 2D strain: a multicenter study by the Israeli Echocardiography Research Group. *Circ Cardiovasc Imaging* 2010;**3**:47–53.
15. Mor-Avi V, Lang RM, Badano LP, Belohlavek M, Cardim NM, Derumeaux G *et al.* Current and evolving echocardiographic techniques for the quantitative evaluation of cardiac mechanics: ASE/EAE consensus statement on methodology and indications. *J Am Soc Echocardiogr* 2011;**24**:277–313.
16. Ingul CB, Stoylen A, Slordahl SA, Wiseth R, Burgess M, Marwick TH. Automated analysis of myocardial deformation at dobutamine stress echocardiography: an angiographic validation. *J Am Coll Cardiol* 2007;**49**:1651–9.