The demystification of magnetic resonance imaging?

Rainer Hoffmann*

Medical Clinic I, University Aachen, Pauwelsstrasse, 52074 Aachen, Germany

Online publish-ahead-of-print 23 May 2006

This editorial refers to 'Determination of interobserver variability for identifying inducible left ventricular wall motion abnormalities during dobutamine stress magnetic resonance imaging' by I. Paetsch, on page 1459

The interobserver variability in test interpretation is a key parameter to evaluate the diagnostic reliability of a diagnostic examination. It has been evaluated for most clinically relevant cardiac imaging modalities. In many instances the results have been rather sobering. A considerable variability in the visual assessment of coronary artery stenosis severity has been shown between different readers of coronary angiograms. A high reader disagreement has been shown for interpretation of dobutamine stress echocardiograms. The high variability in the interpretation of dobutamine stress echocardiograms has become a matter of debate and has been used as an argument to doubt the diagnostic competence of the method. Low image quality, a large variability in test performance, and insufficient standardization in image interpretation without any objective parameter to define the test result has been referred to as the major limitations of stress echocardiography. Lower observer variability due to more standardized image acquisition, better image quality, and objective parameters of test interpretation has been shown for perfusion scintigraphy and has been used as an argument for its favourable diagnostic performance. Regarding stress echocardiography, multiple shortcomings could be reduced over time. Image quality could be improved. Along with digital image acquisition and more standardized interpretation criteria refinements in the stress echocardiographic techniques have resulted in higher interobserver agreement. Still, interobserver agreement remained far from optimal.

Dobutamine stress cardiovascular magnetic resonance imaging (DSMR) incorporates multiple advantages when compared with stress echocardiography. It is performed using standardized acquisition protocols. The acquired imaging views as well as quality of images is less dependent on the individual sonographers performance and the image quality is excellent in most cases. Thus, interobserver agreement is expected to be high considering the most obvious limitations of dobutamine stress echocardiography. However, data on interobserver agreement in the interpretation of DSMR were sparse until recently. In this issue of the Journal, Paesch et al. report on the interobserver variability in the interpretation of DSMR evaluated in a multicenter setting. The study confirms an excellent image quality in the majority of imaging views with the use of DSMR as reported earlier by the investigators as well as other authors. The study confirms also a high diagnostic accuracy for detection of coronary artery disease for all four readers using this technique. However, while the accuracy was almost similar among the four readers, sensitivity for detection of CAD ranged from as low as 71% to as high as 86%, and the specificity had a discordant trend indicating already that readers vary in their interpretation of image data. The surprise of this study is the relatively low interobserver agreement between the four expert observers from different institutions. The disagreement between the four readers was equally distributed and not the effect of one reader with significant deviations from the other readers.

How do we explain this result? Excellent images were thought to ensure high diagnostic accuracy but also to facilitate consistent interpretation because of a high confidence in the perception and interpretation of images. Impaired echocardiographic image quality has been a major cause of lower diagnostic accuracy when comparing dobutamine stress echocardiography with dobutamine stress MRI. Poor imaging conditions with resulting insecurity on the interpretation of images have also been shown to be a major cause of disagreement when analysing stress echocardiograms. The rather disappointing level of interobserver agreement confirms other recently published studies which have also indicated that use of MRI is no warranty to reach nearly perfect agreement between different readers on analysis of regional left ventricular function.

The interobserver agreement was on a level previously reported for dobutamine echocardiography applied with most advanced echocardiographic techniques. Given these results it appears that reader agreement may be method and image-quality dependent on a low level of agreement. However, once a relatively high image quality is reached this aspect loses importance and other aspects come into play. Thus, we do not reach almost perfect reader agreement based on excellent image quality alone.

The difficulty of divergent test results relates to a different issue, involving in particular, the final pathway of image interpretation, which remains a subjective process based on the individual perception. Analysis of regional systolic...
function is a highly complex process. It relates to the analysis of dynamic images considering changes in wall thickness as well as endocardial inward motion and synchronicity of analysed left ventricular segments. There are obviously different thresholds in the subjective evaluation of regional wall motion abnormalities. Although there are doubtful, conservative reviewers who have hard times in deciding about an abnormality, other more liberal readers are quicker in their decision, which is rather driven by an intention to detect abnormalities. What are the reasons for a different perception of visual data? Current research in the artificial intelligence and computer science area attempts to understand the complex human perception and response to visual data. Beyond the initial perception of images, the cognitive analysis of specific patterns of image configurations comes into play. Furthermore, a definite level of expectation towards the test result has an impact on test interpretation. Finally, the level of confidence in the findings can vary. A profound knowledge of these processes is the key to algorithms, which might eventually support or even reliably replace the human subjective visual interpretation.

DSMR might, in particular, encounter a difficulty related to induction of new wall motion abnormalities as a sign of test abnormality. Reader disagreement occurs particularly in cases with rather subtle new wall motion abnormalities, whereas cases with pronounced wall motion abnormalities tend to be read more homogenously between different readers. DSMR tests are normally stopped after induction of even minor new wall motion abnormalities to prevent difficulty in handling emergency situations with the patient being in the MR scanner. Thus, DSMR cases might be stopped in many instances at an earlier stage of ischaemia than for instance dobutamine stress echocardiograms given the stress situation.

Considering the complex analysis process related to the interpretation of moving images, an imaging technique based on dynamic images may be at a disadvantage when compared with a technique which requires analysis of still frame images only, such as nuclear perfusion imaging techniques. It appears more vulnerable by the individual perception of the reader.

High interobserver agreement on test interpretation is required to ensure similar treatment decisions and avoid clinical misdirections. As long as we are dependent on subjective visual interpretation reading circumstances should be optimized. This relates to factors such as monitor brightness, contrast, playback rates, and side-by-side display of rest and stress images, which might have an impact on the individual perception of imaging data and confidence in reported test results. We might have to consider reporting the confidence level in our findings as an additional information to prevent inadequate rating of test results with subsequent inappropriate clinical decisions or treatments.

Objective quantitative measures of regional left ventricular function are desirable and may assist the evaluation of regional wall motion in the future. These techniques are becoming available for different imaging techniques and have reached a high level for MR imaging. However, they may function only as a supportive parameter to bolster our own subjective analysis of image data.

Conflict of interest: none declared.

References


