Quantitative Image Analysis in Clinical Trials - Current Standards and Future Developments

By Dr Anton van Weert, Managing Director of Heart Core bv and Professor Hans R Head of the Laboratory of Image Processing (LKEB) at the Leiden University Medi Center (LUMC) and General Director of Heart Core bv

The use of imaging is an established method in defining primary and secondary endpoints in clinical trials. The tremendous technological improvements in the performance of imaging equipment and software tools have played an important role in the application of imaging for study outcome purposes. Eight to 10 years ago, standard X-ray angiography was still stored mainly on 35 cinefilm and plain film and required precious film development processes in-hospital to guarantee good quality of the angiograms. At that time replacement of this proven concept by digital cath lab systems with DICOM compatible standards linked to a digital storage and archiving infrastructure started and took place in a few years. Although the spatial resolution of digital images with 5122 matrix sizes is slightly belc cinefilm, the better contrast resolution and the advantages of computerised evaluation processing and in combination with the option for semi-automatic quantitative assessment of the images on worksta dedicated software, facilitated the change to a digital standard.

More recent technological developments are directed towards non-invasive imaging approaches. The us example, fast MR and multi-slice CT scanners has become widespread in many therapeutic areas, for d purposes as well as for clinical research applications. These three- and four-dimensional imaging techn through subsequent acquisitions in the same imaging session, allow for the assessment of multiple para an object. In addition, clinicians can gain increasingly detailed information on the appearance and status diseased organ and the short and long-term monitoring of the efficacy of the treatment. Hence, imaging accepted as a valuable tool in the assessment of inclusion and exclusion criteria for patients participatin clinical trials and the subsequent offline interpretation of these images at an independent image analysis lab.

The classical method of evaluating the efficacy of a particular treatment is to set up a reading session in two clinicians, such as radiologists or cardiologists, review the images in a randomised and blinded fash the event of a discrepancy between the two readers, a third expert will adjudicate. Although these interp may be important from a clinical point of view, and sometimes even the only option for image review, the concrete quantitative data usually requires large numbers of subjects to be evaluated for the study endp
The availability of online software packages to support image interpretation and semi-automated image processing has become an essential element of the imaging era in the same way that imaging equipment in hospitals nowadays. For offline post processing and image analysis, the developments may be even more impressive. Where in the past, simple point-to-point or caliper measurements were conducted to generate absolute data, at present state-of-the-art software programs have been developed to perform complex threedimensional analyses to evaluate the condition of a diseased part of the body and the evolution of the disease over time. In some applications, this can even be done by automated matched analyses, whereby the images at follow up can be mapped with those acquired at screening or prior to the treatment.

Because of the much smaller variability in quantitative image analyses, the population sizes can be decreased significantly if objective measurements are obtained. This is why the use of quantitative imaging endpoints in clinical trials has grown so rapidly in the past 15 years; it has been driven by overwhelming developments in both imaging equipment and validated software for offline quantitative analysis based on automated contour techniques.

It is important that quantitative image analyses are conducted centrally at an independent core laboratory in the same way that blinded readings are, according to standard operating procedures. Staff must be well trained which, given the complexity of the imaging techniques, the subsequent interpretation and the software tools for offline analysis, will usually take at least one month, but can take more than three, depending on the imaging modality. However, even when qualified, specific image analysis guidelines must be described, taught and applied to establish low inter- and intra-observer variability for each study.

Based on experience with a variety of imaging modalities used in different therapeutic areas, the current standards in the application of quantitative image analysis in clinical trials will be described below, using magnetic resonance imaging (MRI) as an example. Subsequently, application of these principles in evolving imaging techniques for pulmonology, orthopaedics and neurology will be described.

**Cardiac MRI**

To evaluate the efficacy of a new drug in the improvement of left ventricular (LV) function, cardiac MRI has become the reference standard for the quantitative measurement of mass, cardiac volumes, wall thickening/thinning, functional parameters, like the ejection fraction and wall motion, and perfusion. When compared to echocardiography, the superior accuracy of MRI-derived parameters allows for a tremendous reduction in the number of subjects required in clinical trials.

Consequently, results regarding the safety and efficacy of a new potential drug treatment are available in a shorter period of time and decisions to either continue or initiate new trials can be made at an earlier stage. The result, the number of patients exposed to drugs that in the end appear to be non-effective or show unacceptable adverse effects can be reduced and effective treatments may become available to the market sooner, an important issue from an ethical point of view. At the same time, considerable cost-savings can be made during the clinical research programme.

In order to implement quantitative cardiac MRI in a clinical trial, the first step is the development of the study protocol. It is evident that the definition of clinical criteria is reflected in the inclusion and exclusion criteria. However, it is also important that MRI related matters are covered. For example, patients with an ICD (implantable cardioverter defibrillator), pacemaker or aneurysm clip must be excluded from the trial. Also, baseline characteristics assessed by MRI, like baseline LV ejection fraction, LV mass and LV volume, may be important. This can be measured online by the investigator, but will preferably be confirmed at the central core lab in order to assure independent and standardised image post-processing and quantitative analysis.

For standardisation purposes, the acquisition guidelines for MRI must be well described and documented. These guidelines include technical settings of equipment, like the pulse sequence, spatial resolution (slice thickness), field of view and coil settings, as well as the scan protocol for optimal imaging of the heart. Each investigation site must confirm that all requirements with respect to image acquisition can be met, and preferably verified by evaluation of test scans.

The logistics of the imaging materials is done either electronically or by courier service. Electronic transfer
done by secure FTP transfer or by using a dedicated web-based transfer, both fully audit trailed and in compliance with 21 CFR Part 11 regulations. Whenever they are needed, blinding of the images must be performed at the core lab prior to starting the analysis process. The next step is to verify the compliance image acquisition with the guidelines and technical settings. Immediate feedback to the investigative site is important in order to confirm that acquisition was performed well or to communicate how acquisition and guideline compliance could be improved. In addition, it is important to verify that acquisition was done in same way at both baseline and follow-up visits, so that comparison of the quantitative analysis data is possible. Moreover, this compliance can be increased by communication to the sites prior to the patients' follow-up visits, that the settings are the same with baseline.

The offline analysis is done using validated software for automated contour detection of all slices from base to apex at, for example, rest and stress and end-diastole and end-systole. Analysis can only be done by well-trained technicians or reading specialists according to standard operating procedures. Inter- and intra-observer variability must be documented and performed either within each study or at the core lab at regular intervals. Given the complexity of the analyses and the aforementioned prerequisites, it is clear that readers on an ad hoc basis cannot do this.

A second technician must verify that analysis was done according to the SOPs and excludes subjective interpretation as far as possible. If, after this, QC process a disagreement is noticed, the first technician will perform the analysis again or a third technician will adjudicate. After analysis is approved, the quantitative data is processed by data management and transferred to the database.

**Future Perspectives**

Provided that well-validated software is available, offline quantitative image analysis is the preferred method when imaging is involved in clinical trials. The relatively recent developments in cardiac MRI analysis has that even complex images can be used for study endpoints. Also, in other therapeutic areas, software algorithms for automated quantitative image analysis have recently become available, and it is expected that these an increasingly important role in clinical trials in the coming years. Three examples are described below.

**MRI of the Brain**

For the detection of white matter lesion (WML) load by MRI several methods are available, varying from interpretation (increase versus decrease) to computer-assisted detection with still significant human intervention. Recently, new software became available for quantification of WML to investigate its role in normal ageing, dementia and late-onset depression. For optimal analysis, three types of MR images must be acquired: fast spin-echo imaging; PD, T2 and FLAIR, on which WML is seen as hyper-intensities (see Figure 2). FLAIR is applied to discriminate between CSF and WML, but has its limitations due to low sensitivity in the infratentorial area and may present some hyper-intense artefacts that look like lesions.

The automated analysis is done in three steps: lobe template generation, fully automated brain stripping delineation and quality control. For each subject, all segmentation results are saved for true threedimensional verification (which can be time-consuming), but also mosaic images are generated showing selected slices for faster quality control. In a validation study, 120 patients were processed semi-automatically by well-trained readers and automatically, using the above approach, to compare the volumes of the intra-cranial mask.

The mean degree of overlap was 97 per cent +/-15 per cent, with an intra-class correlation of 0.949 (alpha = 0.997). On average, there was no overlap of the automatic mask with the semi-automatic mask in only 3 cent (+/-2 per cent); vice versa this was 2.6 per cent (+/-2 per cent). This indicates that a fully automated quantification method is available for large clinical trials, and is expected to become the preferred future application given its speed, reproducibility and objectivity.

**Orthopedic Implants**

To evaluate the success of a treatment for an orthopaedic implant (such as total hip or knee replacement) it is important that the loosening of an implant is detected as soon as possible. In clinical practice, the loosening of prostheses is assessed indirectly in successive radiographs by measuring radiolucency lines around the prosthesis and position differences of the prosthesis relative to the bone. Radiolucency lines indicate the presence of a fibrous layer. However, these measurements are not very accurate: radiolucency may occur in areas that
over-projected by the metal of the implant and the amount of radiolucency may be under-estimated. A more accurate approach is quantitative roentgen stereophotogrammetric analysis (RSA) (see Figure 3).

This technique involves small roentgen opaque markers being introduced in the bone and attached to the prosthesis to serve as well-defined artificial landmarks. Two synchronised roentgen foci are used to obtain a stereo image of the bone and the prosthesis. Using a calibration object that holds tantalum markers at a known positions, the positions of the roentgen foci are assessed. The co-ordinates of bone and prosthesis markers are then accurately measured, after which the three-dimensional position of the markers is reconstructed with RSA software.

Finally, the change in the position of the prosthesis markers relative to the bone markers is determined, translation and rotations of the prosthesis can be calculated. The importance of quantification became obvious after the introduction of Boneloc cement in 1991. After the market launch, several clinics reported a much higher incidence in the loosening of implants after using Boneloc cement as compared to conventional practice, and this was confirmed by two clinical trials involving quantitative RSA. Unfortunately, at that point in time, Boneloc had been used in more than 1,000 cases in Norway alone, and after a period of four and a half years, the revision rate of the prostheses was 14 times larger than for prostheses fixed with conventional cement.

Since then, many clinical trials involving relatively small patient cohorts showed the correlation between migration (three to 12 months) of the implant as measured by RSA and the risk of early implant revision. However, despite these convincing facts, discussions are still ongoing, and regulation on quantitative RSA is not yet mandatory as part of the data submitted to regulatory bodies before market launch. Besides the ethical aspects, such an approach would also allow a more objective comparison of new implants or therapy approaches with current clinical practice, similar to clinical trials involving new drugs.

New developments may enhance this discussion; whereas RSA has been applied with tantalum markers attached to the implant, software for so-called model-based RSA has been recently introduced. By model RSA, a three-dimensional model of the implant, obtained either by using the computer aided design (CAD) reversed engineering, is used for matching with the RSA radiographs and the subsequent quantification of degree of rotation, migration and translation of the implant relative to the bone. As a result, the costly and time consuming validated production of prostheses with tantalum markers is no longer required. In addition, it excludes any marker interference with the performance of the implant.

**CT in Lung Emphysema**

Only a relatively low number of patients suffer from lung emphysema, a disease that results in the progressive decreasing of physical density in lung tissue. Lung densitometry, as measured by multi-slice computed tomography (MSCT), has been demonstrated to be more sensitive than lung function tests in the assessment of progression of emphysema. Cross-sectional data showed significant correlation with pathology and lung function tests but at a high level of reproducibility.

However, longitudinal studies require an even higher level of reproducibility. As described above, in relation to MRI, this can only be achieved by the consistent standardisation of CT image acquisition, continuous validation of the equipment settings used during the acquisition, and the availability of validated software for offline quantitative analysis (see Figure 4). The standardisation and verification of technical settings can be verified using a MSCT scan from a phantom that simulates changes in lung density.

For application in clinical trials, offline analysis and quality control feedback is a prerequisite, given the need for quality control feedback with respect to the technical settings of the different MSCT scanners on the market. The consistency of CT acquisition from patient to patient and at follow-up with respect to baseline, and know the offline quantitative analysis process. Since all these ingredients are already available, more proof will be available from future longitudinal clinical trials in which lung density as measured by MSCT is the primary endpoint, as compared with functional lung tests.

**Conclusion**

In many therapeutic areas, quantitative image analysis is already an established method of assessing the primary endpoint of a clinical trial. This is a result of the enormous improvements in both imaging equipment...
offline post-processing software tools. In contrast to blinded readings, application of quantitative image analysis usually requires fewer subjects to participate in a trial and yields hard data to compare new therapies with current clinical practice.

Although blinded readings are still the only option in several therapeutic areas and will remain important to obtain the clinical observations, the role of quantitative analysis will undoubtedly strongly increase in the coming decade. With the complexity of imaging equipment, its settings and output options, and the offline analysis software tools, there will be an important role for independent core labs to guide the investigational sites during the trial and to process the images in a validated and standardised environment. This was an important factor in cardiac MRI imaging becoming an established method of evaluating the efficacy of treatment in II-IV trials, both now and for the future.

The authors can be contacted at a.van.weert@heartcore.nl and j.h.c.reiber@lumc.nl