Published in www.clinicalleader.com; 25 January 2012.

Atychiphobia: A Fear That Makes Sense When Considering Independent Reader Performance In Clinical Trials

By Kohkan Shamsi, MD, PhD, Principal, RadMD and Rick Patt, MD, Principal, RadMD

Medical imaging in clinical trials has been used as a measure of efficacy for more than 20 years. It was recognized by the FDA (FDA's 1994 *Points to Consider for Developing Medical Imaging Drug and Biologic Products and Draft Guidance 2010)* that variability in evaluating trial images necessitated a more standardized, controlled and independent review process. This independent review ("blinded read") is meant to improve the quality of imaging data evaluations from clinical trials. The performance of the reader(s) is a, and sometimes the, critical component in establishing drug or device efficacy. In this review we will discuss factors that can influence reader performance and how to actively manage readers who are blinded to clinical data to reduce bias.

Clinical practice vs. pharmaceutical practice:
In the clinical practice of radiology, a radiologist's image interpretation is designed for individual patient management, and uses subjective terms such as "improved" or "worsening" to describe a disease process. We refer to the pharmaceutical practice of radiology, whereby readers use standardized image evaluation criteria developed to evaluate drug efficacy. And as the images they review are blinded, these readers are not influenced in their review by medico-legal or insurance issues because these images are not impact patient management. Clinical practice rarely meets the requirements to enable selection, measurement, and tracking of lesions, and may use non-standard terminology. Moreover, this standard of care clinical practice varies geographically both within the US and globally. In addition, the pharmaceutical practice of radiology, using standardized imaging efficacy criteria used in trials (eg. RECIST, IWG, etc), is not taught in radiology residency training programs.

Read paradigms:
In the blinded reads performed for diagnostic imaging contrast agents, a read paradigm of 3 readers evaluating all images is commonly used. The performance of these readers is compared to a standard of reference to calculate sensitivity and specificity of the diagnostic method. However, in most therapeutic trials, oncology being an example, there is generally no standard of reference (established truth) against which the imaging results can be compared. When two or more readers evaluate the same group of images there exists a potential for disagreement. A process for adjudicating differences between readers has been established in oncology trials whereby two primary readers independently read all images and a third independent reviewer is asked to review subjects' images when the primary readers disagree. The number of patients that require adjudication by the 3rd reader is known as adjudication rate.

Reader performance and adjudication rates:
In the 15+ years of the widespread use of the dual primary readers / single adjudicator blinded read paradigm, it has become accepted that there are multiple factors which impact adjudication rates in trials. A few of these include: imaging modality; disease process; lesion morphologic appearance and its impact on the ability to precisely measure lesions; blinded read study design; efficacy evaluation criteria; image quality; study population; reader qualification and experience reader training etc.

Regulatory authorities have not elucidated specific acceptable or non-acceptable adjudication rates, and rates of 25-40% or greater are routinely observed in oncology registration trials. Recently adjudication rates have come under scrutiny by the FDA and Oncology Drug Advisory Committee (ODAC) suggesting that a specific rate in a specific trial may be considered as a negative performance indicator. A few examples include:

- Trabectadin studied for ovarian cancer: ODAC 2009 questioned 29% adjudication

rate.

- Pralatrexate studied for relapsed or refractory peripheral T-cell lymphoma. - FDA "flags" Phase 3 study and asks for ODAC in 2009 input due to high adjudication rate (51%).
- Bevacizumab studied as single therapy in previously treated glioblastoma. The adjudication rate was 46%. ODAC in 2009 noted "The degree of discordance underscores the difficulties in accurately assessing response or progression in GBM."

Reader performance management:

Reader selection:
To ensure optimal reader performance, active reader performance management is required. Reader management begins with understanding the clinical trial protocol design, drug/device mechanism of action and evaluation criteria. Readers should be selected based on their experience with the disease process;, expertise in the imaging modality; and availability for trial timelines and training/retraining sessions. Availability of reader may sound trivial but is an important factor as independent read is a time consuming process, most readers already have full-time jobs, and an extension in reading time line can delay the approval process that can be very costly.

Reader training
FDA in their August 2011 draft guidance on medical imaging in clinical trials, and at other occasions and in public forums, has emphasized the need for transparent and standardized training. Differences in clinical and pharmaceutical image evaluations and lack of training about the drug development process mandate pre-training of the readers on basics of utilizing imaging in clinical development. Before commencement of reading, the readers should be trained on GCP, role of imaging in clinical trials, and evaluation criteria including modifications, if any. Images are "interpreted" by the radiologist – a term which acknowledges that differences between readers can occur. The basic training can be performed prior to commencing a blinded read; however, study specific training should be performed as close to beginning the study reads as possible. While technical training on the use of a read system can be performed by technologist or by a project manager, training on image interpretation using specific evaluation criteria must be performed by a peer physician specialist who is well versed not only with the imaging criteria but also with drug development and regulatory guidelines. An operating room nurse does not train a surgeon to perform a hip replacement. Peer expert training is an essential component of ensuring optimal reader performance.

Reader testing and retraining
Once readers are trained, they are tested on sample cases. This gives them an opportunity to clarify ambiguities with the expert peer trainer in the evaluation criteria. Also, this helps the trainer to understand the reading style of the readers. Once reader start reading a study, their performance should be actively monitored by re-reviewing cases to evaluate intra-reader variability, and determine if the readers are following the training rules that were established during the training process. Readers tend to fall back into the clinical reading mode and quite a few times have to be reminded about the pharma read vs. clinical read. When a large drift is detected, the readers are retrained on those specific issues in which they are seen to make. We need to reiterate here that the reader monitoring does not change their evaluations in any way but ensures that they are following rules that were established to prove the efficacy of drug/device.

Final word:
Reader performance is frequently overlooked by the sponsors of the trials because of the perception that there is little difference between the clinical and pharmaceutical practice of radiology. A prospectively designed comprehensive reader performance management program is both an insurance policy and risk management program in one of the last steps in evaluating data from a trial. As they say – the devil is in the details, and if you are unaware of the details of how readers in your trial are managed, the devil may well appears in the results.