Medical Monitor's Conundrum: Making Sense of Site/ Central Discordance in Radiology Assessment

Concerns over disagreements between the interpretation of medical images performed by investigators and those performed by blinded independent reviewers (BICR) accompany most trials that involve imaging biomarkers. The following is intended to be a brief introduction into this topic and provides some guidance on how to manage such disagreements.

What is Site/Central Discordance?

Intra- and inter-reader variability is well known in radiology and continues to exist even with standardised response assessment criteria such as RECIST 1.1, both at the level of the site and central review. In fact, depending on study design and assessment criteria, the variability between readers can range between 30% and 70%. Site/central discordance refers to the discrepancy in the treatment response that is made at the clinic by the treating physician versus the evaluation that's done by BICR at an imaging core lab. Since such discordance is unavoidable, it is critical to proactively have a plan in place to address when it happens.

In 2011, the Oncology Drug Advisory Committee (ODAC) from the FDA provided briefings on an analysis of five oncological trials across different indications (breast cancer, renal cell cancer, and sarcoma). The assessment of progression-free survival (PFS) was compared between the investigator and the BICR. The analysis looked at the differences in the type of response and the timepoint of progression in these trials. Discordance was noted in all five trials, both for type of response and timing of the PFS event. The discrepancy was similar in both arms (experimental and control). Interestingly, the discordance varied not only among the different indications but also within a single indication of breast cancer.

Dr. Peter Eggleton of Merck recently stated at the Calyx Engage industry forum, "I don't see site/central discordance as a problem. I see site/central discordance as a natural consequence of having different processes in place. And if I accept that there is a level of site/central discordance above which you are very uncomfortable, there is also one below which I am very uncomfortable. I was once handed a study with a site/central concordance of 96%. I instituted a deep dive into this because I didn't believe that it was possible for two different processes to come up with such a similar answer. We suspected that information was leaking across from the site to the BICR. Concordance can be too low as well as too high. There is not a competition to get 100% concordance."

Why does it Occur?

Numerous factors contributing to the discordance can broadly be classified into three groups:

- Factors related to the protocol
- Image-read workflow
- Variability in radiology review

Protocol Factors — Inherent Factors Related to the Protocol which could Potentially Bias the Investigator toward Treatment Decisions.

- **Trial design:** In an open-label trial, there is a potential for bias as the investigator is aware of the patient's treatment arm as compared to a double-blind, randomised, controlled trial.
- **Line of therapy:** If the investigational drug is a first line of therapy, the investigators are highly likely to move to an alternative therapy (as against another protocol with patients with no potential alternative therapy) based on very early evidence of clinical progression, even with equivocal imaging findings, since they are accountable for the treatment and wellbeing of the patient. However, the BICR will document findings as equivocal in alignment with the IRC (Imaging Review Charter) guidance.
- **Study indication:** Certain malignancies are known to have incredibly challenging imaging manifestations and associated inherent variability. For instance, ovarian cancer manifests with tiny peritoneal metastases which can be difficult to quantify, leading to high variability.
- **Assessment criteria:** Long-standing, well-established standardised criteria, like RECIST 1.1, will have fewer gaps in interpretation and, hence, disagreement compared to newer, more complex criteria. For example, the Lugano criteria used for lymphoma involve a highly complex multi-modality assessment that is integrated with controlled clinical data, thus potentially more disagreement.

Image-read Workflow – Some Points Highlighting the Differences in Approach to Image Interpretation and Response Assessment between Clinic and BICR.

- **Consistency of readers:** Image reviews at sites are often done at academic centres, where a scan can be read by multiple subspecialised radiologists, including trainees. For example, if a head, neck, chest, abdomen, and pelvis scan is done for a subject, the head and neck may be interpreted by a neuroradiologist, while chest, abdomen, and pelvis may be read by a body imaging expert. Furthermore, the imaging of a given patient may be read by different radiologists across timepoints, based on clinical rotations at the site. A structured, standardised reporting template customised for the trial may also be lacking. At some sites, the radiologist captures the measurements of the tumour in a clinical report, and these are later transcribed by trial coordinators into the electronic data capture system that is customised for the trial. This process depends on the trial coordinator's training and understanding of the assessment criteria and may have the potential for transcription errors as well.

  In contrast, the BICR is performed by a group of radiologists who are specifically trained on trial-specific rules. The assessment for a subject is done by a single reader (or two readers with an adjudication) not only for the entire imaging anatomy, but also across all the timepoints on a highly customised case report form

(CRF) specifically created for the study. These CRFs often have the algorithmic logic built-in to support responses based on the imaging assessments, charter rules and measurements. This standardised process limits the scope of errors, reduces variability, and helps maintain consistency across the trial.

- **Availability of clinical information:** At the site, the radiologist has access to all clinical data (such as physical examination, ECOG score, tumour markers, etc.) and historical imaging while reviewing the current scan. Correlation of clinical data and comparison with historical imaging findings are sometimes crucial and can change the interpretation of an ambiguous or atypical scan. Furthermore, there is also an opportunity for interaction with the treating oncologist in tumour boards, where a multidisciplinary approach is taken to better define the next steps of therapy. This allows for a comprehensive evaluation of the patient using all available information.

  However, by design, the BICR has either no or very limited access to clinical data. This can be a direct cause of disagreement between site and central radiologist. Dr. Cheryl Sadow of Peritus Imaging, who has acted as an independent reviewer on over 200 trials, stated that, *"The challenge central readers face in not having access to clinical information is the need to infer what has transpired during the course of the study. Without knowledge of symptoms or biopsy information to suggest otherwise, in general, the central reader will be conservative in assuming new disease is related to the known malignancy. For example in Figure 1, in the CT images presented for a subject enrolled in a prostate cancer trial, a newly enlarged paraesophageal node (red arrow) would generally be assessed as new disease, despite the atypical location for prostate metastasis. The site reader has access to the development of new clinical symptoms and might suggest a biopsy of this node based on the atypical location, which was subsequently confirmed as a second primary esophageal malignancy."*

- **Training and quality checks:** Lastly, although site radiologists are excellent clinicians, they are not all trained on the intricacies and nuances of various standardised clinical trial assessment criteria. Whereas central readers not only have excellent qualifications as radiologists but are also extensively trained in an ongoing manner throughout the course of the trial on specific criteria for the study. In addition, there is an ongoing quality check of reads that are performed for central reads to ensure the alignment with the charter and protocol guidelines. Training sessions and an ongoing quality check can be a challenge to implement consistently across all sites and hence contribute to the discordance in assessment.

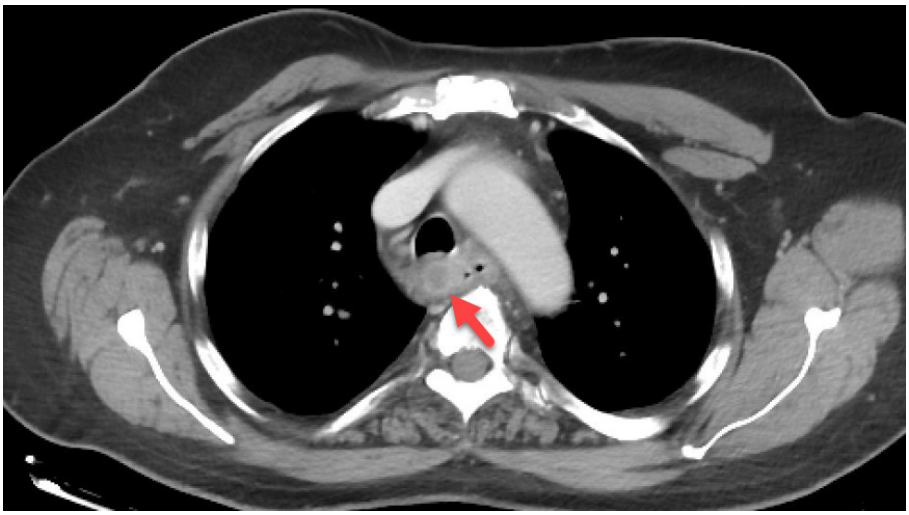


Figure 1: Follow up CT chest in a patient enrolled in advanced prostate cancer trial with evidence of new mediastinal node? Metastasis? Secondary Malignancy. Biopsy needed for confirmation.

It is important to note that differences in the approach of image analysis are due to the differences in the roles and responsibility of a reader at site vs. a core lab. The site reviewers are integral to patient care, including the assessment of complications that can be related to, or unrelated to, therapy. In contrast, central review is intended purely for the assessment of treatment response that is unrelated to the clinical care of patients.

Variability in Radiology Review — Image Interpretation can be Variable among Radiologists.

The elements of this innate variability associated with radiology in the clinical trial setting include differences in lesion selection, inconsistencies in measurements, differences in determining progression based on non-targets, and new lesion selection. In 2010 there was a study analysing the discrepancy rate in interpretation of abdominal and pelvic CTs among experienced radiologists. A total of 90 CTs done for various indications were reviewed and later blinded and re-reviewed by three experienced radiologists. The study concluded that there can be as much as 26% to 32% interobserver and intraobserver discrepancies among the radiologists.[1]

What Can We Do About It?

The management of discordance needs to be addressed in the startup phase of the trial, and adequate measures need to be taken to mitigate discordance and maintain it within a range consistent with study design and assessment criteria. Some of these steps are summarised below:

1) Ensure consistent and high-quality image acquisition:

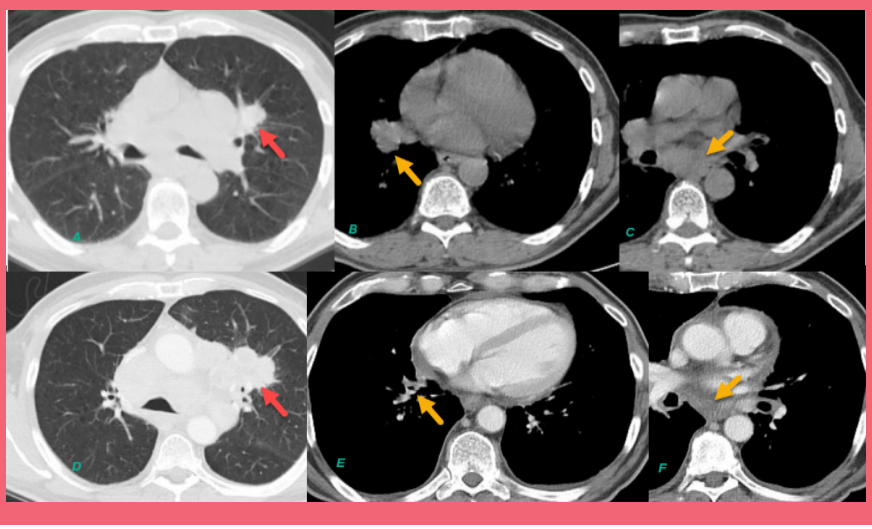
- **A thorough review** of the study protocol by a team of experts at the core lab, including radiologists, ensuring a **comprehensive** inclusion of all the modalities.
- Collaboration between the sponsor and the core lab to provide investigators with protocol-specific imaging guidelines for **standardisation.**
- Implementing robust quality control mechanisms to query for image quality.

2) Tackle complicated assessment criteria by:

- Identifying potential gaps and consulting key opinion leaders to clarify the approach in dealing with such uncertainties.
- Accurately documenting the assessment guidelines both for the central and site radiologists, with training of the trial coordinators on the criteria.

Practical Example — Potential Differences in NSCLC Read Outcomes

"Axial CT chest images of a subject enrolled in non-small cell lung cancer trial, baseline images (top row) and follow-up images (bottom row). (A) Baseline CT image in lung window shows a left upper lobe lung mass (red arrow) selected as target disease; two mediastinal nodes (yellow arrows) seen on soft tissue window (B, C) were selected as non-target disease by the site reviewer due to lack of intravenous contrast. On the follow-up, (D) CT image in lung windows shows the target lung mass increased significantly resulting in progressive disease by target disease, even though the non-target mediastinal nodes had decreased in size as seen on the soft tissue window (E, F). For the central review, due to strict guidelines and ongoing reminders for measuring as many lesions as possible, the central reviewer measured lung mass and mediastinal nodes as target disease at baseline, to include maximum target lesions possible. Hence despite the increase in size of lung disease, because the mediastinal nodes had decreased in size significantly, overall assessment was partial response on follow-up." - Dr. Nisha Sainani



Conclusion

Variability within radiology review is well understood. Site/central discordance is inevitable, and its existence needn't negatively impact trial outcomes. A discordance within acceptable range (based on the indication and assessment criteria) is expected and, in fact, indicates that the site/central reading systems are working independently without any bias. However, proactive steps need to be taken to manage discordance within an acceptable range. It is crucial to evaluate the discordance early on and implement mitigation steps at the initiation, as well as in an ongoing fashion during the course of the study. Ignoring the reality of site/central discordance or overreacting to control it will never work; instead, awareness and understanding of the risks and how to mitigate them can best set your study for success.

David Leung of Bristol-Myers Squibb recently stated,

"Some tend to think of discordance as an intrinsically bad thing: an error, a sign of human imperfection. But imagine a miracle drug that cures all, resulting in more than a 95% shrinkage of all tumours. Conversely, imagine another drug that is terrible where all tumours clearly progress. In both cases you will find no discordance between any readers, site or central. In real life, however, we have drugs that are somewhere in between. There will be subtleties that are difficult to interpret and will result in differences in interpretation among readers. So, I firmly believe that as long as we understand the reason for discrepancy, we will find that in most cases they are not caused by error or human imperfection, but by challenges of judgement and limitations of criteria."

Contact hello@calyx.ai for more information

REFERENCES

1. Abujudeh et al; Abdominal and pelvic computed tomography (CT) interpretation: discrepancy rates among experienced radiologists. Eur Radiol. 2010 Aug;20(8):1952-7.

Dr. Surabhi Bajpai

Calyx's Dr Surabhi Bajpai is a board-certified radiologist with over 12 years of radiology experience. Before Calyx, she spent four years as a research fellow in Abdominal Imaging, Massachusetts General Hospital, Boston which is affiliated with Harvard University.

Dr. Manish Sharma

Calyx's Dr. Manish Sharma is a board-certified radiologist with over 17 years of experience in medical imaging and clinical trials currently focused on evolving reader training, variability, and monitoring using data-driven analytics.