Analysis of the Cause of Discordance Between Two Radiologists in the Assessment of Radiographic Response and Progression for Subjects Enrolled in Breast Cancer Clinical Trials Employing Blinded Independent Central Review

Kristin Borradaile MS1, Robert Ford MD1, Michael O’Neal MD1, Kevin Byrne MD1

When radiologists function as independent reviewers, there is the potential for them to identify different lesions and thereby assess disease extent at different rates. There is also the potential for discordance between independent central review (BICR) of radiographic exams for registrational oncology studies when the primary endpoint is based on tumor measurements such as lesion size, shape, and density. This discordance may affect the evaluation of the subject's overall extent of disease. There is also the potential for discordance between BICR radiologists.

Methods

BICR data was blinded and pooled to identify cases in which the two primary readers were discordant in outcome (overall best response, best response at any time point, or progression). Radiographs from 459 subjects were reviewed to determine the cause of discordance and whether it resulted from a justifiable interpretation difference where neither reader was incorrect in their assessment, or if it resulted from an assessment error by one reader. We acknowledge there may have been bias introduced into this process, as the interpretations were judged by radiologists from the same facility as the original reviewers.

Results

In 37% of cases (168/459), discordance resulted from a difference in lesion selection between readers. In 137 cases, the difference was justifiable, but in 31, the choice of lesions by one reader was not thought to represent the overall extent of disease at baseline. In 30% of cases (139/459), discordance resulted from a difference in the perception of new lesions. In 88 cases, the difference was justifiable, but in 51, one reader's assessment was judged to be incorrect during review. In 13% of cases (58/459), discordance resulted from a difference in the qualitative assessment of progressive disease (PD) based on non-target (NT) disease. In 58 cases, the difference was justifiable, but in 19, one reader's assessment was deemed incorrect during review. In 58% of cases (262/459), image quality issues resulted in a difference in a justifiable interpretation difference between readers. In 9% of cases (41/459), the difference resulted from a difference in lesion measurements between readers. In 37 cases, the difference was justifiable, but in 4, one reader's measurements were deemed incorrect during review. In 1 case, an error resulted from a lack of clinical information during the review. Examples of justifiable discordance are provided.

Discussion

Lesion Selection: When radiologists function as independent reviewers, there is the potential for them to identify different lesion targets. This may affect the evaluation of the subject's overall extent of disease. There is also the potential for the reviewers to identify the same lesions but classify them differently between target and non-target lesions. This may affect the evaluation of the subject's overall extent of disease.

Perception Differences: Due to the inherent subjective nature of radiographic assessments, the presence of image artifacts, and the variability in the interpretation of radiographic findings, there is the potential for discordance between radiologists.

Conclusions

Some factors that cause discordance are process-driven or due to justifiable interpretation differences. This was observed in 77% of cases in this study (554/459). Examples include lesion selection, image quality issues, inter-reader measurement variability, perception of new lesions, and the assessment of NT-PD. Some of these factors may contribute to discordance, such as those identified in 23% of cases (105/459). This compares favorably to perceived naturalistic inter-reader and inter-laboratory discordance rates. The BICR process of adjudication by a third reader identifies and mitigates these justifiable discordances and interpretive errors.