Double reading performance and the impact of adjudication on progression-free survival estimations: Findings from a lung clinical trial.

Hubert Beaumont, Antoine Iannessi; Median Technologies, Valbonne, France

Background: The FDA recommends Blinded Independent Central Review (BICR) with double read for clinical trials with imaging. However, inter-reader variability is a concern in these trials. While studies have investigated the variability of RECIST, the primary response criteria, little attention has been given to the optimal association of readers. The evaluation of therapeutic response in phase III trials relies on the Date of first Progressive Disease (DoPD), with the Discrepancy Rate (DR) as the preferred index for measuring inter-reader variability in DoPD evaluation. Another important index measures readers' bias, assessing their tendency to over or under-estimate diagnoses. In cases of discrepancies, a third reader is brought in for adjudication. However, the impact of adjudication on trials' Progression-Free Survival (PFS) is not well-documented. Our study examines the variability in a lung clinical trial using RECIST, analyzing double reading performance, reader association prediction, and the impact of adjudication on PFS estimations. **Methods:** We retrospectively analyzed five phase III lung clinical trials using the RECIST 1.1 criteria in BICR with double reads. The trials involved 7 readers organized into 11 teams, each reader having participated in multiple trials and performed over 50 reads, resulting in 1017 patients' reviews. Our analysis included: Calculation of DR and bias for each team, and testing the correlation between DR and bias. Computation of the signed bias for each individual reader. Evaluation of a probabilistic model to predict the DR for each team and the bias for each reader. Comparison of PFS between single and double reads after adjudication and endorsement of one of the readings. **Results:** A multiple comparisons test did not reveal any difference between teams' DR (Marascuilo test; q > 0.05). The average DR across all teams was 39.9% [95%CI; 37.8; 41.9]. However, we did find significant differences in bias when comparing 9/55 pairs of teams (Marascuilo test; q < 0.05). The range of absolute bias values was 20% to 100%. We did not find a correlation between bias and DR (p = 0.64). Additionally, when comparing the average bias value per reader, no differences were observed (Marascuilo test; q < 0.05). We failed to predict teams DR and readers' bias. The endorsement rate of readers ranged [18%; 82%]. After adjudication, we found that 27% of the PFS were lower than the minimum value obtained from the single readers, in one case 20.6% shorter. **Conclusions:** Significant readers' bias has a notable impact on double readings, independent of the DR values. The performance of one reader cannot be generalized based on others. Additionally, adjudication significantly affects the PFS of clinical trials. These findings emphasize the importance of considering readers' bias and the potential consequences of adjudication in clinical trial assessments. Research Sponsor: None.