

Modified RECIST for Monitoring Hepatocellular Carcinoma with Computed Tomography: Inter-reader Variability of the Response

Beaumont H¹, Egels S², Iannessi A³, Bonnard E³, Foley C⁴, Lucidarme O².

¹ Median Technologies, Valbonne, FR. ² Hopital Pitié Salpêtrière, Paris FR. ³ Centre Antoine Lacassagne, Nice, FR.

⁴ GlaxoSmithKline, Brentford, UK

Hepatocellular carcinoma (HCC) is the most common primary liver cancer. In clinical trials, computer tomography (CT) is a widely used imaging technique for the monitoring of HCC patients, and modified RECIST (mRECIST) have been suggested as appropriate criteria in part because they distinguish the viable part from the necrotic part of the tumor.

Based on longitudinal changes of tumor burden, this study investigates the causes of inter-reader variability in evaluating the therapeutic response of HCC when relying on mRECIST.

24 patients with advanced (unresectable and/or metastatic) HCC enrolled in a phase I/II multicentre international study were retrospectively reviewed. From the originally selected target lesions, 41 were randomly selected. Original mRECIST expert evaluations were aggregated to the retrospective readings of three radiologists having different expertise (non mRECIST experts). All the 150 readings performed by each readers were made in using an electronic calliper. Precision of measurements between couple of readers was analysed by assessing standard deviation (SD) and reproducibility coefficient (RDC). The agreement of readers responses were compared by using Kappa coefficient statistic. The variability between non-experts and the variability between mRECIST expert and non-experts were analysed. Reader's disagreement at declaring progressive (PD) and responding (PR) patient, were visually classified between variability of the measure, difference in the perception of tumour boundaries and differences in using either RECIST or mRECIST criteria for a given lesion.

SD of measurements between non-experts ranged [24.9%; 36.3%] and RDC was 16.6 [13.85; 23.95]. Kappa coefficients was 0.41 [0.28; 0.55]. SD of expert against non-experts ranged [33.2%; 41.1%] and RDC was 18.8 [16.02; 24.53]. Kappa coefficients was 0.20 [0.06; 0.35].

Pooling the four readers together, rate of discrepancy at declaring respectively PD or PR by patients was 47.8% (11/23) and 34.8% (8/23). 90.9% (10/11) of discrepancies at declaring PD were due to different perceptions of tumour boundaries. Discrepancy at declaring PR was, in 62.5% of the cases, correlated to the application of mRECIST, 25% was correlated to different perception of tumours boundaries.

Inter-reader variability of HCC measurements was large which leads to poor agreement in the response. The main cause of discrepancy at declaring PD originated from the complexity of HCC patterns and the poor definition of tumours boundaries while discrepancies at detecting PR come from

the variability of readers at selecting the viable part of the tumours. In the context of HCC clinical trials, reliability of endpoints depend on the criteria involved. Present study must be completed with an inter-reader analysis of the variability between mRECIST experts.