

BLINDED INDEPENDENT CENTRAL REVIEW OF ONCOLOGY TRIALS: THE MONITORING OF READER PERFORMANCE

Hubert Beaumont¹, Antoine Iannessi¹, Jennifer Cillario¹, Yan Liu¹

¹ Median Technologies, Valbonne, France

BACKGROUND

For image-based evaluations, RECIST 1.1 [1] remains the most used criteria for assessing therapeutic response in clinical trials of solid tumors. The variabilities of evaluations are generally mitigated by double reading the images with a third reader adjudicating any discrepancies [2]. However, blinded

independent central review (BICR) with double read and adjudication is a complex process that needs to be closely monitored. The rate of inter-reader discrepancies is one of the metrics of choice for detecting quality issues in trials [3].

OBJECTIVE

To provide reference value metrics to monitor reader performance with double read plus adjudication in clinical trials using RECIST 1.1.

METHOD

From the list of clinical trials recorded in our database, we selected a subset of trials according to the following inclusion criteria:

- 1) Response criteria: RECIST 1.1
- 2) Phase II and III
- 3) Trial setting: BICR + adjudication.
- 4) Trial endpoint: Overall Response Rate, Progression Free Survival and Date of Progression (DoP)
- 5) Reader monitoring enabled.

We analyzed, per trial and per reader, the rate of inter-reader discrepancies and the rate of reader endorsement by the adjudicator.

The adjudicator analyzed the double reads and endorsed the most pertinent evaluation.

We compared the discrepancy rate between indications using the Marascuillo procedure.

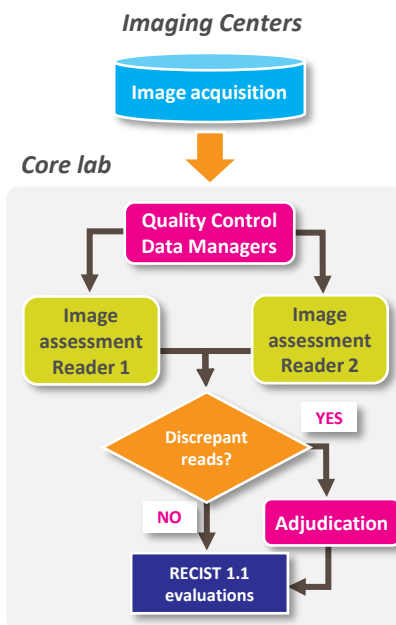


Figure 1: Read diagram – double reader with adjudication

Out of our database, 5 trials conformed to the inclusion criteria. Their indications were: Lung (1), Skin (1), Biliary track (1), Gastric (1) and Multiple (1) solid tumors. A total of 1,561 patients (mean=312/trial) and 5,986 time points (mean=1,197/trial) were analyzed by 12 readers; 8 adjudicators were involved.

Trial ID	Indication	Nb. Pat (N)	Nb TP/Pat (N)	Disc./Pat (%)
Trial 1	Skin	108	7.07	59.3
Trial 2	Lung	688	3.84	53.1
Trial 3	Gastric	371	2.08	44.7
Trial 4	Biliary track	287	5.23	63.8
Trial 5	Multiple	107	2.82	33.0

Table 1: Discrepancy rate for our selected clinical trials was computed considering that at least one time point (TP) response per patient (Pat) was discrepant.

- Per reader, the discrepancy rate ranged from 27.4% to 68.5% (mean=50.1%). Per trial, the mean discrepancy rate was 50.8% (range=33.0-63.8%).
- The discrepancy rate was found to be significantly different between indications: Biliary (63.8%) vs Multiple cancers (33.0%) ($p < 0.001$).

Figure 2: Example of inter-reader discrepancies.

Change of tumor burden assessed by two different readers: Due to measurement variation or target lesion selection at baseline, one reader declared an earlier partial response (PR).

—◆— Diameters
— Partial Response



RESULTS

Table 2: Reader endorsement rate.

For 3 of the 5 trials, adjudications were triggered for discrepant DoP and Date of Response. For each of the 11 readers involved in these trials, we calculated their mean endorsement rate.

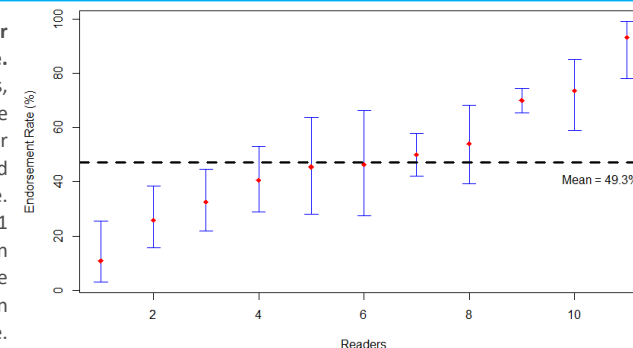
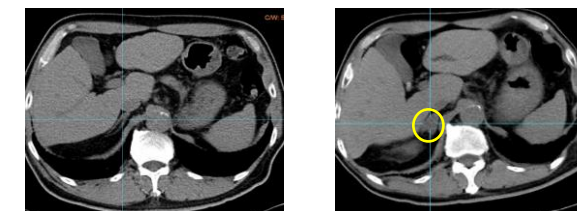


Figure 3: Example of inter-reader discrepancies.

One new lesion (yellow circled) was missed by one of the readers.



CONCLUSIONS

- **Observed discrepancy rate** varies with trial indication from 30 to 60% and increases with the number of time points/patients.
- **Reader Endorsement Rate** is key to ensuring consistent readings and to triggering corrective actions
- **Next: Adjudication rate** should be considered with respect to trial endpoints

REFERENCES

- [1] Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009; 45(2):228-247. doi:10.1016/j.ejca.2008.10.026
- [2] Clinical Trial Imaging Endpoint Process Standards, Food and Drug Administration – guidance for industry, April 2018.
- [3] Cohen KL, Monitoring Reader Metrics in Blinded Independent Central Review of Oncology Studies. *J Clin Trials*. 2015; 5:4