CrossMark

# PET-guided clinical trials in Hodgkin lymphoma: to agree or not to agree, that is the reviewer's question

A. Gallamini[1] · M. Meignan[2]

In image-based clinical trials, when the study endpoint depends on treatment response assessment by imaging, blinded independent central review (BICR) is considered mandatory to overcome potential image reader bias leading to a systematic over- or under-interpretation of tumor shrinkage [1, 2].

[18]F–Fluorodeoxyglucose Positron Emission Tomography (FDG-PET) performed interim during treatment (iPET) proved the most accurate tool to predict doxorubicin-vinblastine-bleomycin-dacarbazine (ABVD) treatment outcome in Hodgkin Lymphoma (HL) [3, 4] and several trials have been published in which treatment of this disorder was iPET-adapted [5–12]. This turned out to be feasible as soon as new simple and reproducible rules for qualitative PET/CT scan interpretation by visual assessment became available for clinical trials [4], the so-called Deauville 5-point scale (DS). Moving from a binary (positive/negative) to a discrete scale such as DS, variability among reviewers increased, but this was offset by an enormous advantage for patient management [13].

Nonetheless, reproducibility of DS scoring proved good or very good, albeit with some exceptions. In the RATHL study, for example, out of 51 patients scored 4 by local investigator, only 34 (66%) were reclassified as score 4 in the central core laboratory of the study by consensus review [14]. Another aspect impacting reviewer concordance is the availability of a set of practical instruction for a stepwise method to proceed on the review and to exclude from the analysis the most common sources of false-positive results [15].

A central question is how the final decision is taken in case of discrepancy among readers: two methods of final report adjudication are consensus or independent decision. In central consensus review, the final decision is made after discussion between a couple of reviewers (as in UK studies) or among all the members of the entire panel (as in US and German trials) of reviewers including in some trials hematologists or radiotherapists. In both cases the reviewer concordance rate can be calculated before the final report adjudication. In BICR the final judgment is taken simply by an arithmetical count of the majority of agreed opinions and, very important, reviewers do not influence each other in taking the final decision.

The choice of BICR for central PET review by the core lab of some cooperative lymphoma groups was based on the assumption that (1) "true" discordance among reviewer does exist in some difficult cases; (2) consensus among reviewers is not a pre-requisite for final result attribution; (3) disagreement among reviewers should be tracked; (4) as there is no limit, in theory, on the number of reviewers in BICR, the higher the reviewer number, the lower is variability of the method; and (5) in consensus central review the logistic aspects of a face to face or telephone call meeting are sometimes a true hurdle when a result of PET review is expected within 48 h from image upload. Nonetheless, in both review systems the reproducibility of the method is warranted by some indices such as Cohen's k [16] or Krippendorf's alpha [17], reporting the concordance rate between a couple of the entire panel of reviewers, respectively. In consensus review, however, the flaw of a modified final judgment by one or more reviewers blunts the relevance of these indexes.

The UK PET center network adopted a central consensus review for the National Cancer Research Institute (NCRI)

✉ A. Gallamini
andreagallamini@gmail.com;
andrea.GALLAMINI@nice.unicancer.fr

1 Research, innovation and statistics department, Lacassagne Cancer Center, 33 Rue de Valombrose, 06129 Nice, France

2 LYSA imaging center, Hôpital Henri Mondor, Creteil, France

trials on HL: the RAPID [5] and RATHL [9] trials. The images were transmitted to the core laboratory (Cole Lab) at St. Thomas' Hospital, King's College, London, for central review. Two experienced reporters independently scored the scans with the use of DS. Differences in opinion, if any, were resolved by consensus. In the RATHL trial a network of national core laboratories in the United Kingdom, Italy, Sweden, Denmark, and Australia reported the PET scan images using DS, adopting a mixed independent and consensus review. Two readers at each local core lab, unaware of the patient's clinical status, scored independently the scans and disagreement was resolved by consensus reading, and, in the rare case of persisting disagreement, a third doctor from another core lab adjudicated the scan result [9, 13, 14]. A similar approach has been successfully adopted by the US Alliance group in the S0816 trial [10], where PET/CT scans were submitted for central review to the CALGB (Cancer and Leukemia Group B) imaging core lab. The latter endeavored Internet-based visual and virtual conferences that allowed the simultaneous display of images and mutual communication between participating sites and the core lab in a secure manner. The central PET/CT review was completed in less than 2 days in 78% and in less than 4 days in 95% of the patients. As in NCRI trials there was one adjudicator in the CALGB Core Lab, for cases where major discrepancies existed between the local site and the central PET/CT interpretation. Similar to UK and US trials, in the German HD15 trial by the German Hodgkin Lymphoma Study Group (GHSG), a multidisciplinary panel consisting of a medical oncologist, a radiologist, a radiation oncologist, and a nuclear medicine physician, accompanied by a statistician, reviewed all PET/CT and CT scans as well as any available x-rays. However, different from UK and US core labs, PET/CT comparison of iPET with baseline PET was not possible as in the GHSH trials only one PET scan was funded. In the HD 15 trial the images were interpreted by a modified DS system using the mediastinal blood pool structures as a reference background for a positive scan. The central review panel in consensus made the final iPET adjudication [18].

The Core Laboratories of the French LYSA (Lymphoma Study Association), the Italian FIL (Italian Foundation on Lymphoma) and of EORTC (European Organization for Research and Treatment of Cancer) had a totally different approach from all the above reported studies, as BICR was adopted for iPET central review. The EORTC first pioneered in the H10 trial the use of BICR for central iPET reading [7]. In this trial, for technical reasons, centralized reviews for the LYSA group started from the trial onset, while for EORTC and FIL groups it was one year later. The LYSA group (formerly GELA) pioneered an on-line reading system through a network of workstations (WS) across LYSA PET sites physically wired by a virtual private network (VPN), commercialized by Keosys® (Saint Herblain, France). Images were distributed to six experts to be reported on screens displaying images with the same color scale and generated by the same software. The final result (a mathematical calculation of the local nuclear medicine physician and of two, four, or six experts readings) was returned to the peripheral site within 72 h from image upload. Later on, in the LYSA AHL 2011 trial, the exchange tool was no longer a VPN WS network, but instead a web-based platform by Imagys®: images uploaded in the system were readily available without the need of image downloading and could be reported online everywhere with the same software by three expert reviewers on their personal computers [19]. In more than 90% of the cases the result of the scan was posted to the peripheral clinical sites in 48 h.

Similar to the EORTC and LYSA platform, in the FIL HD0607 trial [8] readers reviewed independently the iPET images and inserted the review in the WIDEN® system (Dixit, Torino, Italy). The latter is a web-based platform that calculated automatically the final result of the review by the majority of concordant scores and forwarded the result of the review to the clinical sites participating in the study. Real-time independent review was carried out; the average and median times for diagnosis exchange were 48 h and 38 h, respectively [20]. LYSA and FIL central image review systems are similar but differ in the image displaying system in that for the LYSA imaging platform the use of the same software (viewer) allows an identical image display through all the workstations of the platform, while in the FIL WIDEN® system images are transferred by DICOM transfer protocol and, more importantly, reviewers report the images as they are accustomed to in the daily practice in their workstations. Moreover, LYSA expert review panel consists in a group of trained experts that was created in 2007 when central image reviewing for clinical trials was first set-up; however, for the time being, a generational turnover of newly trained experts is lacking. This problem has been originally solved by the FIL imaging commission, along with that of skill dissemination across the nuclear medicine (NM) community. New NM experts are taught and trained for the definite task required by the study protocol, with a training set of PET scan images similar to those to be reported in the trial, A "learning curve", obtained by the reported PET scan images with increasing skill and self-confidence from the first reviewed images till the last reported ones, is available in the website of WIDEN® to document the specific skill reached by the NM experts [21]. In conclusion, the LYSA and FIL central review methods have been conceived with the multidisciplinary contribution of clinicians, NM experts, physicists, engineers and biostatisticians for academic studies, with adoption of BICR to overcome source of errors among reviewers and facilitate the skill dissemination in the NM community. This was obtained in both groups thanks to a continuous recruitment or new NM reviewers; in other lymphoma research groups, adopting consensus instead of independent central image review, the

review performance proved very good, but the problem of expert turnover and skill dissemination still remains an unmet need.

## REFERENCES

1. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? J Clin Oncol. 2008;26:3791–6.

2. Amit O, Bushnell W, Dodd L, et al. Blinded Indipendent central review of progression-free survival endpoint. Oncologist. 2010;15: 492–5.

3. Cheson BD, Fisher RI, Barrington SF, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. J Clin Oncol. 2014;32(27):3059–68.

4. Barrington SF, Mikhaeel NG, Kostakoglu L, et al. Role of imaging in the staging and response assessment of lymphoma: consensus of the international conference on malignant lymphoma working group. J Clin Oncol. 2014;32:3048–58.

5. Radford J, Illidge T, Counsell N, et al. Results of a trial of PET-directed therapy for early-stage Hodgkin's lymphoma. N Engl J Med. 2015;372:1598–607.

6. Straus DJ, Pitcher B, Kostakoglu L, et al. Initial Results of US Intergroup Trial of Response-Adapted Chemotherapy or Chemotherapy/Radiation Therapy Based on PET for Non-Bulky Stage I and II Hodgkin Lymphoma (HL) (CALGB/Alliance 50604). Blood. 2015;126:578. [abstr.]

7. André MPE, Girinsky T, Federico M, et al. Early positron emission tomography response-adapted treatment of stage I and II Hodgkin Lymphoma: final results of the randomized EORTC/LYSA/FIL H10 trial. J Clin Oncol. 2017;35(16):1786–94.

8. Gallamini A, Rossi A, Patti C, et al. Interim PET-adapted chemotherapy in advanced Hodgkin Lymphoma (HL). Results of the second interim analysis of the Italian GITIL/FIL HD 0607 trial. Hematol Oncol. 2015;33(Suppl. 1 June 2015):163. Abstract 118

9. Johnson P, Federico M, Kirkwood A, et al. Adapted treatment guided by interim PET CT scan in advanced Hodgkin's lymphoma. N Engl J Med. 2016;374:2419–29.

10. Press OW, Li H, Shoder H, et al. US Intergroup trial of response adapted for stage III to IV Hodgkin Lymphoma using early interim Fluorodeoxyglucose -Positron Emission Tomography imaging: Southwest Oncology Group S 0816. J Clin Oncol. 2016;34:2020–7.

11. Borchmann P, Havenkamp H, Lohri A, et al. Progression-free survival of early interim PET-positive patients with advanced-stage Hodgkin's lymphoma treated BEACOPP ESCALATED alone or in combination with Rituxinab (HD18): an open-label, international randomized phase 3study by the German Hodgkin Study Group. Lancet Oncol. 2017;18(4):454–63.

12. Casasnovas O, Brice P, Bouabdallah R, et al. Randomized phase III trial comparing an early PET driven treatment de-escalation to a not PET-monitored strategy in patients with advanced stages Hodgkin Lymphoma: interim analysis of the AHL2011 Lysa Study. Blood. 2015;126:577. [abstr.]

13. Barrington SF, Qian W, Somer EJ, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin Lymphoma. Eur J Nucl Med Mol Imaging. 2010;37(10):1824–33.

14. Barrington SF, Kirkwood AA, Franceschetto A, et al. PET-CT for staging and early response: results from the response-adapted therapy in advanced Hodgkin Lymphoma study. Blood. 2016;127(12): 1531–8.

15. Biggi A, Gallamini A, Chauvie S, et al. International validation study for interim PET in ABVD-treated, advanced stage Hodgkin lymphoma: interpretation criteria and concordance rate among reviewers. J Nucl Med. 2013;54:683–90.

16. Cohen J. A coefficient for agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.

17. Hayes AS, Krippendorf K. Answering the call for a standard reliability measure for coding data. Commun Method Meas. 2007;1: 77–89.

18. Kobe C, Kuhnert G, Kahraman D, et al. Assessment of tumor size reduction improves outcome prediction of positron emission tomography/computed tomography after chemotherapy in advanced-stage Hodgkin lymphoma. J Clin Oncol. 2014;32: 1776–81.

19. Meignan M, Itti E, Bardet S, et al. Development and application of a real-time on-line blinded independent central review of interim PET scans to determine treatment allocation in lymphoma trials. J Clin Oncol. 2009;27:2739–41.

20. Chauvie S, Biggi A, Stancu A et al: WIDEN: a tool for medical image management in clinical trials. Clinical Trial 2014;0:1–7.

21. Ceriani L, Barrington S, Biggi A, et al. Training improves the interobserver agreement of the expert positron emission tomography review panel in primary mediastinal B-cell lymphoma: interim analysis in the ongoing International Extranodal Lymphoma Study Group-37 study. Hematol Oncol. 2016. https://doi.org/10.1002/hon.2339.