Supplemental Appendix

Article Selection Process

Our literature search using MEDLINE and EMBASE identified 585 and 595 potentially relevant articles, respectively (Supplemental Fig. 1). We excluded 446 studies for the MEDLINE search and 470 studies for the EMBASE search by scanning the titles and abstracts. We then retrieved and reviewed 160 full reports for inclusion and excluded 108 studies: 47 studies using FDG PET for initial staging, 17 studies using FDG PET for monitoring response to treatment as early prognostic tool, 4 studies using FDG PET as pre–high-dose-chemotherapy assessment, 1 study for the use of γ-camera coincidence methodology for FDG imaging (S1), and 39 studies for other reasons. We made correspondence attempts to the primary investigators of 43 studies that did not meet all the inclusion criteria but had some relevant participants: 13 studies including mixed categories of patients with HD and NHL (26,27,29,S2–S11), 11 studies enrolling patients who completed first-line therapy as well as salvage therapy (28,31,34,S12–S19), 9 studies including multiple FDG PET results for some patients (30,32,33,S20–S25), 5 studies using insufficient data to calculate diagnostic accuracy estimates (25,S26–S28), 5 studies with different unit of analysis (S29–S32), and one study
including NHL with unknown histologic subcategories (S33). Twenty-nine paper authors (69%) responded, and we included 11 studies: 2 studies with relevant subgroup data available from the published papers (26,29) and 9 studies for which the principal investigators provided relevant unpublished data (25,27,28,30–34). No additional studies were identified through the SCOPUS and Biological Abstracts search or a manual search of the bibliography of the reviewed studies, review articles, or textbooks.

Quality Assessment of Published Studies

Only a few studies appropriately recruited the study participants, making the applicability of the study results to clinical practice limited (37). More than half the studies did not report the information regarding whether the test was interpreted independently from reference standard, and most studies did not describe whether the interpreters of the reference standard were blinded to test results. This absence of blinding may lead to overestimation of diagnostic accuracy (37). Six studies clearly incorporated PET results into reference standard, and 11 studies did not report whether the PET results were considered for the final assessment of patients during clinical follow-up. This incorporation of test result into reference standard may also lead to overestimation of diagnostic accuracy (37). In five studies, several patients
received radiotherapy after undergoing PET (17,20,24,25,28), and another study included a patient who completed the last cycle of chemotherapy after undergoing PET (18). These studies have possibility of “treatment paradox” (S34), which may shift some patients categorized as TP to FP as well as FN to TN, affecting diagnostic accuracy. In a “worst case” scenario, if the added therapeutic interventions were perfectly effective and administered exclusively to patients with residual disease with negative PET results, this would make exclusive shifts from FN to TN, leading to overestimation of both sensitivity and specificity.

Further, although we considered clinical follow-up the best available reference standard (S35) as the follow-up reported in the literature is relatively short for malignant lymphoma, this could miss late relapses to affect both sensitivity and specificity (37). Most studies did describe key points on the test procedures of PET to reproduce in clinical practice, while most studies did not report how they followed-up patients, or whether or not they followed formerly recommended response assessment guidelines such as the Cotswalds criteria for HL (S36) or the International Workshop criteria for NHL (2), decreasing the applicability of the studies (37).

Only four studies reported whether they adopted indeterminate category to interpret PET results. The absence of reporting indeterminate or uninterpretable results may distort the diagnostic accuracy as well as lose information for the better understanding of the test.
Supplemental References


tomography using fluorine-18-fluorodeoxyglucose in patients treated for malignant


S7. Kumar R, Xiu Y, Potenta S, et al. 18F-FDG PET for evaluation of the treatment response

tomography and magnetic resonance imaging useful in the prediction of relapse in

tomography (PET) with respect to computed tomography in the follow-up of lymphoma

tomography (PET) in the management of lymphoma patients. *Ann Oncol*.


S22. de Wit M, Bohuslavizki KH, Buchert R, Bumann D, Clausen M, Hossfeld DK.


S26. Steinert HC. PET/CT in lymphoma patients. *Radiologe* [German].


S27. Cui RX, Zhou Q. 18FDG PET in the management of malignant lymphoma.


