

¹⁸F-FDG PET을 이용한 림프종 치료 반응 평가: ¹⁸F-FDG PET의 진단 성능 특성과 구간 우도비

원광대학교 의과대학 핵의학교실¹, 내과학교실²
김창근¹ · 김대웅¹ · 박무림²

Assessment of Tumor Response to Therapy in Lymphoma Using ¹⁸F-FDG PET: Diagnostic Performance of ¹⁸F-FDG PET and Interval Likelihood Ratio

Chang Guhn Kim, M.D.¹, Dae-Weung Kim, M.D.¹, and Moo Rim Park, M.D.²

Departments of ¹Nuclear Medicine and ²Internal Medicine, Wonkwang University School of Medicine, Iksan, Jeonbuk, Korea

In this paper, the authors intended to summarize briefly the features of lymphoma with regard to ¹⁸F-FDG PET for assessment of tumor response to therapy, to describe why assessment of treatment response should be performed, to review what method so far has been used in monitoring treatment response, to discuss what limitations of morphologic imaging criteria for assessing tumor response are, in compared with ¹⁸F-FDG PET, and to introduce recently proposed criteria for assessing tumor response in malignant lymphoma. And also the authors emphasize the need to understand the characteristics of diagnostic performance of ¹⁸F-FDG PET in several clinical settings in order to interpret ¹⁸F-FDG PET results appropriately, and to encourage the use of interval likelihood ratio to enhance clinical implications of test results which, in turns, allows referring physicians to understand the meaning of interpretation with easy. Until recently, treatment response has been assessed according to the morphologic criteria. Metabolic imaging with ¹⁸F-FDG PET was adopted to have important role for treatment assessment in IWC+PET criteria proposed recently by IHP. To accomplish this role, we should perform and interpret ¹⁸F-FDG PET according to IWC+PET criteria. It is important for referring physicians to understand the various limitations of ¹⁸F-FDG PET and pitfalls in PET interpretation, and to understand that clinical information are needed by nuclear medicine physicians to optimize the interpretation of ¹⁸F-FDG PET. (Nucl Med Mol Imaging 2009;43(5):369-385)

Key Words: Lymphoma, therapy response, positron emission tomography, fluorodeoxyglucose.

서 론

이 글에서 저자들은 치료 반응 평가를 위한 ¹⁸F-FDG PET과 관련하여 림프종의 특성은 무엇인지, 종양의 치료 반응 평가는 왜 하는지, 치료 반응 평가는 지금까지 어떻게 해 왔는지, 형태학적 변화에 바탕을 둔 기존의 반응 평가의 제한점과 ¹⁸F-FDG PET의 장점에 대해 기술하고, 최근에 권고된 림프종의 치료 반응 평가 기준을 소개하고자 한다.

또한, ¹⁸F-FDG PET 진단 성능의 특성을 구체적인 예시를 들어 설명하고 우도비(likelihood ratio)에 대해 기술하고자 한다. 진단 성능의 특성과 우도비 개념을 판독에 반영함으로써 임상 의사가 검사 결과의 임상적 의미를 보다 잘 이해하게 될 것이다. 림프종의 치료 반응 평가 기준을 구체적으로 제시하거나, 여러 연구자들의 치료 반응 평가 기준을 비교, 분석하는 것이 이 글의 목적이 아니다. 이 글을 준비하는 동안 치료 반응 평가를 주제로 한 Journal of Nuclear Medicine 2009년도 5월호 보충판이 발행되었고, 좋은 논문들이 수록되어 있다. 그 중에서도 Weber와 Wahl의 논문을 독자들이 참조하시기 바란다.

림프종의 특성

악성 림프종은 역사적으로는 Reed-Sternberg (RS) 세포

• Received: 2009. 9. 20. • Revised: 2009. 9. 25.
• Accepted: 2009. 9. 30.
• Address for reprints: Chang Guhn, Kim, M.D., Department of Nuclear Medicine, Wonkwang University School of Medicine, 344-2 Shinyong-Dong, Iksan 570-711, Korea
Tel: 82-63-859-1821, Fax: 82-63-852-1310
E-mail: leokim@wonkwang.ac.kr
※ 이 논문은 2009학년도 원광대학교 교비 지원에 의해서 연구되었음.

유무에 따라 호지킨 림프종과 비호지킨 림프종으로 크게 구분하였다. 그러나 병리조직학의 발전에 따라 호지킨 림프종은 RS 세포가 관찰되는 고전적인 호지킨 림프종과 림프구 세포 또는 lymphocyte predominant (LP) 세포가 관찰되는 nodular lymphocyte predominant 호지킨 림프종으로 구분되게 되었다. 비호지킨 림프종 또한 1990년대까지 세포의 크기, 여포 구조의 유무 및 임상 경과에 따른 Working Formulation 분류를 사용하였으나 면역조직화학 염색과 세포유전학, 분자생물학이 발전하면서 기존의 분류법으로는 명명할 수 없는 새로운 아형들이 보고되었고, 동일한 형태의 조직형 일지라도 림프구 기원에 따라 임상 경과가 다르다는 것 등을 알게 되어 새로운 분류법이 필요하게 되었다.

이에 World Health Organization (WHO)에서는 림프구계 종양의 범주 안에 악성 림프종과 림프구 백혈병을 따로 구분하지 않고 세포 기원에 따라 크게 B세포 림프종양과 T/NK세포 림프종양으로 구분하였고, 각각을 골수 분화 단계인 precursor 세포 림프계 종양과 보다 분화된 세포인 mature 또는 peripheral 림프계 종양으로 구분하였다. 또한 lymphoblastic lymphoma, acute lymphocytic leukemia, chronic lymphocytic leukemia 그리고 small lymphocytic lymphoma (SLL)처럼 발현되는 임상 양상만 다를 뿐 기원 세포가 동일하고 치료가 동일한 질병은 하나의 질환으로 분류하였다. 그러므로 우리가 기존에 알고 있던 비호지킨 림프종은 mature B세포 또는 T/NK세포 종양으로 나뉘게 되었다. 2008년에 개정된 WHO 분류¹⁾에 의하면 mature B세포 림프종이 29아형, mature T/NK세포 림프종이 18아형으로 구분되어 있다. 더욱이 각각의 아형은 임상 경과가 다양하며 점차 아형에 따라 각기 다른 치료법이 개발되고 있어 비호지킨 림프종은 더 이상 하나의 질환명이 아니라 림프구 악성종양군을 통칭하는 의미로 받아들여야 하겠다.

미국인에서 악성 림프종의 발생 빈도는 매년 인구 10만 명당 B세포 림프종 26.13명, T세포 림프종 1.79명, 호지킨 림프종 2.67명이다. 전체 림프구 종양에서 호지킨 림프종은 8.2%이며, 비호지킨 림프종은 diffuse large B-cell lymphoma (DLBCL) 31%, follicular lymphoma (FL) 22%, MALT lymphoma 7.6%, peripheral T-cell lymphoma (PTCL) 7.6%, SLL 6.7%, mantle cell lymphoma (MCL) 6% 순으로 발생한다.²⁾ 우리나라는 국가암정보센터(www.cancer.go.kr)의 2003-2005년 자료에 의하면 비호지킨 림프종은 남자 10만명당 6.5명이 발생하며, 남자 암환자의 2.2%에 해당하는 10번째 호발암이다. 특히 14세 미만에서는 남녀 모두 세 번째, 15-34세 남자에서 네 번째로 호발하였다. 암 사망률 또한 남자 암환자에서 10번째였다. 우리나라에서

흔한 조직형은 2005-2006년에 31개 센터의 3995명을 대상으로 한 분석에 의하면 DLBCL 41.27%, MALT lymphoma 18.38%, extranodal NK/T cell lymphoma 6.08%, PTCL 5.79%, plasma cell myeloma 3.53%, FL 2.9%, anaplastic large cell lymphoma 2.83%, MCL 2.41% 순으로 서구와는 다른 발생 양상을 보인다.

이와 같은 비호지킨 림프종의 다양한 아형의 특성은, 치료반응평가지 ¹⁸F-FDG PET 결과에 반영되어 여러 연구자마다 다양한 결과를 보고하는 반면에, 호지킨 림프종은 비교적 일관성 있고 균일한 결과를 보여준다.

임상적으로 악성 림프종이 의심되면 수 많은 조직형을 감별하기 위하여 반드시 림프절 절제 조직검사를 시행해야 한다. Lactate dehydrogenase (LDH)를 포함한 기본적인 혈액검사와 흉부와 복부골반 CT스캔과 골수조직검사는 모든 악성 림프종 환자에서 시행하는 필수 검사 항목이다. ¹⁸F-FDG PET/CT 스캔은 National Comprehensive Cancer Network (NCCN, www.nccn.org) 가이드라인에 의하면 DLBCL 와 호지킨 림프종에서만 필수 검사이며 나머지 조직형에서는 선택적으로 검사하도록 되어있다. MUGA 스캔이나 심초음파검사는 anthracycline 계열 약제를 치료로 사용할 때 검사하게 된다. 병기 결정은 호지킨, 비호지킨 림프종 모두 Ann Arbor 분류의 Cotswold 기준(1989)을 사용하고 있다.³⁾ 그러나 위장관 MALT lymphoma는 TNM 분류와 유사한 병기 분류법을 사용하기도 한다.

비호지킨 림프종의 예후인자는 잘 알려진 international prognostic index (IPI), 병리조직학적 아형, B 증상과 bulky disease의 유무 등이 있고, 호지킨 림프종의 예후인자는 international prognostic score (IPS)의 7가지 인자가 있다.^{4,5)} 이들 예후인자를 바탕으로 치료방법을 선택하게 된다. 그러나 최근 치료 초기 또는 도중에 시행한 ¹⁸F-FDG PET 검사 결과가 위의 IPI 또는 IPS보다 더 강력하고 독립적인 예측인자라는 연구결과들이 인식되고 있다.^{6,7)}

현재 B세포 림프종들은 rituximab을 추가로 사용하면서 치료 성적이 향상되고 있다. DLBCL은 rituximab+CHOP (cyclophosphamide, doxorubicine, vincristine, prednisone)이 표준 치료제이며, FL은 CHOP 또는 rituximab+CVP이, 그리고, MCL 또한 rituximab+combination chemotherapy가 그 표준 치료제이다. 그러나 T세포 림프종은 CD52 단클론 항체인 alemtuzumab이 도입되었으나 독성 등으로 인해 큰 성과는 거두지 못하고 있다.

치료에 대한 종양의 반응평가는 왜 필요한가?

항암치료제 신약 개발을 위한 임상 시험을 수행할 때, 임상 1상에서는 적은 수의 환자를 대상으로 신약의 독성 평가가 주요 관심사이고 이를 통해, 이어지는 임상 2상에서 사용할 치료 용량을 결정하게 된다. 임상 1상에서는 환자수가 적어서 신약의 효능을 평가하는 통계적 검증력이 부족하므로 임상 2상에서 보다 많은 환자를 대상으로 약의 효능을 평가하게 된다. 신약의 효능이 확인되거나 가능성이 시사되면 보다 많은 환자를 대상으로, 전형적으로 다기관, 무작위 임상 3상 시험을 하게 된다. 신약 효능의 결정적인 증거는 환자의 생존을 향상이다. 생존율을 평가하기 위해서는 progression free survival (PFS) 또는 overall survival (OS) 같은 지표가 필요한데, 오랜 경과 관찰 기간이 소요되고 막대한 경제적 비용이 소요된다. 통상적인 신약 개발 시 임상 1, 2, 3 상 시험을 거쳐 미국 식품 의약국 (FDA)의 승인을 얻기까지 7-8년의 세월과 약 천 만불 이상의 금전적 경비가 소요된다고 한다. 신약 개발에 소요되는 비용이 너무나 막대하므로 최근에 시판되는 새로운 항암 치료제는 한달 치료 비용이 약 2천 만원에 달한다. 따라서 임상 2상 시험에서 PFS 또는 OS 와 같은 임상적 이득을 대신할 수 있는 또는 예측할 수 있는 평가 지표가 절실히 필요하다. 이상적인 지표(surrogate)는 가능한 치료 초기에, 객관적으로 그리고 정확하게 신약의 효능을 반영할 수 있고, 임상 시험을 계속 진행할 수 있는 약인지 그렇지 않은지를 구별할 수 있어야 한다. 이러한 지표들 중에서 치료에 대한 '종양의 반응'이 신약 효능의 객관적 지표로 간주되고 있다.

종양의 반응이란 간단히 말해서 동일한 임상적, 병리 조직학적 특성임에도 불구하고 어떤 환자는 치료 효과가 있고, 어떤 환자는 그렇지 않은 현상을 말한다. 종양들 사이에 현재 우리가 이해하지 못 하는, 임상적으로 의의 있는 생물학적 차이가 있고, 치료를 한 후에야 이를 알 수 있다. 특정 항암제가 특정 환자에게 효과가 있을 것인지 정확하게 예측할 수 있는 지표가 현재는 없으므로 치료 전 평가된 예후 인자에 따라 치료제를 선택하게 된다.

1960-70년대에 신약 개발을 위한 임상 시험을 수행하면서 치료제의 효능을 증명할 수 있는, 종양의 치료 반응 평가 방법의 필요성을 절감하게 되었고 임상 시험 관리 승인 기관의 승인을 받기 위해서, 또한 임상 시험 결과들을 서로 비교하기 위해서도 치료 반응 평가의 표준화가 요구 되었다.

항암 치료제는 일반적으로 일부 환자에게만 효과가 있다. 폐암, 유방암, 소화기암과 같은 대부분의 고형암들의 항암 치료제에 대한 치료 반응율이 20-40%에 불과하고 특정 항

암제가 특정 환자에게 효과가 있을 것인지 미리 예측할 수 없으므로 모든 환자에게 치료를 시작해야 한다. 따라서 치료 효과 여부가 판정될 때까지 60-80%에 이르는 대부분의 환자가 독성이 강한 항암 치료제를 투약 받게 되고 그로 인한 합병증의 위험이 커지고, 고가의 항암제를 별 이득 없이 소모하게 된다. 따라서 종양의 치료반응을 치료 초기에 또는 치료 도중에 평가하여 이런 환자들을 가려낼 수 있는 지표가 필요하다.

림프종 특히 호지킨 림프종의 경우 치료 반응율이 높아서 약 80-90%가 1차 관해에 도달한다. 그러나 화학 요법제의 독성으로 인한 합병증으로 장기간 생존율이 위협받게 된다. 그래서 치료에 잘 반응 하는 환자에 대해서는 용량을 줄이거나 독성이 보다 적은 치료제로 교체 치료함으로써 합병증으로 인한 위험을 줄일 필요가 있다.

반면 고형암에서와 마찬가지로, 치료에 반응하지 않는 환자에 대해서는 치료제를 강화 하거나 다른 대체 치료제로 바꾸어 치료함으로써 효과 없는 치료를 계속함으로써 생기는 합병증의 위험을 줄이고 치료 성공률을 높일 수 있다. 이와 같은 목표를 달성하기 위해서는 치료반응을 가능한 초기에, 객관적으로 그리고 정확하게 평가할 수 있는 방법이 필요하다.

지금까지 치료 반응 평가는 어떤 방법으로 해 왔나?

이와 같이 치료 반응 평가의 표준화 필요성이 대두됨에 따라 WHO의 주관으로 종양학자, 혈액학자, 방사선학자 등 전문가 그룹이 모여서 신약 개발을 위한 임상 시험 시 적용할 수 있는 고형암의 치료 반응 평가 기준을 개발하여 1979년 *Handbook of Criteria for Solid Tumor Response*를 출판하였고 1981년 Miller 등에 의해 발표되었다.⁸⁾

WHO의 반응 평가 기준은 측정 가능한 종양과 측정 불가능한 병변으로 나누어 각각 평가 하였다(Table 1). 측정 가능한 종양을 대상으로 가장 긴 직경과, 이것과 수직으로 교차하는 가장 긴 직경을 곱한 값(the product of the longest perpendicular diameter, bidimensional)을 측정하고 이 값들을 모두 합한다(the sum of the product of perpendicular diameter, SPD). 측정된 값들의 치료 전, 후 변화를 4 카테고리, 즉 complete response (CR), partial response (PR), no changes (NC)/stable disease (SD), progressive disease (PD)로 구분하였다.

CR: 모든 병변이 사라질 경우 (최소 4주 동안), PR:

SPD가 치료 전에 비해 50% 이상 감소될 경우 (최소 4주 동안), PD: 하나 이상의 병변의 SPD가 25% 이상 증가할 경우 또는 새로운 병변이 출현한 경우, NC/SD: PR과 PD 기준에 미치지 못할 경우.

측정 불가능한 병변에 대해서도 같은 방법으로 분류하였다. 측정 불가능한 병변은 복부 종양, 림프관성 폐전이, 흉막삼출, 복수, 골수병변 등을 포함한다. WHO 반응 평가 기준을 개발할 당시에는, 1972년 처음으로 두부 CT가 출시되고 몇 년 후 body CT가 출시되었으나 아직 널리 보급되기 전이었으므로 X선 단순촬영을 이용하여 종양의 크기를 측정하거나 촉진 가능한 림프절들을 캘리퍼스로서 측정하던 시기였으므로 복강 내 종양이 측정 불가능한 병변으로 간주된 것은 놀라운 사실이 아니다. 이 후 20년 동안 WHO 치료 반응 평가 기준은 많은 연구자들에 의해 임상 시험 시 널리 이용되어 왔고 임상 시험을 관리 승인하는 기관에 의해 종양의 반응을 평가하는 표준으로 인정되어 왔다. 그 동안 이 기준을 사용해 오면서 WHO 치료 반응 평가 기준에 명확히 언급 명시되지 않은 부분이 있었고 다소 애매하게 표현된 부분들이 잘못 해석되어 연구자마다 다양한 기준을 사용하여 연구 결과를 서로 비교하기가 어렵게 되는 문제점이 점차 부각되기 시작하였다. 1994년 The European Organization for Research and Treatment of Cancer (EORTC), The National Cancer Institute of USA (NCI), The National Cancer Institute of Canada Clinical Trial Group 이 협동하여 기존에 사용되던 치료 반응 평가 기준들을 재검토한 후 Response Evaluation Criteria in Solid Tumor (RECIST) group의 평가를 거쳐 2000년에 공식적으로 RECIST 기준이 발표되었다.⁹⁾

그 동안 CT, MRI 등 영상 진단 기술이 발전함에 따라 이를 반영하여, RECIST 기준은 WHO의 기본 골격을 그대로 유지한 채 세세한 부분들을 좀더 구체적이고 명확히 규정하였다. 가장 큰 차이점은 측정 가능한 병변에서 가장 긴 직경 하나만을 측정한다는 점이다(single longest diameter, unidimensional). 측정 가능한 병변의 크기는 전통적인 영상법에서는 20 mm 이상, 나선형 CT에서는 10 mm 이상으로 규정하였다. 측정 가능한 모든 장기의 병변 중 가장 직경이 크고, 가장 정확하게 반복 측정할 수 있는, 대표적 병변을 표적 병변으로 규정하고 표적 병변은 한 장기당 5개 이하, 총 10개까지 측정하도록 하였다. 모든 병변(최대 10개)에 대해서 측정된 장경의 합을 구하고, 치료 후 측정된 값과 비교하여 그 변화 정도에 따라서 4 가지 카테고리 분류하였다.

CR: 모든 병변이 사라질 경우(최소 4주 이상), PR: 최소 30% 이상 감소할 경우(4주), PD: 최소 20% 이상 증가할 경우(치료 및 경과 관찰 과정에서 2회 이상 검사한 경우) 또는 새로운 병변이 출현할 경우, SD: PR과 PD 기준에 도달하지 못한 경우

위에서 기술한 측정 가능한 병변의 크기에 미달하는 작은 병변들을 포함하여 다른 모든 병변들을 비표적 병변으로 규정하고 이 병변들의 유무와 개수는 기록하나 측정할 필요는 없다고 하였다. 비표적 병변에 대해서도 CR, PD, non-PD의 3 카테고리로 분류하였다.

위와 같이 판정된 결과들을 객관적 반응이라 지칭하였고 치료를 시작한 후부터 병이 진행하거나 재발이 발견될 때까지 관찰된, 가장 좋은 객관적 반응을 종합최선반응이라고 하였다.

WHO 및 RECIST 기준은 치료 반응 평가를 위해 종양의 크기를 측정할 때, CT, MRI, 초음파와 같은 형태학적 영상 검사에 거의 전적으로 의존하고 있다. ¹⁸F-FDG PET은 아직 널리 보급되지 않았고, 검증되지 않았다는 이유로 치료 반응 평가의 검사로 채택되지 않았다.

거의 비슷한 시기인 1999년, EORTC는 순전히 ¹⁸F-FDG PET의 결과를 바탕으로 한 치료 반응 평가 기준을 최초로 제시하였다.¹⁰⁾ 기존의 WHO와 RECIST 기준과 마찬가지로 치료 반응을 4가지 카테고리 분류하였다. 이 기준에서는 1주기 치료 후 최소 15-25%, 2주기 치료 후에는 25% 이상의 ¹⁸F-FDG 섭취 감소가 있을 경우 partial metabolic response (PMR)로 규정하고, 25% 이상 ¹⁸F-FDG 섭취가 증가하였을 경우 progressive metabolic disease (PMD)로 규정하였다. 이 기준은 적용대상 종양이 림프종인지 고형암인지 구체적으로 명시하지는 않았다.

RECIST 기준을 새로 다듬은 새로운 RECIST 기준 버전 1.1 이 2009년 발표되었다(Table 1).¹¹⁾ 버전 1.1 명칭이 말해주듯이 기존의 RECIST 기준의 골격은 그대로 유지되고 다른 점은 다음과 같다. 시뮬레이션 결과 측정할 표적 병변의 개수가 수개만 되어도 치료 반응 분류가 거의 일치한다는 점이고 림프절의 측정 시 횡단면 영상에서 장경 보다는 단경을 측정하도록 권고한 점 등이다. 또한 새로운 ¹⁸F-FDG PET 양성 병변이 출현할 경우 PD로 판정하는 새로운 기준이 채택되었다.

2009년 ¹⁸F-FDG PET을 이용한 고형암의 치료반응평가 기준이 Wahl 등에 의해 제안되었다(PERCIST 1.0).¹²⁾ 이 기준에서는 ¹⁸F-FDG 섭취가 30% 이상 감소하였을 경우 PMR로 규정하고, 30% 이상 증가하였을 경우 PMD로 규

Table 1. Comparison of WHO and RECIST 1.1 Criteria for Target Lesions Only

Characteristic	WHO	RECIST 1.1
Measurability of lesion at baseline	1. Measurable, bidimensional (product of LD and greatest perpendicular diameter) 2. Nonmeasurable/evaluable (e.g., lymphangitic pulmonary metastases, abdominal masses)	1. Measurable, can be accurately measured at least one dimension (longest diameter) as 20 mm either conventional technique or 10 mm with spiral CT; nodes: target short axis \geq 15 mm, nontarget 10- to 15-mm nodes, normal < 10 mm 2. Nonmeasurable, all other lesions including small lesions
Objective responses	1. Measurable disease (change in sum of products of the LD and greatest perpendicular diameters, no maximal number of lesions specified); CR; disappearance of all known disease, confirmed at \geq 4 wk; PR; \geq 50% decrease from baseline, confirmed at \geq 4 wk; PD; \geq 25% increase of one or more lesions or appearance of new lesions; NC; neither PR nor PD criteria met	1. Target lesions (change in sum of LDs, maximum of 2 per organ up to 5 total): CR; disappearance of all target lesions (for at least \geq 4 wk) PR; at least 30% decrease in sum longest diameter, taking as reference baseline sum longest diameter (confirmed at 4 wk) PD; at least 20% increase in sum longest diameter of target lesions, taking as reference the smallest sum longest diameter or the appearance of one or more new lesions SD; Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD
Overall best response		

정하였다.

한편, 림프종의 경우 1980년대까지 Ann Arbor 병기 분류 기준에 따라 치료 후 결과를 평가 하였고 주로 이학적 검사, 임상 소견, 검사실 소견, X선 소견에 의존 하였다. CT, MRI 등 영상 진단법이 널리 보급됨에 따라 Ann Arbor 병기 기준에 바탕을 둔, 호지킨 림프종의 치료 반응 평가 기준이 1989년 새로 제안 되었다(Cotswold 기준).³⁾ 이 기은 병기 설정뿐만 아니라 치료 반응 평가에 CT를 이용하도록 공식적으로 채택하였다. 치료 반응 평가는 CR, CRu (CR/unconfirmed), PR, PD로 분류하였고 그 평가 판정 기준은 WHO 기준을 그대로 원용하였다. 예를 들어 PR은 측정 가능한 모든 병변에서 측정된 SPD가 최소 50% 이상 감소되고 증상이 사라진 경우로 하였고, PD는 측정 가능한 1 개 이상의 병변에서 25% 이상 증가하거나, 새로운 병변이 출현하거나, 증상이 재발한 경우로 하였다. 갈륨 스캔이나 폴 스캔이 임상적으로 의심되는 병변이나 병의 범위를 확인하기 위해 사용될 수는 있으나 일상적인 검사로 요구되지는 않았다.

표준화된 치료 반응 평가 기준이 임상 시험을 수행하기 위해서, 연구 결과를 상호 비교하기 위해서, 가능성이 있는 신약을 확인하기 위해서 필요함에도 불구하고 비호지킨 림프종에 적용할 수 있는, 국제적으로 공인되고 널리 받아들여진 치료 반응 평가 기준이 없었다. 이에 비호지킨 림프종의 치료 반응 평가 기준 권고안인 International Working

Group response criteria (IWC)가 1999년에 발표되었다 (Table 2).¹³⁾ 이 기준 역시 WHO 기준을 부분적으로 원용하여 CT와 같은 형태학적 영상 진단법에 의해 종양의 크기 변화를 측정하였다. 다만 재발 또는 PD의 경우 SPD가 최소 50% 이상 증가하였을 때 또는 새로운 병변이 출현했을 때 등으로 규정하여 WHO와 Cotswold 기준의 25% 보다 더 강화 되었다. 이는 고형암과는 다른, 독특한 비호지킨 림프종의 특성을 반영하여 고형암의 반응 기준을 다소 변경하여 사용한 것으로 보인다. 치료 반응은 CR, CRu, PR, SD, relapsed disease, PD로 분류하였다. 이 기준에서도 갈륨 SPECT에 대해서 언급하고 있으나 널리 보급되고 재현성이 검증 될 때까지는 치료 반응을 평가하는 임상 시험에서 일상적인 검사로 채택하지 않는다고 하였다.

이 기준이 1999년 발표되기 전까지는 정상 림프절의 크기, 치료 반응 평가를 어느 시점에 해야 하는지, 또 얼마나 자주 해야 하는지, 그 평가 기준이 무엇인지에 대해 다양한 기준들이 연구 그룹마다, 기관마다 사용되고 있었다. 이 기준은 발표된 후 임상 의사와 임상 시험 관리 기관에 의해 신속하게 그리고 광범위하게 받아들여졌고, 많은 신약 개발을 위한 임상 시험을 승인 받기 위해 사용되었다. 그러나 8년 동안 이 기준이 이용되는 동안, 관찰자간 및 관찰자내 변이가 상당히 큰 경향이 있고, CRu 카테고리를 적용하는데 혼선이 있었고, 그 동안 ¹⁸F-FDG PET의 이용이 증가하였으며 면역 조직 화학 염색법, 유동 세포 분석법의 발전

Table 2. International Workshop Criteria for Response Assessment in Lymphoma

Response Category	Description
Complete response (CR)	Complete disappearance of all detectable disease on CT with previously involved nodes on CT >1.5 cm in their greatest axial diameter regressing to <1.5 cm, and nodes of 1-1.5 cm regressing to <1 cm. In addition, resolution of disease-related symptoms, normalization of biochemical abnormalities, and normal bone marrow biopsy.
Complete response unconfirmed (CRu)	Corresponds to CR criteria, but with a residual mass >1.5 cm in greatest axial diameter that has regressed by >75% in the SPD.
Partial response (PR)	At least 50% reduction in the sum of the product of the greatest diameters (SPD) of the six largest nodes with no increase in the size of the other nodes and no new sites of disease. Hepatic and splenic nodules should also decrease by at least 50% in the SPD.
Stable disease (SD)	Response is less than a PR, but is not progressive disease.
Progressive disease (PD)	More than 50% increase in the sum of the product of the greatest diameters of any previously abnormal node, or appearance of any new lesions during or at the end of therapy.
Relapsed disease (RD)	The appearance of any new lesion or increase in size of >50% of previously involved sites or nodes in patients who achieved CR or CRu.

으로 개정의 필요성이 대두 되었다. 이에 International Harmonization Project (IHP)가 소집되어 5개 분과 위원회를 구성하고 비호지킨 림프종과 호지킨 림프종 모두에 적용할 수 있는 개정된 치료 반응 평가 기준(IWC+PET)과 권고안을 2007년 발표 하였다.^{14,15)}

치료반응평가 방법으로서의 형태학적 영상 진단법과 ¹⁸F-FDG PET의 장단점

앞서 소개한 종양의 치료반응평가 방법 및 기준들은 모두 종양의 형태학적 변화 즉, 치료 전 후 종양의 크기 변화를 측정하는 데에 바탕을 두고 있다. 지난 20-30년 동안 종양생물학, 분자생물학, 유전학 등이 발전하여 종양의 분자생물학적 특성, 종양과 인체와의 상호작용과정이 많이 밝혀지게 되었다. 종양의 반응은 여러가지 형태로 반영될 수 있는데 예를 들면, 혈청 종양 표지자, 분자 생물학적 표지자, 유전자 발현 감정법 등을 들 수 있다. 그럼에도 불구하고 30년 전에 제정된 형태학적 기준 즉, 오로지 종양의 '크기 변화'가 그 동안 큰 변화 없이 지금까지 사용되고 있다. 형태학적 변화 즉, 종양 크기의 변화는 치료반응을 평가하는 중요한 지표인 것은 틀림 없으나 ¹⁸F-FDG PET의 이용이 증가하면서 그 제한점이 상대적으로 부각되었다.

형태학적 기준의 제한점을 ¹⁸F-FDG PET과 비교하여 요약하면 다음과 같다.

1. 형태학적 기준의 한계점을 극명하게 보여준 대표적 예 중의 하나가 림프종 치료 후 많은 환자에서 관찰되는, 잔존 종괴의 평가문제이다. 임상적으로 완전 관해에 도달한 환자에서 잔존 종괴가 CT에서 관찰되고 이 종괴는 수 주에서

수 개월, 수 년까지 지속적으로 관찰되었다. 이 종괴가 생존 림프종인지, 치료와 관련된 섬유화 및 괴사 조직인지 CT로는 판정하기 어려웠다.¹⁶⁾ 특히, 호지킨 림프종은 치료에 별 영향을 받지 않는 섬유화 조직, 반응성 염증 세포들로 이루어져 있고, 악성 림프종 세포는 극히 일부만 포함하고 있으므로 비록 효과적인 치료에 의해 림프종 세포가 모두 사멸되더라도 뚜렷한 종괴의 수축을 초래할 것으로 기대하기는 어렵다.¹⁷⁾ 그래서 1989년 Cotswold 기준에서 CRu의 카테고리가 새로 만들어 졌고, 1999년 IWC 기준까지 유지되었다. ¹⁸F-FDG PET은 이러한 잔존 종괴가 생존 림프종인지, 섬유화 또는 괴사조직인지를 보다 명확하게 구분할 수 있어, 2007년 발표된 IWC+PET 기준에서는 CRu 카테고리가 사라지게 되었다.

2. 최근에 림프종뿐만 아니라 고형암에서도 사용하는 표적 치료제는 세포감소적(cytoreductive)이라기 보다는 주로 세포증식억제적(cytostatic)인 효과를 보여준다. 치료가 성공적으로 수행되었다 하더라도 종양의 크기에는 큰 변화를 초래하지 않고, 따라서 형태학적 영상 기준으로는 치료에 변화가 없는 SD로 판정되기 마련이다.^{3,18)} 종양의 크기가 수축되지 않았음에도 불구하고 치료에 성공한 환자들의 생존률이 향상되었기 때문에 이런 임상 상황에서는 형태학적 기준이 종양의 치료 반응을 반영한다고 말하기 어렵다. 최근에는 위와 같은 치료제들이 전통적인 항암제와 병합하여 사용하는 추세이므로 이 문제가 중요한 이슈가 되고 있다. ¹⁸F-FDG PET은 종괴의 크기 변화에 의존하지 않고 종양 세포의 포도당 대사의 변화 또는 생존 암세포의 수를 반영하므로 이와 같은 상황에서는 ¹⁸F-FDG PET이 치료반응 평가의 도구로서 더 적합할 수 있다.

3. 치료 후 종양의 수축은 즉각 일어나지 않고 시간이 걸리므로, 치료 초기 또는 치료 도중에 치료 반응을 평가하여 그에 따라 치료제 또는 치료 방법을 조절, 변경하는 방법으로 형태학적 기준을 적용하기는 매우 어렵다.

4. CT 영상에서 종양의 크기 변화를 측정하여 치료 반응을 분류할 때 판독자간 및 판독자내 변이가 크다. 폐암 환자를 대상으로 RECIST와 WHO 기준을 각각 적용했을 때, SD를 PD로 잘못 분류하는 판독자내 비율이 각각 9.5%, 20.5%이었고, 판독자간 비율은 각각 29.8%, 42.5%였다.¹⁹⁾

림프종 치료반응평가에서 ¹⁸F-FDG PET의 이용

림프종에서 ¹⁸F-FDG PET은 치료 도중과 치료완료 시점에서 치료효과 판정을 위해 이용될 수 있고, 추적기간 동안 재발의 발견에 이용될 수 있으며, 이러한 역할들에 대한 많은 연구들이 현재 진행되고 있다.²⁰⁾

1. 치료 도중 림프종의 치료반응평가

림프종은 많은 치료 전 예후 판정 인자가 확립되어 있어서, 이러한 예후 판정 인자와 병기를 이용하여 최적의 초기 치료 전략을 계획할 수 있다. 하지만, 이 초기 치료 전략이 항상 성공적이지 않으므로, 치료 도중 현재 치료에 반응이 있는 환자와 없는 환자를 구분할 수 있다면 반응이 없는 환자에서 무익한 치료를 중단하고 새로운 치료를 시작하여 치료 관련 이환률과 사망률은 높이지 않으면서 완치율을 높일 수 있다. 기존의 영상의학적인 방법은 종양의 크기가 줄어드는 형태학적 변화를 관찰하는 방법이므로, 치료 후 괴사나 섬유화와 잔존암을 정확히 구분하지 못하고, 치료 후 종양이 줄어들지 않는다는 상대적으로 긴 시간이 소요되므로 치료 도중 반응 평가에는 제한적으로 사용되었다. 반면에 기능적 영상인 ¹⁸F-FDG PET은 초기 대사 변화를 관찰할 수 있어 치료 도중 반응 평가가 가능하다.

비호지킨 림프종과 호지킨 림프종을 대상으로 한 여러 연구에서 1-3주기의 항암치료 후 ¹⁸F-FDG PET상에서 초기 대사 변화는 최후 치료효과 및 PFS 등과 밀접한 연관성을 보였다. Mikhael 등의 연구에서, 치료 도중 ¹⁸F-FDG PET 양성인 환자들은 1년 PFS율이 10-50%인 반면, PET 음성 환자들은 79-100%의 PFS율을 보였고, ¹⁸F-FDG PET 양성인 환자들의 재발률이 병기에 상관없이 높았다.²¹⁾ Hutchings 등의 후향적 연구에서는 5년 PFS율이 치료 도중 ¹⁸F-FDG PET 양성인 환자들의 경우 39%인 반면, ¹⁸F-FDG PET 음성인 환자는 92%로 유의하게 높았다.²²⁾ 이러한 결과는 Hutchings 등과 Gallamini 등의 전향적 연구

를 통해 다시 입증되었다.^{23,24)}

그 외에도 치료 도중 림프종의 치료효과 판정에 있어서 ¹⁸F-FDG PET의 역할에 대한 많은 연구가 있었고 ¹⁸F-FDG PET의 유용성을 입증하였지만, 연구마다 대상 환자군, 연구방법, ¹⁸F-FDG PET 스캔 전까지 시행한 항암 요법의 주기 수, ¹⁸F-FDG PET 스캔 분석 방법, 치료반응 판정기준 및 추적기간 등이 다양하여 하나의 일률적인 판정 기준을 제시하기는 어렵다.

2. 치료완료 시점에서 치료효과 판정

치료완료 후 치료효과 판정은 PFS와 OS와 같은 치료에 의한 임상적 이득을 측정할 수 있는 중요한 요소이며, 치료 계속 여부나 치료 방법을 결정하는데 중요한 기준이 된다. 기존의 영상의학적인 방법이 잔여암과 괴사 또는 섬유화를 구분하는데 좋지 않은 성적을 보인 반면, ¹⁸F-FDG PET은 CT상 관찰되는 잔존암의 유무와 상관없이 PFS와 OS를 잘 예측할 수 있음이 여러 연구를 통해 입증되었다.²⁵⁻²⁷⁾

하지만, 치료완료 시점에서 ¹⁸F-FDG PET 음성인 경우에도 현미경적인 질병이 존재함을 배제할 수는 없다. 따라서, ¹⁸F-FDG PET의 결과에 따라 계획된 항암치료나 방사선치료를 생략하는 것은 시험적으로 시도되어야 한다. 또한 ¹⁸F-FDG 섭취가 사라졌다는 소견이 계획된 추가 치료를 생략하는 근거가 될 수 없다는 것을 의미한다. 이러한 사실은 Picardi M 등의 연구에 의해 확인할 수 있다.²⁸⁾ Picardi M 등은 항암치료 완료 후 종괴의 ¹⁸F-FDG 섭취가 사라진 160명의 호지킨 림프종 환자들을 추가 방사선치료를 받는 군과 받지 않는 군으로 무작위적으로 나누어 추적 검사하였다. 그 결과, 추가 방사선치료를 받지 않은 군은 18개월 동안 14%에서 재발이 발생한 반면, 추가 방사선치료를 받은 군은 2.5%에서만 재발이 발생하였다. 이 연구는 ¹⁸F-FDG PET 음성 환자 일부에서 현미경적인 질병이 잔존하므로 처음 계획되었던 방사선치료를 생략하면 재발률이 높아진다는 사실을 확인시켜주었다.

3. 추적기간 동안 재발의 발견

추적기간 동안 재발의 조기 발견은 명백한 재발 이전의 미세질병을 치료를 할 수 있게 하므로, 치료 효과를 높이고 환자의 예후를 개선시킬 수 있는 중요한 요소이다.

추적기간 동안 재발을 발견하는데 있어 ¹⁸F-FDG PET의 역할에 대한 연구는 상대적으로 많지 않다. 21명의 호지킨 림프종 환자를 대상으로 한 Dittmann 등의 연구에서 ¹⁸F-FDG PET은 CT에 비해 추가적인 정보를 제공하지 못하였다.²⁹⁾ 반면, Jerusalem 등은 치료가 완료된 36명의 호지

킨 림프종 환자를 대상으로 2-3년간 ¹⁸F-FDG PET 추적검사를 시행하였고, 이 기간 동안 11명의 환자에서 ¹⁸F-FDG PET 양성 소견이 관찰되었다. 이 환자들 중 실제로 재발이 발생한 5명의 환자는 임상 증상, 검사실 결과, CT 등에서 재발 소견을 보이기 전에 ¹⁸F-FDG PET을 이용해 재발을 발견할 수 있었고, ¹⁸F-FDG PET에서 위양성을 보인 나머지 6명의 환자는 4-6주 후 반복 시행한 ¹⁸F-FDG PET에서 재발이 없음을 확인할 수 있었다.³⁰⁾ Zinzani 등은 151명의 종격동 림프종 환자를 대상으로 후향적 연구를 시행하였고, 30명의 환자에서 재발을 의심하는 ¹⁸F-FDG PET 양성 소견이 관찰되었다. 30명 중 실제 재발 환자는 17명이었으며, 나머지 13명은 다른 양성 질병이 발견되었다. 재발된 17명 환자 중 3명만이 다른 검사 방법에서 재발을 의심하는 소견이 관찰되었다.³¹⁾

¹⁸F-FDG PET의 림프종 치료반응평가 기준

앞에서 기술한 1999년 EORTC 기준은 악성뇌종양, 수모세포종, 두경부종양, 유방암, 결장직장암 그리고 림프종 등에 관한 연구 결과를 종합하여 제시되었다.¹⁰⁾ 따라서 이 기준은 악성 림프종만을 대상으로 제시된 기준이 아니고, 기준 결정에 인용된 연구 결과들이 한정되어 있어 대규모 추가 연구가 필요하다는 단점을 가지고 있었다. 하지만 ¹⁸F-FDG 섭취의 변화 정도(standardized uptake value, SUV)를 치료효과 판정의 기준으로 삼은 첫 번째 기준이고, 치료효과 판정을 위한 큰 틀을 제시하여 이후의 연구와 임상에서 치료효과를 비교할 수 있는 기준을 제시했다는 의의가 있었다.

2007년 IHP는 림프종에서 ¹⁸F-FDG PET의 이용이 급격하게 늘어남에 따라 누적된 연구 자료를 분석하여 림프종에서 ¹⁸F-FDG PET의 이용에 대한 합의 내용을 발표하였다.¹⁵⁾ 합의내용의 주요 골자를 요약하면 다음과 같다.

1. 림프종 치료완료 시점에서 치료반응평가를 위한 ¹⁸F-FDG PET의 사용

림프종 치료완료 시점에서 치료효과 판정을 분석한 여러 연구들의 종합 분석한 결과, ¹⁸F-FDG PET은 잔여암을 발견하는데 호지킨 림프종의 경우 민감도와 특이도가 각각 84% (95% CI, 71% to 92%)와 90% (95% CI, 84% to 94%)이었고, aggressive 비호지킨 림프종의 경우 각각 72% (95% CI, 61% to 82%)와 100% (95% CI, 97% to 100%)이었다. 이러한 성적은 기존의 방사선학적 영상이 잔여암과 섬유화, 괴사 등을 구분하지 못하는 한계를 극복할

수 있는 ¹⁸F-FDG PET의 우수성을 반영하고 있다.

2. 치료 전 ¹⁸F-FDG PET의 필요성

일반적으로 ¹⁸F-FDG 섭취를 잘하는 조직형의 림프종에서 치료 전 ¹⁸F-FDG PET은 의무적인 것은 아니지만, 치료 후 판독을 용이하게 하기 위해서 적극적으로 추천된다. 반면, ¹⁸F-FDG 섭취가 다양한 것으로 알려진 조직형의 림프종에서 치료 전 ¹⁸F-FDG PET은 반드시 필요하다.

3. 치료반응평가를 위한 ¹⁸F-FDG PET의 검사 시점

항암치료 또는 방사선치료 후 최소 2-3개월 동안 ¹⁸F-FDG PET의 판독을 저해할 수 있는 염증 반응이 관찰되므로, 이를 최소화하기 위해 항암치료 완료 후 3주 이후에, 방사선치료 후에는 8-12주 이후에 시행하는 것이 추천된다.

4. 치료반응평가를 위한 ¹⁸F-FDG PET의 판독

¹⁸F-FDG PET상에서 병변의 유무는 영상 평가로 충분하고, 따라서 정량적인 방법이 필수적인 것은 아니다. 일반적으로 잔여암은 정상적인 또는 생리적인 섭취 이외의 부위에서 관찰되는 국소적이거나 전반적인 ¹⁸F-FDG 섭취로 정의할 수 있다.

5. 치료 도중 치료반응평가를 위한 ¹⁸F-FDG PET의 이용

치료 도중 시행한 ¹⁸F-FDG PET은 치료 방법을 바꿀 수 있는 충분한 근거가 되고, 치료 초기의 ¹⁸F-FDG PET 소견 정상화와 환자의 예후는 밀접한 연관이 있다. 치료 도중 시행한 ¹⁸F-FDG PET을 판독할 때는, 더 일관되고 정확한 판정을 위해 SUV와 같은 정량적인 방법을 이용해 판독하는 것이 추천된다. 또한 치료 도중 ¹⁸F-FDG PET 검사는 다음 번 항암주기에 가장 가까운 날짜에 시행하는 것이 추천된다.

IHP는 위와 같은 합의 내용뿐만 아니라, 1999년에 발표된 IWC 기준을 통합하여 새로운 기준인 IWC+PET을 발표했다.¹⁴⁾ 이 기준은 기존의 IWC 기준이 사용하였던 종괴의 크기변화를 여전히 판정기준으로 사용하고 있지만 ¹⁸F-FDG 섭취 유무 역시 치료반응 평가의 중요한 기준으로 사용하여, 림프종의 치료반응 평가에 ¹⁸F-FDG PET이 필수적이라는 사실을 반영하고 있다(Table 3). 즉, 림프종에서 항상 문제가 되어왔던 30-50%의 환자에서 관찰되는 잔존 종괴를 ¹⁸F-FDG PET 결과를 근거로 림프종이 여전히 존재하는지의 여부를 확인할 수 있게 한 점이다. 따라서

Table 3. Revised Response Criteria for Malignant Lymphoma (IWC+PET)

Response	Definition	Nodal Masses	Spleen, Liver	Bone Marrow
Complete remission (CR)	Disappearance of all evidence of disease	(a) FDG-avid or PET positive prior to therapy; mass of any size permitted if PET negative (b) Variably FDG-avid or PET negative; regression to normal size on CT	Not palpable, nodules disappeared	Infiltrate cleared on repeat biopsy; if indeterminate by morphology, immunohistochemistry should be negative
Partial remission (PR)	Regression of measurable disease and no new sites	≥50% decrease in SPD ^{o)} of up to 6 largest dominant masses; no increase in size of other nodes (a) FDG-avid or PET positive prior to therapy; one or more PET positive at previously involved site (b) Variably FDG-avid or PET negative; regression on CT	≥50% decrease in SPD of nodules (for single nodule in greatest transverse diameter); no increase in size of liver or spleen	Irrelevant if positive prior to therapy; cell type should be specified
Stable disease (SD)	Failure to attain CR/PR or PD	(a) FDG-avid or PET positive prior to therapy; PET positive at prior sites of disease and no new sites on CT or PET (b) Variably FDG-avid or PET negative; no change in size of previous lesions on CT		
Relapsed disease or progressive disease (PD)	Any new lesion or increase by ≥50% of previously involved sites from nadir	Appearance of a new lesion(s) >1.5 cm in any axis, ≥50% increase in SPD of more than one node, or ≥50% increase in longest diameter of a previously identified node > 1 cm in short axis Lesions PET positive if FDG-avid lymphoma or PET positive prior to therapy	>50% increase from nadir in the SPD of any previous lesions	New or recurrent involvement

^{o)}sum of the product of the diameters.

예전 IWC 기준의 CRu 카테고리는 사용하지 않게 되었다.

IWC+PET 기준은 여러 연구를 통해 그 유용성이 확인되었는데, L. Brepoels 등은 ¹⁸F-FDG 섭취가 높은 비호지킨 림프종과 호지킨 림프종에서 IWC+PET 기준은 항암치료에 의해 완치가 될 환자와 재발할 환자를 이전의 IWC의 기준에 비해 더 정확하게 구분할 수 있었다고 보고하였다.^{32,33)} 그러나 ¹⁸F-FDG 섭취가 낮은 비호지킨 림프종의 경우 새로운 기준이 도움이 되지 않았고 추가적인 연구가 필요한 것으로 결론 내렸다.

종양의 치료 반응 양상

항암치료나 방사선 치료를 매번 시행할 때마다 종양 세포가 일정비율로 사멸된다면 생존 종양 세포 수 또는 FDG 섭취 정도는 치료가 반복됨에 따라 연속적으로 지수함수적으로 감소할 것이다(Fig. 1).³⁴⁾ 종양의 종류나 아형, 특정 종양에 대한 특정 치료제의 효과, 환자와 종양과의 생물학적 상호작용 등에 의해 사멸되는 종양세포의 비율이 달라짐으로 다양한 양성의 반응 곡선을 가정할 수 있다.

각각의 반응 곡선은 특정 종양의 반응을 나타내거나, 같은 종양에 대해 다른 치료제의 반응을 각각 나타내거나, 같

은 종양, 같은 치료제에 대한 환자의 반응을 각각 나타내는 반응곡선이라고 할 수 있다. 따라서 치료 반응은 종양의 종류, 생물학적 특성, 치료제의 치료 효과, 치료 시작 후 반응을 평가한 시점 등에 따라 달라짐으로 현재로서는 일률적으로 어떤 평가 기준을 제시하기는 어렵다. 판독기준 또한 검사 목적 즉, ¹⁸F-FDG PET을 특정 임상상황에서 민감도를 향상시키는데 주안점을 둘 것인가 특이도를 향상시키는데 주안점을 둘 것인가에 따라 달라질 수 있다.

반응곡선에서 눈여겨 볼 점은, 몇 개의 종양이 6-8치료주기 동안에 ¹⁸F-FDG PET음성 결과에 도달하나 그 의미가 같지 않다는 점이다. 반응속도가 서로 다른 종양들이 ¹⁸F-FDG PET음성에 함께 포함되어있기 때문이다. 현재 사용되는 대부분의 ¹⁸F-FDG PET 장비의 해상력 한계는 약 4.5-10 mm이고 부피로는 0.1-1.0 g에 해당된다. 일반적으로 종양세포수가 10⁹ 개로 성장하면 그 부피가 1 cm³가 된다고 알려져 있다. 따라서 ¹⁸F-FDG PET이 검출할 수 있는 종양세포의 수는 10⁸-10⁹ 이다. 따라서 ¹⁸F-FDG PET 음성 결과는 최대 10⁷ 개의 생존 종양 세포가 존재하는 종양으로부터 종양세포가 완전히 사멸된 종양까지를 모두 포함하고 있음을 의미한다. 통상적인 치료가 완료된 후, 10⁷개에 가까운 종양세포가 남아있는 종양은 ¹⁸F-FDG PET 음성

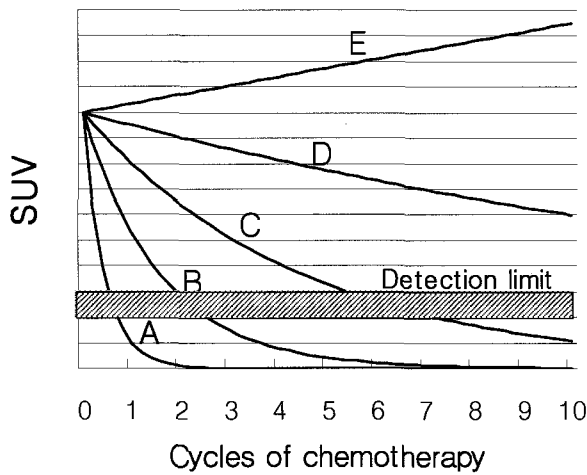


Figure 1. Hypothetical tumor response curve shows that tumor response varies from a rapidly responding (A), through a slowly or poorly responding (C), to an almost totally nonresponding tumor (E), depending on the tumor type, effectiveness of specific therapy regimens and host-tumor interactions. A rapidly responding tumor becomes ¹⁸F-FDG PET negative even after 1-2 cycles of treatment. In contrast, a slowly or poorly responding tumor becomes PET negative after 6-8 cycles of treatment, but may still contain significant amount of viable tumor cells which may start to regrow and eventually recur after completion of treatment. This tumor may be discriminated from a good responding tumor by performing ¹⁸F-FDG PET after 2-4 cycles of treatment.

결과에도 불구하고 다시 재성장하여 재발할 가능성이 높다. 만약 이런 현상이 치료 초기에 일어난다면 ¹⁸F-FDG PET 음성 결과가 최종 결과를 예측하지 못하는 경우가 생길 수 있다. 이런 현상은 고형암에서 흔히 관찰되는데, ¹⁸F-FDG PET 음성 결과가 최종 결과를 예측하는 확률이 50-70%로 낮다(음성 예측도). 이에 비해 림프종의 경우에는 1-2 치료 주기 후 ¹⁸F-FDG PET 음성 결과가 최종 결과와 일치하는 확률이 매우 높아 음성 예측도가 95% 수준이다. 이는 치료 초기에 평가된 치료 반응 속도가 치료 완료시까지 지속적으로 유지됨을 의미하고 아마도 림프종의 특성 또는 치료제의 탁월한 치료 효과에 기인한 것으로 보인다. 이와 같이 치료 반응은 단순히 ¹⁸F-FDG PET 음성에 도달했느냐의 여부보다는 치료 후 초기에 얼마나 신속하게 반응하느냐가 더 중요한 요소로 생각된다. Fig. 1에서, 6-8 치료 주기에 ¹⁸F-FDG PET 음성을 보인 종양이 2-4 치료 주기에서는 ¹⁸F-FDG PET 양성으로 판정될 수 있음을 보여주듯이 어느 시점에 평가하는가에 따라서 반응 평가 결과가 달라질 수 있다.

현재 사용중인 ¹⁸F-FDG PET 장비의 해상력과 방사능을 검출하는 민감도로는 현미경적 잔존 종양 세포의 존재를 배제할 수 없어, 위음성 결과가 나올 수 있으므로 이를 염

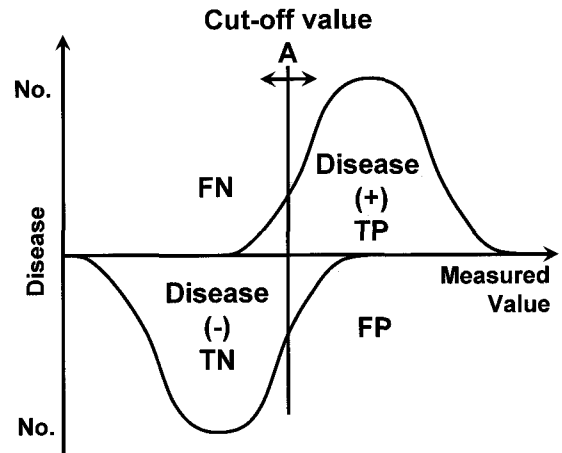


Figure 2. Graphical illustration of diagnostic performance of test shows relationship between 4 parameters (TP, true positive; FP, false positive; TN, true negative; FN, false negative) of diagnostic performance as cut-off value changes.

두해 두어야 한다. 또한 가장 흔한 위음성으로는 고혈당을 들 수 있다. 가장 흔한 위양성의 원인으로는 화학 요법 또는 방사선 치료 후 종양 부위의 염증 반응이고 생리적 섭취 부위, 정상 해부학적 부위의 섭취 등과 구별해야 한다.

위에서 논의한, 반응 평가에 영향을 미치는 여러가지 요인들과 영향의 경중을 고려하여 판단해야 하므로 판독자의 경험과 주관이 개입되는 것은 불가피하다. 위에서 논의한 주제들에 대해 많은 임상 시험들이 진행 중에 있으므로 그 결과들이 충분히 축적되면 보다 구체적이고 일관된 평가 기준이 제시될 것으로 기대된다. 치료 반응 평가를 위한 ¹⁸F-FDG PET 결과를 판독할 때 도움이 되는, ¹⁸F-FDG PET 진단 성능의 특성과 우도비에 대해 다음 항에서 기술하고자 한다.

¹⁸F-FDG PET 진단 성능의 특성

진단적 검사의 진단 성능 특성을 표현하는 변수들과 그 정의를 Table 4와 아래에 요약하였다.

$$\text{Sensitivity} = \frac{\text{true positive (TP)}}{[\text{TP} + \text{false negative (FN)}]}$$

$$\text{Specificity} = \frac{\text{true negative (TN)}}{[\text{TN} + \text{false positive (FP)}]}$$

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})}$$

$$\text{Positive predictive value} = \frac{\text{TP}}{(\text{TP} + \text{FP})}$$

$$\text{Negative predictive value} = \frac{\text{TN}}{(\text{TN} + \text{FN})}$$

$$\text{False positive rate} = \frac{\text{FP}}{(\text{FP} + \text{TN})} = 1 - \text{specificity}$$

$$\text{False negative rate} = \frac{\text{FN}}{(\text{FN} + \text{TP})} = 1 - \text{sensitivity}$$

이 변수들은 진단기준이 변함에 따라 서로 밀접하게 연

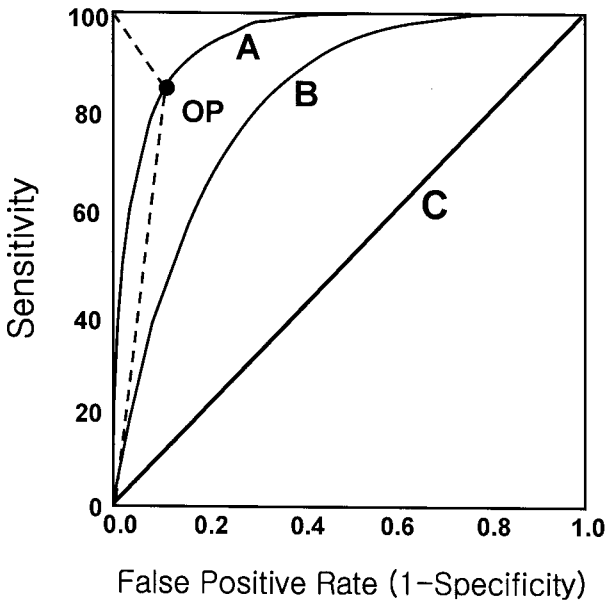


Figure 3. The examples of receiver operating characteristic curve were shown. The larger area under curve (AUC), the better diagnostic test is. Optimal point (OP) balancing sensitivity and specificity can be determined by a point on the curve to which distance from both point of sensitivity 100% and point of specificity 100% is the shortest one. Positive likelihood ratios at each point on the curve can be calculated using formula, sensitivity/(1-specificity).

Table 4. Characteristics of Diagnostic Performance of Test

	Disease (+)	Disease (-)
Test (+)	TP ^{a)}	FP ^{b)}
Test (-)	FN ^{c)}	TN ^{d)}

^{a)}true positive, ^{b)}false positive, ^{c)}false negative, ^{d)}true negative.

관되어 변화한다. 이 변수들이 변화하는 상호 관계를 이해하기 쉽게 그래프로 표현한 것이 Fig. 2에 제시되어 있다. 민감도를 증가시키기 위해서 양성으로 판정하는 진단 기준을 완화하면, 즉 그래프에서 진단기준 수직선 A를 왼쪽으로 옮기면 위양성률이 증가하게 되어 특이도는 감소하게 되고, 반면에 위음성이 줄어들기 때문에 음성 예측도는 증가하게 된다. 반대로 검사 결과를 양성으로 판정하는 기준을 강화하면 즉, 그래프에서 진단 기준 수직선 A를 오른쪽으로 옮기면, 양성 예측도와 특이도는 증가하나 반면에 위음성이 증가하게 되어 음성 예측도는 감소하게 된다. 그래프에서 보듯이 진단 기준이 변화에 따라 즉, 민감도가 증가하면 양성 예측도는 감소하고 필연적으로 특이도는 감소하나 음성 예측도는 증가하게 된다. Fig. 2에서 정상군과 비정상군의 분포 곡선 중 겹치는 면적이 적을수록, 측정값의 분

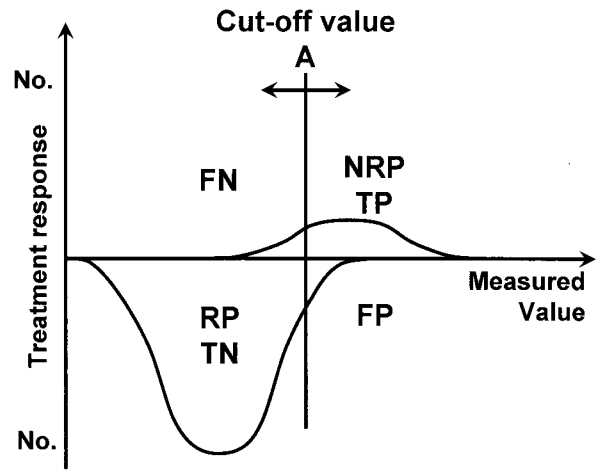


Figure 4. Graphical illustration of diagnostic performance of ¹⁸F-FDG PET for assessment of treatment response in lymphoma shows very high specificity and negative predictive value because of high pre-test probability (frequency of responders to therapy). Note the small overlapping area between nonresponders (NRP) and responders (RP) and wide range of measured value (wide width on x axis; TP, true positive; FP, false positive; TN, true negative; FN, false negative).

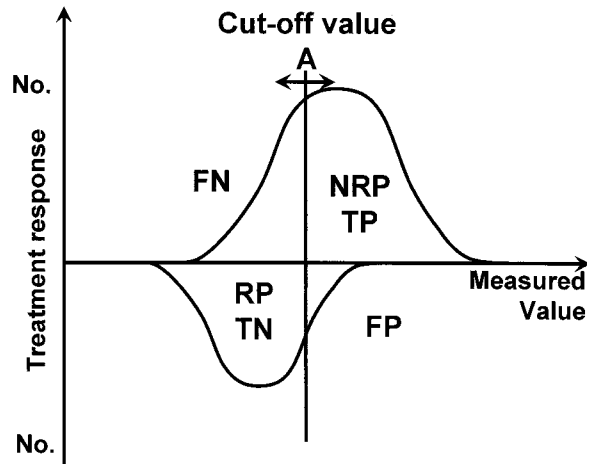


Figure 5. Graphical illustration of diagnostic performance of ¹⁸F-FDG PET for assessment of treatment response in solid tumors shows high positive predictive value but very low negative predictive value mainly due to high pre-test probability for nonresponder (NRP) in contrast to that of lymphoma in Figure 4 (RP, responders; TP, true positive; FP, false positive; TN, true negative; FN, false negative).

포 범위가 넓을수록 좋은 검사법이라 할 수 있다.

이와 같이 진단 기준에 따라 민감도와 특이도가 변화하는 상호 관계를 그래프로 표현한 것이 receiver operating characteristic (ROC) 곡선이다(Fig. 3). ROC 곡선은 진단 검사의 진단 성능을 평가하는 방법으로서, 곡선 아래부분의 면적이 넓을수록 우수한 검사이다. ROC 곡선 A로 표현되

Table 5. Prognostic Values and Diagnostic Performance of ¹⁸F-FDG PET in Lymphoma

Authors	Year	Chemo. cycles ^{a)}	Pts no ^{b)}	Type	PET results (Pts no)				RP ^{d)} (%)	Sens ^{e)}	Spec ^{f)}	PPV ^{g)}	NPV ^{h)}
					Positive		Negative						
					Total	Tx. fail ^{c)}	Total	Tx. fail					
Jerusalem	2000	2 or 3	28	NHL ^{g)}	5	5	23	7	57	42	100	100	70
Mikhaeel	2000	2-4	23	NHL	8	7	15	0	70	100	94	88	100
Spaepen	2002	3 or 4	70	NHL	33	33	37	6	44	85	100	100	84
Hoekstra	1993	1 or 2	13,13	HL ^{g)} ,NHL	10	7	16	1	69	88	83	70	94
Kostakoglu	2002	1	13,17	HL,NHL	15	13	15	2	50	87	87	87	87
Torizuka	2004	1 or 2	3,17	HL,NHL	16	14	4	2	20	88	50	88	50
Mikhaeel	2005	2 or 3	121	NHL	52	37	69	12	60	76	79	71	83
Haioun	2005	2	90	NHL	36	22	54	11	63	67	75	61	80
Kostakoglu	2006	1	23,24	HL,NHL	16	14	31	0	70	100	94	88	100
Hutchings	2005	2 or 3	85	HL	13	8	72	4	86	67	93	62	94
Hutchings	2006	2	77	HL	16	11	61	3	82	79	92	69	95
Zinzani	2006	2	40	HL	8	7	32	1	80	88	97	88	97
Gallamini	2006	2	108	HL	20	18	88	3	81	86	98	90	97
Gallamini	2007	2	260	HL	50	43	210	10	80	81	97	86	95

^{a)}cycles of chemotherapy, ^{b)}number of patients, ^{c)}failure of therapy, ^{d)}responder; ^{e)}sensitivity, ^{f)}specificity, ^{g)}positive predictive value, ^{h)}negative predictive value, ^{g)}non-Hodgkin's lymphoma, ^{h)}Hodgkin's lymphoma.

는 검사가 가장 우수한 검사이고, 직선 C로 표현되는 검사는 직선 아래부분의 면적(Area under curve, AUC)가 0.5이므로 동전을 던져 양성 또는 음성으로 판정하는 것과 다름 없는 경우이다. 진단 기준을 설정할 때, 민감도와 특이도 어느 한 쪽으로 치우치지 않는 최적의 균형값을 취하게 된다. Fig. 3에서 민감도 100%인 점(Y축 맨 상단)과 위양성 0%인 점(Y축 맨 하단), 두 점 모두로부터 가장 가까운 거리에 있는, ROC 곡선 상의 어떤 점이 민감도와 특이도가 각각 최대가 되는, 최적의 균형을 이루는 값이 된다. ROC 곡선이 주어지면 곡선상의 어느 지점에서의 우도비를 구할 수 있고 이를 이용해서 검사 후 병의 확률을 계산할 수 있다.

진단 기준 값이 일단 결정 되면 ROC 곡선에서 보듯이 민감도와 특이도는 검사 대상 집단의 크기에 전혀 무관하고 영향을 받지 않는다. 그러나 양성 예측도와 음성 예측도는 민감도와 특이도가 각각 동일 수준을 유지하더라도 검사 대상군의 유병률 또는 검사 전 확률에 따라 달라진다. Fig. 4와 Fig. 5에서와 같이 민감도와 특이도가 각각 동일한 수준을 유지하더라도 정상군(치료 반응군)과 비정상군(치료 실패군)의 크기가 달라짐에 따라 양성 예측도와 음성 예측도는 각각 달라지게 된다. 예를 들어, 민감도 90%, 특이도 80%가 동일하게 유지되더라도 검사 대상 환자 중 정상군의 수가 비정상군보다 훨씬 많을 때는 음성 예측도가 증가하고 반대로, 비정상군의 수가 정상군보다 훨씬 많을 때는 양성 예측도가 증가하게 된다.

림프종 환자에서 항암 화학 요법과 방사선 치료에 대한 종양의 반응을 평가한 ¹⁸F-FDG PET의 진단 성능에 대한 여러 보고들이 Table 5에 요약되어있다. 림프종 특히, 호지

킨 림프종은 항암 화학요법의 치료 효과가 매우 좋아서 치료 반응율이 80-90%에 이른다. ¹⁸F-FDG PET이 림프종의 치료 반응을 평가하는데 매우 이상적인 검사법이라면 검사 대상 환자의 약 20%를 치료에 반응하지 않는 환자, 즉 양성으로 판정할 것이고, 약 80% 환자는 치료에 잘 반응하는 환자, 음성으로 판정할 것이다. 이 글에서는 ¹⁸F-FDG PET 검사 결과가 비정상, 즉 비정상적인 ¹⁸F-FDG 섭취가 관찰될 때를 양성으로 간주하고, 비정상적인 ¹⁸F-FDG 섭취가 관찰되지 않거나 진단 기준 이하로 FDG 섭취가 현저하게 감소되었을 경우를 음성으로 규정하고자 한다. 다른 일반 진단 검사의 양성과 음성의 개념과 동일하다. ¹⁸F-FDG PET을 이용하여 치료 반응을 평가하는 여러 연구에서 ¹⁸F-FDG PET 검사 결과를 양성으로 판정하는 정의가 서로 달라서 혼동하기 쉬우므로 무엇을 양성으로 정의하였는지를 확인하고 주의하여야 할 필요가 있다.

Table 6은 호지킨 림프종에서 ¹⁸F-FDG PET의 진단 성능의 예를 보여준다. 호지킨 림프종 100명의 환자를 대상으로 항암화학요법 치료를 하여 그 중 80명이 일차 완전 관해에 도달하였고, 그 중 76명이 ¹⁸F-FDG PET 검사 결과 ¹⁸F-FDG 섭취가 관찰되지 않아 음성으로 판정되었고 나머지 20명은 치료에 반응하지 않고 재발 하였으며 그 중 16명은 섭취가 지속적으로 관찰되어 양성으로 판정 되었다. 따라서 호지킨 림프종 환자의 화학요법 치료에 대한 종양의 치료 반응을 평가할 때 치료에 반응하지 않는, 즉 치료 실패 환자를 찾아내는 ¹⁸F-FDG PET의 민감도는 80%, 특이도는 95%, 양성예측도 80%, 음성예측도는 95%이다. Fig. 4와 Table 6에서 보듯이 호지킨 림프종의 경우 치료 성공률

Table 6. Diagnostic Performance of ¹⁸F-FDG PET for Assessment of Tumor Response to Therapy in Lymphoma

	NRP ^{a)}	RP ^{b)}	Total
PET (+)	16	4	20
PET (-)	4	76	80
Total	20	80	

Sensitivity = 80% (16/20), specificity = 95% (76/80), positive predictive value = 80% (16/20), negative predictive value = 95% (76/80).

^{a)}non-responder, ^{b)}responder.

Table 8. Calculation of Interval Likelihood Ratios

	Disease (+)	Disease (-)
Test (+)	TP ^{a)}	FP ^{b)}
Test (-)	FN ^{c)}	TN ^{d)}

^{a)}true positive, ^{b)}false positive, ^{c)}false negative, ^{d)}true negative.

이 높기 때문에 즉, 검사 전 가능성이 80%로 높기 때문에 음성 예측도가 95%로 매우 높게 나타난다. 이런 양상을 그 그래프로 표현한 것이 Fig. 4이다. 일반적으로 특이도가 높으면 음성 예측도가 낮게 되는데, 이 경우 특이도와 음성 예측도가 모두 매우 높다는 것은 호지킨 림프종 자체의 생물학적 특성, 화학요법의 탁월한 치료 효능 또는 이 두 요인이 복합적으로 작용하여 ¹⁸F-FDG PET 검사에 반영된 것으로 설명된다. 따라서 호지킨 림프종 환자에서 ¹⁸F-FDG PET 검사 결과가 음성으로 판정되면 거의 틀림 없이 병이 없으며(음성예측도 95%), 또한 병이 없는 환자(치료에 반응한 환자)는 거의 다 찾아낸다(특이도 95%)라고 말할 수 있다. '보이지 않으면(거의 틀림없이 병이) 없고'(치료에 대한 반응이 있다) 또한 '안보이는 것은(거의 다) 찾아낸다'.

따라서 호지킨 림프종에서의 ¹⁸F-FDG PET의 역할은 80%에 달하는 치료에 잘 반응하는 환자를 찾아내서 계획된 치료를 계속 진행 하거나, 덜 강력하고 합병증이 적은 처방계획으로 교체하여 치료함으로써 ¹⁸F-FDG PET 반응 맞춤 치료를 하는데 있다고 할 수 있다. 호지킨 림프종의 치료 반응 평가시 ¹⁸F-FDG PET 검사는 특이도와 음성 예측도가 모두 90-95% 이상으로 매우 높은 진단 특성을 보여 주므로 확진적 검사의 성격이 매우 강하다고 할 수 있다. 림프종 특히, 호지킨 림프종은 완치 가능성이 매우 높으나 화학요법 치료에 의한 2차 합병증으로 인해 생존률이 심각하게 위협받는 질환이다. 그래서 치료에 잘 반응하는 환자는 치료제를 조절하여 합병증으로 인한 위험을 최소화하는, 즉, 위험 관리 맞춤 치료가 최근의 패러다임이고 이런 관점에서 ¹⁸F-FDG PET 검사가 중요한 역할을 할 것으로 기대

Table 7. Diagnostic Performance of ¹⁸F-FDG PET for Assessment of Tumor Response to Therapy in Solid Tumor

	NRP ^{a)}	RP ^{b)}	Total
PET (+)	49	3	52
PET (-)	21	27	48
Total	70	30	

Sensitivity = 70% (49/70), specificity = 90% (27/30), positive predictive value = 94% (49/52), negative predictive value = 56% (27/48).

^{a)}non-responder, ^{b)}responder.

된다. 따라서 호지킨 림프종 환자의 ¹⁸F-FDG PET 검사 결과를 판독할 때는 이런 점을 고려하여 치료에 반응하는 환자를 정확하게 가려내는 즉, 음성 예측도를 향상시키는 데 주안점을 두는 것이 바람직하다.

이와는 대조적으로 폐암, 식도암, 위암, 유방암과 같은 고형암에서는 림프종과 상황이 반대가 된다. 림프종과는 달리 폐암, 유방암, 위장관암과 같은 고형암들은 화학요법 또는 화학방사선 요법에 대한 치료 효과가 높지 않아서 병리 조직학적으로 완전 반응에 이르는 비율이 20-40%로 보고되고 있다. 고형암 환자 100명을 대상으로 화학요법 또는 화학방사선요법 치료를 했을 경우 70명의 환자(검사 전 확률 70%)가 치료에 반응하지 않을 것이고 그 중 49명이 ¹⁸F-FDG PET 검사 결과 ¹⁸F-FDG 섭취가 지속적으로 관찰되어 양성, 치료에 반응하지 않는 환자로 판정되었고, 나머지 30%의 환자가 치료에 반응하여 그 중 27명이 ¹⁸F-FDG PET 검사 결과 음성으로 판정되었다고 하면 고형암 환자에서 치료 반응을 평가할 때, 치료에 반응하지 않는 ¹⁸F-FDG PET 양성 환자를 찾아내는 ¹⁸F-FDG PET의 민감도는 70%이고 특이도는 90%이다. 또한 양성 예측도는 94%이고, 음성 예측도는 56%에 불과하다(Table 7). 즉, 고형암에서 ¹⁸F-FDG PET 검사 결과 비정상적인 ¹⁸F-FDG 섭취가 관찰되어 ¹⁸F-FDG PET 양성으로 판정한 환자 중 그 판정이 참일 확률은 94%인 반면, 그 반대로 판정했을 때 그 판정이 옳을 확률은 56%에 불과하다. 고형암의 치료 반응 평가 시 ¹⁸F-FDG PET의 진단 성능 특성이 Fig. 5에 그래프로 표현되어 있다. 다시 말해 림프종과는 달리 고형암에서는 PET 결과가 양성이면 최종 치료 실패를 예측하는 확률이 높으나, 검사 결과가 음성으로 판정되었을 때 최종 결과와 일치하는 확률은 56%에 불과하다는 뜻이다. '보이면(병이) 있다' 그러나 '보이지 않으면(병의 존재가) 반반이다'.

이와 같이 고형암에서는 림프종과는 반대로 ¹⁸F-FDG PET 검사의 역할은, 치료환자의 대부분을 차지하는, 치료에 반응하지 않고 치료 실패 가능성이 높은 환자를 가려내

Table 9. Interval Likelihood Ratios of ¹⁸F-FDG PET for Characterization of Solitary Pulmonary Nodule

Category	+ ILRs ^{a)}	- ILRs ^{b)}
Definitely benign	1.49 (1.33-1.66) ^{c)}	0.03 (0.01-0.12)
Probably benign	1.81 (1.55-2.11)	0.15 (0.09-0.25)
Indeterminate	0.29 (0.03-2.71)	1.01 (1.00-1.02)
Probably malignant	3.15 (1.89-5.25)	0.77 (0.69-0.84)
Definitely malignant	9.90 (5.37-18.26)	0.42 (0.35-0.50)
Definitely benign + probably benign ^{d)}	-	0.10 (0.06-0.16)
Indeterminate + probably malignant + definitely malignant ^{d)}	5.18 (3.72-7.22)	-
Sensitivity (%)	91.7 (86.6-95.0)	-
Specificity (%)	82.3 (75.4-87.6)	-

^{a)}Interval likelihood ratios associated with positive (abnormal) test result, ^{b)}Interval likelihood ratios associated with negative (normal) test result, ^{c)}95% confidence intervals, ^{d)}combined reading.

어 효과가 없는 처방계획의 사용을 중단하여 합병증으로 인한 위험을 감소시키고, 치료 초기에 처방계획을 바꾸거나 치료 강화를 통해 환자 개개인의 화학요법 민감도와 방사선 요법 민감도에 따라 ¹⁸F-FDG PET 반응 맞춤형 치료를 하는데 있다고 할 수 있다.

일반적으로 고형암에서는 민감도가 낮은 반면 양성 예측도가 높고 음성 예측도가 림프종에 비해 현저히 낮다. 다소의도적인 그러나 균형을 맞춘 노력을 통해 민감도와 음성 예측도는 향상시킬 수 있으나 위양성율이 증가하여 양성 예측도가 감소하는 역기능을 감수해야 한다.

이와 같이 ¹⁸F-FDG PET 검사의 진단 성능 특성은 암의 종류 및 생물학적 특성, 특정 암에 대한 특정 치료법의 효능, 유병률, 치료의 목적에 따라 달라지므로 이를 고려하여 판독에 반영해야 한다.

구간 우도비(Interval Likelihood Ratio)

림프종의 치료반응 평가 시 대부분의 연구자들이 육안적 판정이거나 반정량적인 방법을 사용하던 간에 ¹⁸F-FDG PET 검사 결과를 양성 또는 음성으로 판독하고 있다. 최근에 발표된 IWC+PET 기준 또한 양성 또는 음성의 이분법으로 판정하는 방식을 적용하고 있다. ¹⁸F-FDG PET으로 측정된 치료에 대한 중앙의 반응은 생래적으로 연속변수이므로 인위적으로 양성과 음성 두 카테고리 나누는 것은 근본적으로 한계가 있다. 다양한 범위에서 측정된 값 속에 내포된 중요한 정보를 잃을 수 있기 때문이다. 일상 진료에서 정상에서 크게 벗어난 검사 결과를 접할 때 그 측정값에 내재된 의미를 우리는 직관적으로 파악하고 환자의 생리적, 병적 상태의 경중을 가늠하게 된다. 구간 우도비는 이러한 직관성을 이론적으로 체계화하여 비정상성의 정도를 보다 구체적으로 표현하는 해석방법이라고 할 수 있다.

구간 우도비는 진단적 검사의 특성을 기술하는 한 방법

으로서 검사 결과들을 여러 구간으로 등급화하여 각 구간마다 질병의 확률을 평가하는 방법이다.^{35,36)} Table 8에 가장 간단한 2×2 표로부터 구간 우도비를 계산하는 방법이 제시되어있다. 양성 구간 우도비는 민감도(진양성률)/위양성률이고 음성 구간 우도비는 위음성률/진음성률이다. 또한 양성 및 음성 구간 우도비를 아래와 같이 구할 수도 있다.

$$\begin{aligned} \text{False positive rate (FPR)} &= \text{FP}/(\text{FP}+\text{TN}) \\ &= 1-\text{specificity} \end{aligned}$$

$$\begin{aligned} \text{False negative rate (FNR)} &= \text{FN}/(\text{FN}+\text{TP}) \\ &= 1-\text{sensitivity} \end{aligned}$$

$$\begin{aligned} \text{Positive likelihood ratio} &= [\text{TP}/(\text{TP}+\text{FN})]/[\text{FP}/(\text{TN}+\text{FP})] \\ &= \text{sensitivity}/\text{FPR} = \text{sensitivity}/(1-\text{specificity}) \end{aligned}$$

$$\begin{aligned} \text{Negative likelihood ratio} &= [\text{FN}/(\text{TP}+\text{FN})]/[\text{TN}/(\text{TN}+\text{FP})] \\ &= \text{FNR}/\text{specificity} = (1-\text{sensitivity})/\text{specificity} \end{aligned}$$

최근에 단일 폐 결절을 진단하는 ¹⁸F-FDG PET의 진단 성능을 구간 우도비를 이용하여 diagnostic CT와 비교 평가한 연구가 Fletcher 등에 의해 보고되었다.³⁷⁾ 흉부 X-선 검사로 처음 진단된 단일 폐 결절 환자 344명을 대상으로 미국 전역에 분포한 10개 보훈병원에서 ¹⁸F-FDG PET과 diagnostic CT를 각각 시행한 전향적 연구에서 미리 설정된 ¹⁸F-FDG PET과 CT의 진단 기준에 따라 명확한 양성(definitely benign), 양성 가능(probably benign), 불명확(indeterminate), 악성 가능(probably malignancy), 명확한 악성(definitely malignancy)의 다섯 등급으로 나누어 판독하였다. ¹⁸F-FDG PET 검사 결과의 판독 기준을 예를 들면 다음과 같다. 명확한 양성: ¹⁸F-FDG 섭취 증가가 없고 폐 조직과 섭취 정도가 비슷할 때(SUV, 0.6-0.8), 양성 가능: 혈액 풀 섭취보다는 충분히 낮지만 폐 조직의 섭취보다는 증가한 경우(SUV, 0.6-0.8이상, 1.5-2.0이하), 불명확: ¹⁸F-FDG 섭취가 폐 조직의 2-3배 이지만, 혈액 풀 보다는

낮은 경우(SUV, 1.5-2.0내외이나 2.5미만), 양성 가능; 섭취가 혈액 풀 보다 높은 경우(SUV, 2.5이상), 명확한 양성; 섭취가 혈액 풀보다 훨씬 증가된 경우(SUV, 2.5보다 많이 증가).

이와 같이 5 구간으로 나누어 판독하여 계산된 구간 별 양성 및 음성 구간 우도비는 Table 9에 제시되어 있다. Table 9에서 양성 구간 우도비가 9.90이라는 의미는 명확한 양성이라고 판독했을 때, 양성 폐 결절을 명확한 양성이라고 판독하는 확률보다 폐암의 확률이 약 10배 높다는 뜻이다. 다시 말해, 명확한 양성이라고 판독한 병변은 양성 폐 결절의 가능성보다 폐암의 가능성이 10배 더 높다는 의미이다. 양성 가능이라고 판독했을 때, 양성 결절을 양성 가능이라고 판독하는 확률 보다 약 3배가 더 높다는 의미다. 즉, 양성 가능이라고 판독한 병변은 양성 폐 결절 보다는 폐암의 가능성이 3배 높다는 뜻이다. ¹⁸F-FDG PET 검사 결과가 폐 결절은 ¹⁸F-FDG PET 양성(폐암) 이고 이 검사의 민감도는 92%, 양성 예측도는 86%이다' 라고 판독하는 것에 비해 '이 폐 결절은 양성 폐 결절일 가능성보다 폐암일 가능성이 10배 높고, 또는 이 폐 결절은 양성결절보다는 폐암일 가능성이 3배 높다' 라고 판독하는 것이 임상 의사가 검사 결과의 의미를 더 이해하기 쉽고 피부에 와 닿는 판독이라고 느낄 것이다.

또한, 우도비를 구하면 검사 후 확률을 곧바로 구할 수 있다. 검사 후 확률은 Fagan의 normogram이나 ROC 곡선을 이용하고, 이 두 가지가 없으면 Baye theorem의 likelihood ratio version을 이용한 아래의 수식으로부터 구할 수 있다.

$$(\text{pre-test odds}) \times (\text{구간 우도비}) = (\text{post-test odds})$$

이 수식을 이용하기 위해서는 pre-test 및 post-test odds를 각각 pre-test probability와 post-test probability로 바꿔야 한다. Odds는 어떤 사건이 발생할 확률을 그 사건이 발생하지 않을 확률로 나눈 것으로 아래와 같이 주어진다.

$$\text{odds} = \text{probability}/(1-\text{probability})$$

단일 폐 결절의 진단성능 연구에서 연구 대상군의 폐암 유병률은 53%였다(검사 전 확률). 명확한 악성으로 판독한 경우 구간 우도비는 약 10이고 검사전 확률은 0.53이므로 post-test odds는 다음과 같다.

$$0.53/0.47 \times 10 = 11.3 (\text{post-test odds})$$

위의 첫번째 수식을 재배열 하면 아래와 같이 된다.

$$\text{probability} = \text{odds}/(\text{odds}+1) = 11.3/(11.3+1) = 92 (\%)$$

따라서 명확한 악성이라고 판독했을 때 검사 후 확률은 92%가 되고 검사 전 확률 53%에 비해 약 40%가 향상된 것이다. 마찬가지로 양성 가능이라고 판독했을 때 구간 우

도비는 약 3이므로 검사 후 확률은 77%가 된다.

연속성을 갖는 변수를 단 하나의 판독 기준으로 양성 또는 음성 두 구간으로 나누어 결과를 보고하는 것 보다 여러 구간으로 나누어서, 판독 기준에 가까운 영역에서 멀리 떨어져 분포하는 변수들에 내재되어있는 의미를 포착할 수 있는 구간 우도비 방법이 이분법 판독 방식보다 큰 장점을 가지고 있다. 물론 양성예측도 또는 음성예측도가 판독 결과 질병의 유무에 대한 확률을 얘기해주는 하지만 판독 기준에 근접한 경계영역의 변수와 멀리 떨어져있는 변수를 동일하게 취급하여 양성 또는 음성으로 판독하므로 멀리 떨어져 분포하는 변수들의 가치가 희석되는 단점이 있다. 사실 Hutchings와 Mikheel이 호지킨 림프종과 비호지킨 림프종 환자에서 2-3주기의 항암화학요법 후 치료 반응을 평가할 때 양성과 음성 그리고 중간 영역에 해당되는 정도 잔존 섭취군(minimal residual uptake, MRU), 세 군으로 나누어 판독한 결과를 보고한 바 있다. 비호지킨 림프종 121명을 대상으로 한 후향적 연구에서 121명 중 MRU가 19명(15%)이었고, 세 분류 사이에 PFS와 OS 사이에 통계적으로 유의한 차이가 있음을 확인한 바 있다.^{38,39)} 이 결과는 ¹⁸F-FDG PET 검사 결과를 단순히 양성과 음성의 이분법으로 판독하는 것보다 몇 개의 분류로 나누어 판독하는 것이 임상적으로 유용한 등급화가 가능할 수 있음을 시사하는 소견이다. 특히 림프종 환자에서는 고형암과 달리 치료 초기에 시행한 ¹⁸F-FDG PET 검사에서도 정량적 변수들이 현저하게 변화하므로 여러 분류로 나누는 것이 고형암보다 용이할지도 모른다. 그러나 몇 개의 구간으로 나누어야 하는지 또 각 구간의 한계값은 얼마로 해야 할 지에 대해서는 아직 확립된 자료가 없다. 연속성 변수를 인위적으로 양성 또는 음성, 정상 또는 비정상의 이분법적으로 판독함으로써 정량변수 속에 내재된 중요한 정보를 활용하지 못한다는 한계점과 구간 우도비 개념의 장점을 인식하여 Wahl 등 일부 연구자들이 고형암에서 여러 구간으로 나누어 판독하는 방식을 시도하고 있다.^{38,40)} 비록, 몇 개의 등급으로 나누어 판독하는 방식을 적용하는 데 필요한, 구체적인 확립된 자료가 현재 없고 가까운 장래에 임상 연구 자료들이 충분히 축적되기는 쉽지 않아 보이기 때문에 다소 주관적일지라도 ¹⁸F-FDG PET 검사 결과를 등급화 하여 판독하여 그 의미가 임상 의사에게 잘 전달되도록 시도하는 것이 바람직한 방향으로 판단된다.

종양의 치료 반응을 평가할 때 종양의 종류와 특정 치료제의 치료 효능에 따라 ¹⁸F-FDG PET의 진단 성능에 차이가 있음을 이해하고, ¹⁸F-FDG PET 검사 또한 치료 반응을 평가할 수 있는 다른 여러 가지 방법, 즉 형태학적 영상검

사, 분자생물학적 표지자, 혈청 내 생물학적 표지자, 유전자 발현 검사, 임상소견, 검사실 소견 중의 한 가지 방법이라는 것을 이해하고 가능한 모든 정보를 고려하여 판독하는 것이 바람직하다.

결론

최근까지 치료반응 평가는 종양의 크기 변화를 평가하는데 주안점을 두어왔고, 종양의 크기 측정은 CT, MRI, 초음파와 같은 형태학적 영상 진단법에 거의 전적으로 의존해왔다. ¹⁸F-FDG PET의 임상응용이 발전함에 따라 형태학적 반응 평가의 한계점이 점차 부각되어 이를 보완하기 위해 ¹⁸F-FDG PET의 적용이 증가하게 되었다. 그러나 ¹⁸F-FDG PET 또한 제한점이 있으므로 그 제한점과 핵의학 전문가가 판독을 할 때 여러 임상 정보가 필요하다는 사실을 임상가가 이해하는 것 또한 중요하다.

최근에 제안된 림프종의 치료반응 기준에서 ¹⁸F-FDG PET 검사가 중요한 역할을 담당하게 되었다. 림프종이 ¹⁸F-FDG PET 결과를 바탕으로 맞춤 치료를 할 수 있는 최초의 질환이 될 가능성이 높아 보이고 기대된다. 이러한 ¹⁸F-FDG PET의 역할이 잘 수행되기 위해서는 반응 평가 기준 권고안에서 제시된, 표준화된 ¹⁸F-FDG PET 검사방법, SUV 측정방법 등에 따라 검사와 판독이 이루어져야 할 것이다. 또한 반응 평가에 영향을 미치는 여러 요인들을 잘 숙지하고, ¹⁸F-FDG PET 진단 성능의 특성과 구간 우도비의 개념을 적용하여 임상 상황에 적절한 판독을 하고 ¹⁸F-FDG PET으로 평가된 치료 반응이 연속성 변수임을 감안하여 정량적 결과에 내재된 의미가 임상외에게 잘 전달되도록 노력하는 것이 바람직한 판독 접근 방향이라고 생각된다.

References

1. Swerdlow SH. WHO Classification of tumours of haematopoietic and lymphoid tissues. 4th ed. Lyon: IARC; 2008. p. 158-319.
2. Fauci AS. Harrison's principles of internal medicine. 17th ed. Seoul: McGrawHill; 2008. p. 688.
3. Lister TA, Crowther D, Sutcliffe SB, Glatstein E, Canellos GP, Young RC, et al. Report of a committee convened to discuss the evaluation and staging of patients with Hodgkin's disease: Cotswolds meeting. *J Clin Oncol* 1989;7:1630-6.
4. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med* 1993;329:987-94.
5. Hasenclever D, Diehl V. A prognostic score for advanced Hodgkin's disease. International Prognostic Factors Project on Advanced Hodgkin's Disease. *N Engl J Med* 1998;339:1506-14.

6. Gallamini A, Hutchings M, Rigacci L, Specht L, Merli F, Hansen M, et al. Early interim 2-[¹⁸F]fluoro-2-deoxy-D-glucose positron emission tomography is prognostically superior to international prognostic score in advanced-stage Hodgkin's lymphoma: a report from a joint Italian-Danish study. *J Clin Oncol* 2007;25:3746-52.
7. Spaepen K, Stroobants S, Dupont P, Vandenberghe P, Thomas J, de Groot T, et al. Early restaging positron emission tomography with (18)F-fluorodeoxyglucose predicts outcome in patients with aggressive non-Hodgkin's lymphoma. *Ann Oncol* 2002;13:1356-63.
8. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;47:207-14.
9. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 2000;92:205-16.
10. Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, et al. Measurement of clinical and subclinical tumour response using [¹⁸F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999;35:1773-82.
11. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228-47.
12. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med* 2009;50S:122S-50S.
13. Cheson BD, Horning SJ, Coiffier B, Shipp MA, Fisher RI, Connors JM, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *J Clin Oncol* 1999;17:1244.
14. Cheson BD, Pfistner B, Juweid ME, Gascoyne RD, Specht L, Horning SJ, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol* 2007;25:579-86.
15. Juweid ME, Stroobants S, Hoekstra OS, Mottaghy FM, Dietlein M, Guermazi A, et al. Use of positron emission tomography for response assessment of lymphoma: consensus of the Imaging Subcommittee of International Harmonization Project in Lymphoma. *J Clin Oncol* 2007;25:571-8.
16. Canellos GP. Residual mass in lymphoma may not be residual disease. *J Clin Oncol* 1988;6:931-3.
17. Kasamon YL, Wahl RL. FDG PET and risk-adapted therapy in Hodgkin's and non-Hodgkin's lymphoma. *Curr Opin Oncol* 2008;20:206-19.
18. Weber WA. Assessing tumor response to therapy. *J Nucl Med* 2009;50S:1S-10S.
19. Erasmus JJ, Gladish GW, Broemeling L, Sabloff BS, Truong MT, Herbst RS, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. *J Clin Oncol* 2003;21:2574-82.
20. Hutchings M, Barrington SF. PET/CT for therapy response assessment in lymphoma. *J Nucl Med* 2009;50S:21S-30S.
21. Mikhael NG, Hutchings M, Fields PA, O'Doherty MJ, Timothy AR. FDG-PET after two to three cycles of chemotherapy predicts progression-free and overall survival in high-grade non-Hodgkin lymphoma. *Ann Oncol* 2005;16:1514-23.

22. Hutchings M, Mikhael NG, Fields PA, Nunan T, Timothy AR. Prognostic value of interim FDG-PET after two or three cycles of chemotherapy in Hodgkin lymphoma. *Ann Oncol* 2005;16:1160-8.
23. Hutchings M, Loft A, Hansen M, Pedersen LM, Buhl T, Jurlander J, et al. FDG-PET after two cycles of chemotherapy predicts treatment failure and progression-free survival in Hodgkin lymphoma. *Blood* 2006;107:52-9.
24. Gallamini A, Rigacci L, Merli F, Nassi L, Bosi A, Capodanno I, et al. The predictive value of positron emission tomography scanning performed after two courses of standard therapy on treatment outcome in advanced stage Hodgkin's disease. *Haematologica* 2006;91:475-81.
25. Juweid ME, Wiseman GA, Vose JM, Ritchie JM, Menda Y, Wooldridge JE, et al. Response assessment of aggressive non-Hodgkin's lymphoma by integrated International Workshop Criteria and fluorine-18-fluorodeoxyglucose positron emission tomography. *J Clin Oncol* 2005;23:4652-61.
26. Brepoels L, Stroobants S, Verhoef G. PET and PET/CT for response evaluation in lymphoma: current practice and developments. *Leuk Lymphoma* 2007;48:270-82.
27. Guay C, Lepine M, Verreault J, Benard F. Prognostic value of PET using ^{18}F -FDG in Hodgkin's disease for posttreatment evaluation. *J Nucl Med* 2003;44:1225-31.
28. Picardi M, De Renzo A, Pane F, Nicolai E, Pacelli R, Salvatore M, et al. Randomized comparison of consolidation radiation versus observation in bulky Hodgkin's lymphoma with post-chemotherapy negative positron emission tomography scans. *Leuk Lymphoma* 2007;48:1721-7.
29. Dittmann H, Sokler M, Kollmannsberger C, Dohmen BM, Baumann C, Kopp A, et al. Comparison of ^{18}F -FDG-PET with CT scans in the evaluation of patients with residual and recurrent Hodgkin's lymphoma. *Oncol Rep* 2001;8:1393-9.
30. Jerusalem G, Beguin Y, Fassotte MF, Belhocine T, Hustinx R, Rigo P, et al. Early detection of relapse by whole-body positron emission tomography in the follow-up of patients with Hodgkin's disease. *Ann Oncol* 2003;14:123-30.
31. Zinzani PL, Tani M, Trisolini R, Fanti S, Stefoni V, Alifano M, et al. Histological verification of positive positron emission tomography findings in the follow-up of patients with mediastinal lymphoma. *Haematologica* 2007;92:771-7.
32. Brepoels L, Stroobants S, De Wever W, Spaepen K, Vandenberghe P, Thomas J, et al. Hodgkin lymphoma: Response assessment by revised International Workshop Criteria. *Leuk Lymphoma* 2007;48:1539-47.
33. Brepoels L, Stroobants S, De Wever W, Spaepen K, Vandenberghe P, Thomas J, et al. Aggressive and indolent non-Hodgkin's lymphoma: response assessment by integrated international workshop criteria. *Leuk Lymphoma* 2007;48:1522-30.
34. Kasamon YL, Jones RJ, Wahl RL. Integrating PET and PET/CT into the risk-adapted therapy of lymphoma. *J Nucl Med* 2007;48S:19S-27S.
35. Sonis J. How to use and interpret interval likelihood ratios. *Fam Med* 1999;31:432-7.
36. Brown MD, Reeves MJ. Evidence-based emergency medicine/skills for evidence-based emergency care. Interval likelihood ratios: another advantage for the evidence-based diagnostician. *Ann Emerg Med* 2003;42:292-7.
37. Fletcher JW, Kymes SM, Gould M, Alazraki N, Coleman RE, Lowe VJ, et al. A comparison of the diagnostic accuracy of ^{18}F -FDG PET and CT in the characterization of solitary pulmonary nodules. *J Nucl Med* 2008;49:179-85.
38. Kasamon YL, Wahl RL, Ziessman HA, Blackford AL, Goodman SN, Fidyk CA, et al. Phase II study of risk-adapted therapy of newly diagnosed, aggressive non-Hodgkin lymphoma based on midtreatment FDG-PET scanning. *Biol Blood Marrow Transplant* 2009;15:242-8.
39. Ryu JS, Choi NC, Fischman AJ, Lynch TJ, Mathisen DJ. FDG-PET in staging and restaging non-small cell lung cancer after neoadjuvant chemoradiotherapy: correlation with histopathology. *Lung Cancer* 2002;35:179-87.
40. Melton GB, Lavelly WC, Jacene HA, Schulick RD, Choti MA, Wahl RL, et al. Efficacy of preoperative combined 18-fluorodeoxyglucose positron emission tomography and computed tomography for assessing primary rectal cancer response to neoadjuvant therapy. *J Gastrointest Surg* 2007;11:961-9.