

수신자 특성곡선 분석방법으로 입증하지 못한 폐종양 F-18-FDG 동적 PET의 유용성

서울대학교 의과대학 핵의학교실

이동수 · 장명진

본 문

폐종양의 악성여부를 감별하기 위한 검사 방법으로 문헌보고를 종합한 예민도와 특이도는 모두 94%, 83%이었다.¹⁻³⁾ 드물게 나타나는 위음성 예로 반흔에 생긴 종양(Scar cancer), 기관지폐포암 등이 있고 위양성 예로 폐렴, 국균증, 히스토플라즈마증, 폐결핵이 있다. 우리 나라에서는 폐결핵이 위양성 예로 문제가 된다.

국내 외의 여러 기관의 보고를 종합하면 고립성 폐결절의 15~51%가 암종이다.⁴⁾ 서울대학교병원에서 가장 최근 보고된 결과에 따르면 고립성 폐결절로 검사한 환자 중 1/3이 양성결절, 2/3가 악성결절이었다. 양성결절 중 결핵종이 60%를 차지하였으므로 우리 나라에서는 고립성 폐결절 중에 결핵종과 암종의 감별이 가장 중요하다. 고립성 폐결절 중에 암종이 차지하는 비율이 1/2~2/3일 때 F-18-FDG PET으로 악성 또는 양성으로 판정하면 각각 예측률이 85~92%, 93~87%이다. 외국의 문헌보고를 종합하여도 열 명에 하나 꼴로 틀릴 수 있다.

육안판독 대신에 준정량적인 방법을 사용하거나 동적 PET으로 시간방사능곡선을 그리는 방법으로 PET의 감별능을 향상하려는 시도가 있다. 김 등⁵⁾은 추적자 동력학 해석을 더욱 정교하게 다듬어 악성과 양성 종양의 감별능을 향상하려 하였다. 준정량지

표인 표준화 섭취율이나 포도당대사율은 연속변수이므로 문턱 값을 정하고 지표의 성능을 조사하는데 수신자 특성곡선 분석 방법을 사용하는 것이 적절하다.⁶⁾

수신자 특성곡선을 이용하여 정량화 방법의 감별 진단성능을 조사하기 위하여 매개변수로 양성과 악성 각각을 진단하려면 곡선 아래 면적을 구하여 비교한다. Hanley와 McNeil⁶⁾은 Wilcoxon 통계량을 사용하여 수신자 특성곡선 아래 면적을 구하여 두 방법의 진단성능이 차이가 있는지 면적들이 정규분포를 한다고 가정하여 비교하기를 제안하였다. 이들은 김 등의 분석⁵⁾에서처럼 같은 환자에서 여러 방법으로 얻은 자료의 곡선 아래 면적을 구할 때 데이터가 쌍을 이룬 것을 고려하여 보정하여 검정력을 높였다.⁷⁾ 김 등은 더 나아가 여러 매개변수의 감별진단 성능의 차이를 조사하여 둘씩 비교하여 보고하였다. 그런데 여러 정량화 방법의 곡선아래면적을 동시에 비교할 때에는 그 정량화 방법 전체에 대한 분산(S)을 구한 뒤 분석을 하여야 한다. 이런 경우의 비교검정을 위하여 DeLong 등⁸⁾이 여러 개의 곡선아래면적을 함께 비교하는 방법을 제안하였다. 이 방법을 김 등의 데이터에 적용하면 다음과 같다.

악성종양 환자(13명)에 대해 r번째 정량화 방법(k개: k=7)으로 얻은 값을 X'_1, X'_2, \dots, X'_k 이라 하고 양성종양 환자(9명)에 대해 r번째 정량화 방법(k개: k=7)으로 얻은 값을 얻은 값을 Y'_1, Y'_2, \dots, Y'_k 라고 한다. Mann-Whitney 또는 Wilcoxon 통계량을 사용하여 각 정량화방법의 수신자 특성곡선 아래 면적을 구할 수 있다. r번째 정량화 방법에 대한 곡선아래면적($\hat{\theta}^r$)은 다음과 같이 구한다.

$$\hat{\theta}^r = \frac{1}{13 \cdot 9} \sum_{i=1}^{13} \sum_{j=1}^9 \psi(X'_i, Y'_j)$$

Received Nov. 2, 1998

Corresponding Author: Dong Soo Lee, M.D., Department of Nuclear Medicine, Seoul National University Hospital 28 Yungundong Chongnogu, Seoul 110-744 Korea
Tel: (02) 760-2501, Fax: (02) 766-9083
E-mail: dsl@plaza.snu.ac.kr

$$\text{이 때 } \psi(X, Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases}$$

$\hat{\theta} = (\hat{\theta}^1, \hat{\theta}^2, \dots, \hat{\theta}^k)$ 의 공분산행렬 추정치(S)를 구한 후 대조도(contrast) $L\theta'$ 에 대해 $(\hat{\theta} - \theta) L' [LSL']^{-1} L(\hat{\theta} - \theta)'$ 이 자유도가 rank(LSL')인 카이제곱 분포를 근사적으로 따른다는 사실을 이용하여 대조도 $L\theta'$ 에 대한 유의성을 검증한다. 대조도를 구성하는데 따라 어느 한 정량화 방법이 다른 것들보다 우수한지 평가할 수 있다. 즉 김 등의 자료 중에 파트락 도표에서 얻은 추적자 섭취상수(Kpat)와 당대사율(MRpat) 사이에 유의한 차이가 있다고 주장한 것이 적절한 결론인지 조사할 수 있다.

김 등은 자료를 분석할 때 두 가지 오류를 범했다. 7가지의 정량화 방법을 임의로 둘씩 비교하여 차이가 있는 것을 두 개의 수신자 특성곡선을 비교하는 통계를 사용하여 주장하였다. 임의로 둘씩 비교하려면 위의 설명과 같이 공분산행렬을 구하였어야 한다. 또한 이 결과로부터 비약하여 '추적자 섭취상수보다 혈중 당농도를 보정한 당대사율의 진단효율성과 정확성이 높다'고 주장하였다. 그런데 이 주장이 통계적으로 옳은지 보려면 당대사율이 섭취 상수들보다 파트락방법 모델, 5매개변수 또는 6매개변수 모델 모두에서 우수하다는 가정을 입증하였어야 한다. 입증하여야 하는 가설이 달라지므로 검증방법도 알맞게 제시하였어야 한다. 실제 임의로 비교한 곡선아래면적 중에 차이가 가장 큰 쌍에서 발견된 차이의 통계적 유의성 즉 z 값이 1.99인 것은 위의 두 가지 가설 어느 것도 입증하였다고 볼 수 없다.

그렇다면 이들의 데이터를 바탕으로 이들이 새로 개발한 방법이 기존 방법보다 우수하지 않다는 것이 입증되는가? 제 2형 오류 즉 차이가 있는데 찾지 못하고 귀무가설을 인정하는 오류를 피하고 차이가 없다는 결론을 내리려면 충분한 증례가 필요하고 그 증례수가 각각 얼마인지 제시되어 있다.⁷⁾ 곡선아래면적이 평균 0.7일 때 0.1의 차이가 유의한지를 검정하는데 김 등의 데이터는 많이 모자라다. 따라서 김 등의 데이터가 이들이 개발한 방법이 도움되지 않는다는 것을 입증한 것도 아니다.

이상의 논의에서 우리는 김 등의 데이터는 이들이 개발한 정교한 정량화 방법이 폐종양 F-18-FDG 동적 PET을 이용하여 악성여부를 감별 진단하는데 도움되는지를 증명하지도 못하였으며 도움되지 않음을 증명하지도 못하였다고 생각한다. 윤 등⁹⁾ Lee 등¹⁰⁾의 보고와 김 등⁵⁾의 보고를 종합하면 우리나라에서 주로 염증성 병변인 폐의 양성 종양이 F-18-FDG PET에 위양성으로 나타나는 것을 정량화 방법을 개선하는 것으로는 해결할 수 없어 보인다.

수신자특성 곡선 방법으로 분석한 결과 폐종양 F-18-FDG PET을 얻을 때 동적 영상획득방법을 적용하는 것이 악성과 양성의 감별진단에 도움된다는 증거가 아직 없다고 생각한다.

참 고 문 헌

- 1) Patz EF, Lowe VJ, Hoffman JM, Paine SS, Burrowes P, Coleman RE, et al. Focal pulmonary abnormalities: Evaluation with F-18 Fluorodeoxyglucose PET scanning. *Radiology* 1993;188:487-90.
- 2) Dewan NA, Gupta NC, Redepenning LS, Phalen JJ, Frick MP. Diagnostic efficacy of PET-FDG imaging in solitary pulmonary nodules. Potential role in evaluation and management. *Chest* 1993; 104:997-1002.
- 3) Gupta NC, Frank AR, Dewan NA, Redepenning LS, Rothberg ML, Mailliard JA, et al. Solitary pulmonary nodules: Detection of malignancy with PET with 2-[F-18]-fluoro-2-deoxy-D-glucose. *Radiology* 1992;184:441-4.
- 4) Ko WJ, Kim CH, Jang SH, Lee JH, Yoo CG, Chung HS, et al. Diagnostic approach to the solitary pulmonary nodule: reappraisal of the traditional clinical parameters for differentiating malignant nodule from benign nodule. *Tuberc Respir Dis* 1996;43:500-18.
- 5) Kim JY, Choi Y, Choi JY, Kim SE, Choe YS, Lee KH, et al. Assessment of quantitative analysis methods for lung F-18-fluorodeoxyglucose PET. *Korean J Nucl Med* 1998;32:332-43.
- 6) Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1983;143:29-36.
- 7) Hanley JA, McNeil BJ. A method of comparing

- the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839-43.
- 8) DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A non-parametric approach. *Biometrics* 1988;44:837-45.
- 9) Yoon SB, Kim B-T, Choi JY, Kim SJ, Choi Y, Choe YS, et al. Role of PET in evaluating indeterminate solitary pulmonary nodule with CT. *Korean J Nucl Med* 1997;31:83-89.
- 10) Lee DS, Yoon SN, Kwark C, Chung J-K, Lee MC, Koh C-S. Inflammatory lung lesions could not be differentiated from malignancy with F-18-FDG PET. *Eur J Nucl Med* 1996;23:1070.
-