

Revisión sistemática y metaanálisis de la eficacia diagnóstica de la PET ^{18}F -FDG en recurrencia tumoral de cánceres de cabeza y cuello

Juan Antonio Pasamontes Pingarrón^a, María Nieves Cabrera Martín^b, Roberto Carlos Delgado Bolton^b, Cristina Fernández Pérez^c, José Luis Carreras Delgado^d y Bartolomé Scola Yurrita^e

^aServicio de Otorrinolaringología. Hospital de Móstoles. Madrid. España.

^bInstituto PET Focuscan. Madrid. España.

^cServicio de Epidemiología y Medicina Preventiva. Hospital Clínico San Carlos. Madrid. España.

^dServicio de Medicina Nuclear. Hospital Clínico San Carlos. Madrid. España.

^eServicio de Otorrinolaringología. Hospital General Universitario Gregorio Marañón. Madrid. España.

Objetivos: En este trabajo se realizó un metaanálisis de la literatura para evaluar la exactitud diagnóstica de la PET ^{18}F -FDG en la sospecha de recurrencia de cánceres de cabeza y cuello.

Material y métodos: Inicialmente se realizó una búsqueda sistemática de la literatura en las fuentes MEDLINE y CANCERLITE hasta mayo de 2007, aplicando unos criterios de cribado a los estudios hallados. Se calcularon los valores de sensibilidad, especificidad, cocientes de probabilidad positivo y negativo y la *odds ratio* diagnóstica. El rendimiento diagnóstico se valoró mediante curvas ROC (*receive-operating-characteristic*).

Resultados: Fueron incluidos 19 artículos en el metaanálisis. La sensibilidad de la PET ^{18}F -FDG fue del 94 %, la especificidad del 80 % y la curva ROC mostró una buena relación entre sensibilidad y especificidad.

Conclusiones: La PET ^{18}F -FDG fue útil en pacientes con sospecha de recurrencia tumoral por cáncer de cabeza y cuello; mostró una sensibilidad alta y una especificidad intermedia-alta.

Palabras clave: ^{18}F -FDG PET. Tumores de cabeza y cuello. Carcinoma de cabeza y cuello. Recurrencia. Metaanálisis.

Systematic Review and Meta-Analysis of Diagnostic Accuracy of ^{18}F -FDG PET in Suspected Recurrent Head and Neck Cancer

Objectives: A meta-analysis of the literature was performed in this article in order to evaluate the diagnostic accuracy of ^{18}F -FDG PET in suspected recurrent head and neck cancer.

Material and methods: A systematic review of the literature was performed using 1987 to 2007 MEDLINE and CANCERLIT databases, applying selection criteria to the studies found. Sensitivity, specificity, positive and negative likelihood ratios, and odds ratio were calculated. The diagnostic accuracy was evaluated with ROC (*receive-operating-characteristics*) curves.

Results: Nineteen articles were included in the meta-analysis. Sensitivity of ^{18}F -FDG PET was 94 %, specificity 80 %, and the summary ROC curve showed a good trade-off between sensitivity and specificity.

Conclusions: ^{18}F -FDG PET was useful in patients with suspected recurrence of head and neck cancer, showing a high sensitivity and intermediate-high specificity.

Key words: ^{18}F -FDG PET. Head and neck tumours. Carcinoma of the head and neck. Recurrence. Meta-analysis.

INTRODUCCIÓN

El diagnóstico de recurrencia tumoral de los cánceres de cabeza y cuello es en muchas ocasiones complejo aun empleando todas las técnicas diagnósticas disponibles a nuestro alcance. Los cambios producidos en los tejidos tumora-

les tras los tratamientos empleados (quimioterapia y/o radioterapia) hacen difícil la valoración mediante las pruebas de imagen utilizadas convencionalmente. La aplicación en España desde 1995 de la tomografía por emisión de positrones (PET) con ^{18}F -FDG como técnica de diagnóstico por imagen que mide metabolismo tumoral supone una alternativa interesante en estas situaciones. Para conocer la utilidad de la PET ^{18}F -FDG en sospecha de recidiva tumoral de pacientes con cáncer de cabeza y cuello, realizamos el presente estudio en el que se llevó a cabo una revisión sistemática de la literatura científica publicada al respecto y a continuación, tras aplicar criterios de selección a los artícu-

Correspondencia: Dr. J.A. Pasamontes Pingarrón.
Santa Engracia, 162, 2.º dcha. 28003 Madrid. España.
Correo electrónico: maria.jantonio@tiscali.es

Recibido el 5-3-2007.

Aceptado para su publicación el 30-11-2007.

los hallados según aspectos de medicina basada en la evidencia (MBE) sobre calidad y relevancia, se practicó un metaanálisis con el fin de obtener resultados estadísticos que permitieran conocer si la PET ^{18}F -FDG muestra utilidad en el diagnóstico de recidiva de tumores de cabeza y cuello.

MATERIAL Y MÉTODOS

El método seguido en esta revisión y metaanálisis, de forma esquemática, fue el siguiente:

- Estrategia de búsqueda bibliográfica.
- Criterios de cribado de estudios (*screening*)¹.
- Criterios de exclusión.
- Criterios de calidad metodológica para pruebas diagnósticas.
- Análisis estadístico de los resultados relevantes de los artículos seleccionados.

Estrategia de búsqueda

Se comenzó con una búsqueda sistemática de la literatura llevada a cabo por dos investigadores, mediante el acceso a las bases de datos médicas MEDLINE y SilverPlatter. Se procedió a realizar la búsqueda de artículos hasta mayo de 2007 y no se limitó la fecha de inicio, ya que se incluyeron todos los artículos publicados al respecto desde sus comienzos. Se emplearon las siguientes palabras de búsqueda²: (*deoxyglucose OR desoxy-glucose OR fluorodeoxyglucose OR fludeoxyglucose OR fdg OR 18fdg OR f-18-dg OR fluoro-2-deoxy-d-glucose OR 2fluoro-2deoxyglucose OR fluoro-d-glucose OR PET OR Positron Emission Tomography*) AND (*head and neck cancer OR recurrent head and neck cancer OR recurrent carcinoma of the head and neck OR recurrent squamous cell carcinoma of the head and neck*). Como resultado de esta búsqueda se encontró un gran número de artículos. En el Hospital de Móstoles (Madrid) se realizó la búsqueda manual de las revistas que contuvieran artículos relacionados con el tema. Además, mediante el servicio de búsqueda de artículos que presta dicha biblioteca, se consiguieron artículos de revistas no disponibles en el centro y que se hallaban en diferentes bibliotecas de universidades y hospitales, tanto de España como del extranjero. Se revisaron también libros de resúmenes de comunicaciones de conferencias y congresos recientes, además de *abstracts* presentados en congresos que hubieran sido publicados como tales en revistas científicas. De igual manera, se accedió, en la Facultad de Medicina de la Universidad Complutense de Madrid y en el Centro de Proceso de Datos, a la literatura y bibliografía no publicadas: tesis doctorales, trabajos de doctorado, etc.

Criterios de cribado de estudios

En primer lugar se descartaron los trabajos que en su título, y a pesar de las palabras clave de búsqueda, no se ajustaban completamente al tema de este estudio. En general, para los criterios de cribado se tuvo en cuenta: *a*) características de los pacientes (por ejemplo, edad, sexo, etc.); *b*) características patológicas; *c*) aspectos tecnológicos;

d) cuestiones metodológicas (por ejemplo, número de pacientes, duración del seguimiento, diseño del estudio, etc.); *e*) resultados medidos, y *f*) tipos de publicaciones.

Los criterios de inclusión establecidos en esta fase preliminar fueron:

1. Estudios publicados en cualquier idioma, tanto artículos completos publicados en revistas de calidad, es decir, citadas por el Science Citation Index (SCI) o incluidas en la base de datos MEDLINE, como *abstracts* presentados en congresos y que hubieran sido publicados como tales.
2. Estudios que presentasen datos primarios.
3. Estudios con al menos 12 sujetos enfermos.
4. Estudios que evaluaran la eficacia y la utilidad de la PET ^{18}F -FDG en el diagnóstico de sospecha de recurrencia tumoral en cánceres de cabeza y cuello.
5. Utilización de fluorodesoxiglucosa (FDG) como radiofármaco.
6. Utilización de cámaras PET dedicadas.
7. Especificación de la prueba de referencia utilizada (anatomía patológica y/o seguimiento clínico).

Los criterios de exclusión en esta fase preliminar fueron:

1. Revisiones.
2. Otros radiofármacos que no fueran FDG.
3. Artículos duplicados o desfasados por estudios subsiguientes (con el mismo nivel de jerarquía y con el mismo propósito) de la misma institución³.
4. Gammacámaras de doble cabezal de coincidencia o *dual-head* (porque presentan menor exactitud diagnóstica y capacidad de detección)⁴.
5. Artículos que presentasen un único caso.
6. Artículos que incluyeran información insuficiente para juzgar sobre: protocolo del procedimiento seguido para obtener las imágenes, si el análisis fue visual o semicuantitativo y resultados estadísticos claros sobre sensibilidad y especificidad.

Tras esta fase previa, los investigadores (JAPP, MNM y JLCD) contrastaron los resultados obtenidos y manifestaron las opiniones individuales en cuanto a la aplicabilidad o no de los criterios de inclusión en relación con los estudios disponibles en la literatura. Una vez se ordenó la literatura, se investigó nuevamente las referencias disponibles en cuanto a su relevancia y se analizó cuidadosamente el documento completo. Entonces, algunos estudios fueron excluidos (exclusión secundaria) porque no se los consideró relevantes o porque su metodología fue deficitaria.

Los *abstracts* a congresos sólo fueron incluidos cuando los objetivos, métodos y resultados de sensibilidad y especificidad del estudio estaban claros. Estos estudios fueron incluidos en el análisis de sensibilidad y especificidad, pero no en el análisis de la calidad metodológica.

En un segundo paso se evaluó de forma crítica la calidad de los trabajos seleccionados.

Criterios de calidad metodológica para pruebas diagnósticas

Para analizarlos se aplicaron a los artículos seleccionados los criterios de calidad metodológica para pruebas

Tabla I. Criterios de calidad metodológica para pruebas de diagnóstico por imagen

| |
|--|
| <p>Criterio 1: descripción del diseño del estudio y criterios de selección de los pacientes</p> <ol style="list-style-type: none"> 1. Diseño del estudio 2. Criterios de selección de pacientes 3. Exclusión de pacientes para estudio final |
| <p>Criterio 2: características de la población</p> <ol style="list-style-type: none"> 1. Media de edad, rango y sexo 2. Enfermedades concomitantes 3. Diabetes mellitus 4. Localización de la recurrencia 5. Características especiales de la institución |
| <p>Criterio 3: indicaciones del paciente que llevan al uso de la PET</p> <ol style="list-style-type: none"> 1. Razones del uso de la PET 2. Razones del uso de la PET y correlación con los resultados 3. Extensión de la enfermedad especificada por paciente |
| <p>Criterio 4: detalles de la tecnología usada durante el estudio y características de interpretación de imágenes</p> <ol style="list-style-type: none"> 1. Técnicas de imagen utilizadas en el estudio y resolución 2. Preparación del paciente 3. Ayuno mínimo de 4 horas 4. Valores de glucemia antes de la inyección de FDG 5. Sondaje urinario (con o sin furosemida e hidratación iv) 6. Corrección de atenuación 7. Fusión de imágenes 8. Explicación de las características generales de los intérpretes 9. Definición de PET-FDG positivo y PET-FDG negativo 10. Análisis visual o semicuantitativo 11. <i>Scans</i> adicionales |
| <p>Criterio 5: confirmación del diagnóstico final</p> <ol style="list-style-type: none"> 1. Confirmación final 2. Asociación entre los hallazgos específicos de la PET y la confirmación final 3. Confirmación histopatológica 4. Confirmación no histopatológica 5. Pacientes perdidos en la confirmación por seguimiento clínico |
| <p>Criterio 6: datos de sensibilidad y especificidad</p> <ol style="list-style-type: none"> 1. Descripción de VP, FP, VN, FN 2. Explicación cualitativa de FP y FN 3. Región específica estudiada por PET-FDG 4. Localización de la recurrencia por PET-FDG 5. Intervalos de confianza 6. Hallazgos erróneos de la PET-FDG 7. Datos expuestos en pacientes y en las lesiones correspondientes |
| <p>Criterio 7: información sobre cambio en el manejo</p> <ol style="list-style-type: none"> 1. PET-FDG cambia directamente el manejo 2. PET-FDG se utiliza para decidir el tratamiento inicial 3. PET-FDG se utiliza para decidir el tratamiento final 4. 2 + 3 se cumplen 5. PET-FDG correcto o incorrecto directamente cambia el manejo del paciente 6. Incremento o disminución de la estadificación por PET-FDG 7. PET-FDG en el algoritmo del manejo del paciente |

Adaptado de Gould, Huebner, Owens y Delgado.

diagnósticos modificados de Huebner, Gould, Owens et al⁵ y Delgado et al⁶. Estos criterios comprenden 7 aspectos: *a*) descripción del diseño del estudio y criterios de selección de los pacientes; *b*) características de la población finalmente estudiada; *c*) indicaciones para la realización de la PET ¹⁸F-FDG; *d*) detalles de los aspectos tecnológicos y de interpretación de las imágenes; *e*) confirmación diagnóstica final; *f*) datos de sensibilidad y especificidad, y *g*) información sobre el cambio en el manejo del paciente.

La tabla I muestra los criterios de calidad metodológica y los diferentes aspectos que incluyen. Se estableció un sistema de puntuaciones que permitiera valorar el cumplimiento de cada criterio y aspecto con 4 puntuaciones: adecuada (A), parcial (P), no referida (N) o no aplicable (N/A), todo ello referido a la exhaustividad con que el artículo presentaba ese aspecto del estudio. Una puntuación de A significa que el aspecto se describe suficientemente y que está de acuerdo con lo que consideramos el ideal. Una puntuación de P significa una descripción parcial del aspecto o un cumplimiento parcial del apartado metodológico. Una puntuación de N se asigna cuando no se hace referencia a dicho aspecto o no se describe en absoluto a pesar de que se considere relevante para los objetivos del estudio. Finalmente, se asignó N/A cuando un estudio no trataba un determinado aspecto de nuestros criterios. Es importante aclarar que los resultados de este análisis no reflejan la fiabilidad de los resultados presentados, sino la cantidad de información disponible y el cumplimiento de unas guías metodológicas que tomamos como el estándar ideal.

Al menos 2 investigadores experimentados en PET ¹⁸F-FDG (JAPP, MNM), de forma independiente, realizaron el análisis del cumplimiento de los criterios de calidad metodológica. En los casos en que no había concordancia del 100 % entre los investigadores, un tercer investigador (JLCD) llevó a cabo un análisis para llegar a un resultado consensuado⁷. El análisis de los estudios por parte de los investigadores no se realizó de forma ciega en cuanto a autor, año, revista, resultados o nombre de la institución.

En los estudios finalmente seleccionados para el metaanálisis se establecieron, en función de su diseño, 2 tipos de estudios o subgrupos con el fin de valorar factores de heterogeneidad. En el primer subgrupo se clasificaron, por un lado, los trabajos en que se solicitó la PET ¹⁸F-FDG ante la sospecha de recurrencia, basada ésta en los hallazgos de la exploración física y/o métodos de imagen convencionales (tipo A), en comparación con los artículos en que la PET ¹⁸F-FDG se realizó de manera sistemática en el seguimiento de pacientes tratados de carcinomas de cabeza y cuello, pero sin sospecha evidente de recidiva (tipo B). El segundo subgrupo se estableció en función del análisis de interpretación de la PET ¹⁸F-FDG, considerando si éste fue únicamente visual (tipo I) o visual y semicuantitativo con medición del SUV (*standardized uptake value*) (tipo II).

Análisis estadístico

En los estudios de evaluación de pruebas diagnósticas, se resume cada estudio por los índices que describen la validez de la prueba. Éstos son sensibilidad, especificidad, cocientes de probabilidad positivo y negativo y *odds ratio*

Tabla II. Datos poblacionales extraídos de los artículos seleccionados para el metaanálisis

| Autores | Total de pacientes (n) | Pacientes estudiados (n) | PET (n) | Rango de edad | Media de edad | Varones (%) | Mujeres (%) |
|--------------------------------|------------------------|--------------------------|---------|---------------|---------------|-------------|-------------|
| Rege et al ⁹ | 60 | 17 | 17 | 23-75 | 58 | 79 | 21 |
| Anzai et al ¹⁰ | 12 | 12 | 12 | 48-80 | 63 | 83 | 17 |
| Greven et al ¹¹ | 31 | 31 | 31 | | | | |
| Fischbein et al ¹² | 44 | 35 | 36 | 30-82 | 59 | | |
| Chia-Hung et al ¹³ | 36 | 36 | 36 | 18-67 | 44,7 | 66,6 | 33,3 |
| Farber et al ¹⁴ | 28 | 28 | 28 | | | | |
| Hanasono et al ¹⁵ | 84 | 34 | 34 | | | | |
| Nowak et al ¹⁶ | 71 | 30 | 26 | | 61 ± 10 | 69 | 31 |
| Lowe et al ¹⁷ | 44 | 30 | 30 | | | | |
| Lonneux et al ¹⁸ | 44 | 44 | 44 | 39-80 | 57,5 ± 10,7 | 88,6 | 11,4 |
| Di Martino et al ¹⁹ | 50 | 13 | 13 | | 60,5 | 2 | 28 |
| Li et al ²⁰ | 43 | 43 | 43 | | | | |
| Kresnik et al ²¹ | 54 | 15 | 15 | | 61,3 | 76 | 24 |
| Wong et al ²² | 143 | 143 | 181 | 13-91 | 59 | 67,8 | 32,2 |
| Tsai et al ²³ | 28 | 28 | 28 | | 48,9 | 78,5 | 21,5 |
| Pasamontes et al ²⁴ | 30 | 30 | 34 | 15-76 | 58 | 73,3 | 26,7 |
| Álvarez et al ²⁵ | 60 | 60 | 60 | 37-83 | 55 | 80 | 20 |
| Cermik et al ²⁶ | 50 | 37 | 48 | 28-89 | 58 | 58,3 | 41,7 |
| Total | 912 | 666 | 716 | 13-91 | 56,8 | 74,4 | 25,6 |

diagnóstica. Los intervalos de confianza de la sensibilidad y la especificidad se calcularon mediante el método exacto para las proporciones binomiales. La homogeneidad de los cocientes de probabilidad positivo y negativo y *odds ratio* diagnósticos se contrastó con la prueba Q de Cochran. Los cocientes de probabilidad y la *odds ratio* diagnóstica se agruparon por el método de DerSimonian Laird (modelo de efectos aleatorios) para incorporar la variación entre estudios.

Para valorar el rendimiento del test diagnóstico basado en los resultados de los estudios analizados, desarrollamos curvas ROC resumen (*summary ROC curve*), de la forma descrita por Moses et al⁸. Para explorar fuentes de heterogeneidad en los estudios (metarregresión) se empleó el método de Moses-Shapiro-Littenberg añadiendo covariables al modelo.

RESULTADOS

Empleando la estrategia de búsqueda anteriormente descrita, se encontró un total de 1.729 artículos, comprendidos desde octubre de 1967 hasta mayo de 2007.

Sin embargo, la primera publicación versada en tomografía por emisión de positrones aparece en mayo de 1987. Por ello, eliminamos los trabajos publicados con anterioridad a dicha fecha, que suponían 47 títulos. De esta manera, el número total de referencias bibliográficas incluidas en la revisión fue finalmente 1.682.

A continuación se descartaron los artículos que no cumplían los criterios mencionados en el anterior apartado, por lo que quedaron finalmente 18 artículos⁹⁻²⁶ para su evaluación definitiva. Únicamente un *abstract* de una comunicación a congreso²⁷ cumplió todos los criterios de inclusión.

Todos estos 18 artículos correspondían a la búsqueda a través de MEDLINE, y en todos los casos se referían a pacientes que fueron diagnosticados y tratados de un carcinoma de cabeza y cuello en los que, ante sospecha de recidiva tumoral, se realizó tomografía por emisión de positrones.

Características de la población estudiada (tabla II)

Entre todos los artículos seleccionados se registró un total de 912 pacientes, de éstos, finalmente, se estudió a 666 en el metaanálisis. A estos 666 pacientes se les realizaron 716 estudios de PET.

El intervalo de edad de la población total estudiada era 13-91 años, con una media de 56,8. El 74,4 % eran varones y el 25,6 %, mujeres.

Valoración de la calidad metodológica

En la valoración de la calidad metodológica de los estudios no se incluyó el *abstract* a congreso²⁶, ya que al no disponer del artículo completo, por no tratarse de una publicación, sus resultados sólo se valoraron en el análisis estadístico. Para los artículos seleccionados se calcularon los porcentajes de puntuaciones A, P y N asignadas a cada artículo. El intervalo de puntuaciones de cumplimiento A

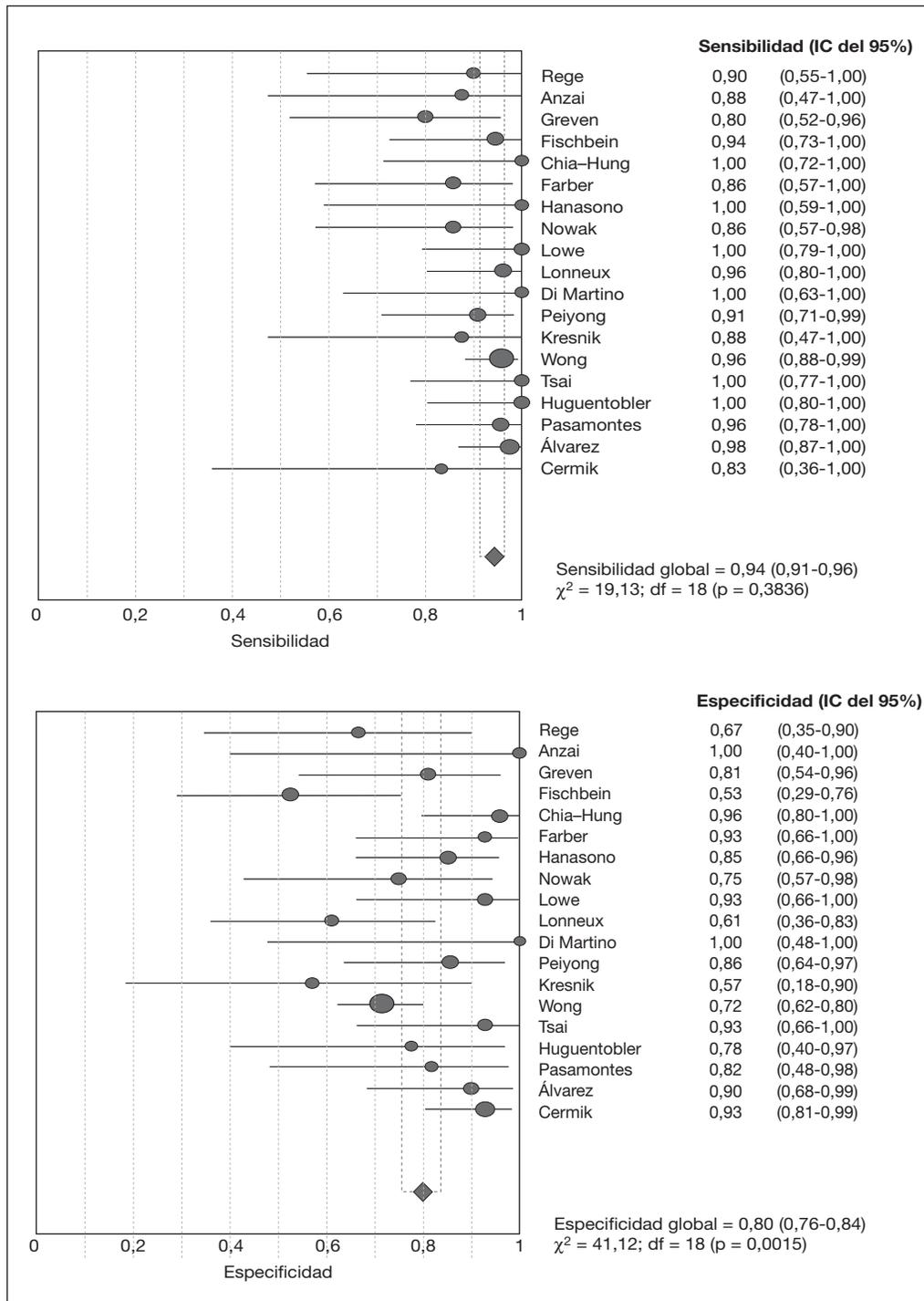


Figura 1. Resultados de sensibilidad y especificidad en los estudios seleccionados para el metaanálisis de la literatura.

en cada artículo fue del 45,2-89,3% con una media del 63,6%. El porcentaje medio de puntuaciones P en cada artículo fue del 16,8%, con un intervalo del 3,4-39,3%. Las puntuaciones de cumplimiento N en cada artículo presentaron una media del 17,8%, con un intervalo del 3,6-29,4%. El cumplimiento de las guías de calidad metodológica para cada artículo se consideró alto cuando el porcentaje de puntuaciones A de cumplimiento era mayor del 70%, aceptable cuando estaba en el 50-70% y bajo cuando era menor del 50%. Cuando se obtuvieron los resultados de cada uno

de los criterios, en los 18 artículos elegidos, se encontró que los criterios 1, 3 y 5 presentaron un cumplimiento general alto (> 70%). El criterio 1 (descripción del diseño del estudio y criterios de selección de muestra) presentó el mayor cumplimiento general (91,3%), el criterio 3 (indicaciones del paciente que llevan al uso de la PET ¹⁸F-FDG) presentó un cumplimiento del 86,05% y el criterio 5 (confirmación del diagnóstico final), un 88,8%. En el resto de los criterios, el porcentaje de cumplimiento oscilaba entre el 30,5 y el 72,8%. Destaca que el criterio al que se hizo menos referen-

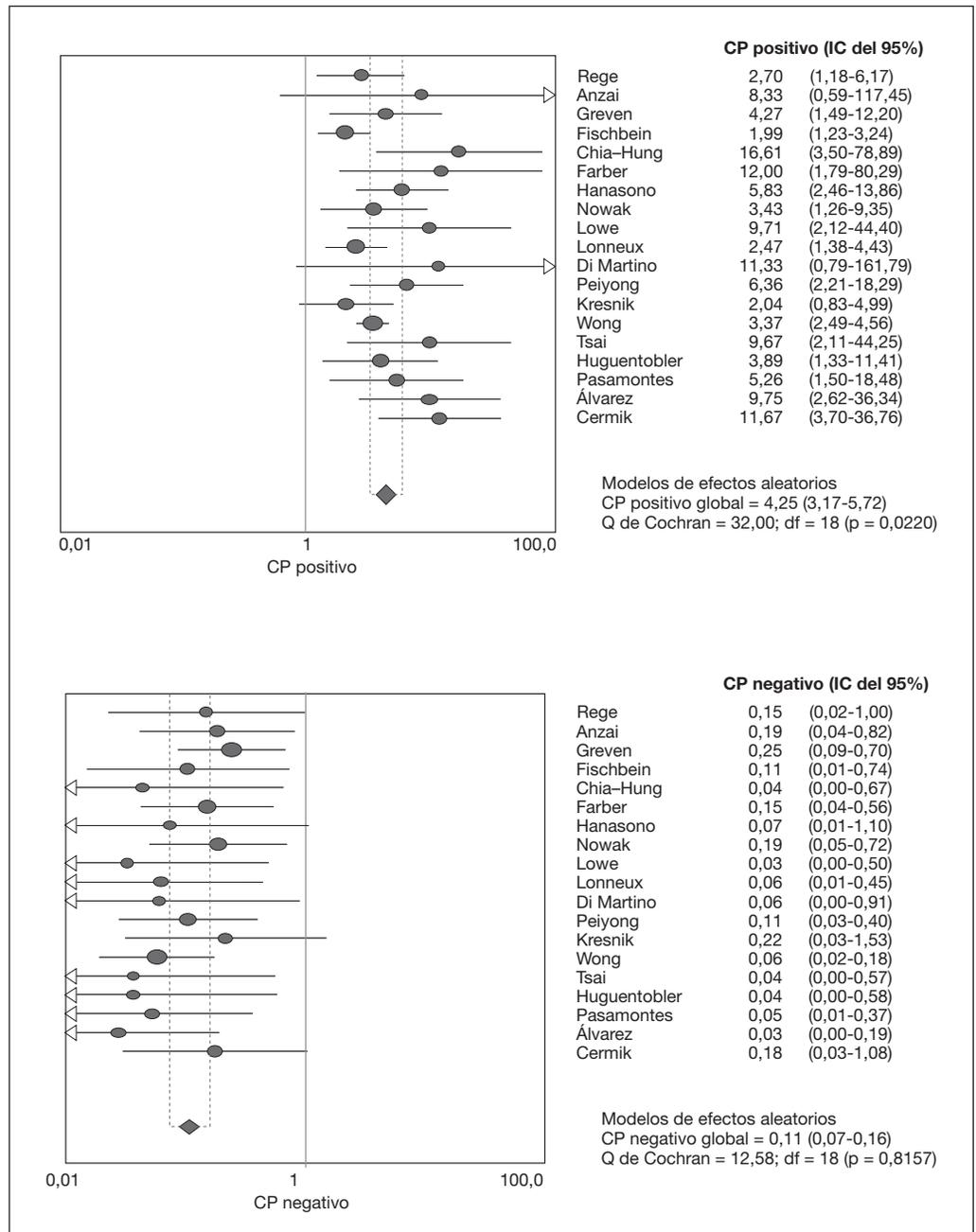


Figura 2. Resultados del CP positivo y el CP negativo de los estudios del metaanálisis.

cia en todos los artículos fue el 2 (características de la población finalmente estudiada), con un índice de N del 56,9%. Por otro lado, el criterio 7, referente a la información sobre el cambio en el manejo del paciente derivado del uso de la PET ¹⁸F-FDG, fue el que contempló el mayor número de apartados con un N/A en el 72,2%.

Valores obtenidos en el metaanálisis

La sensibilidad global de la PET ¹⁸F-FDG en los 19 estudios incluidos en el metaanálisis fue del 94% (intervalo de confianza [IC] del 95%, 91-96) con un rango de 80 a 100. La especificidad global fue del 80% (IC del 95%, 76-84), con un rango de 53 a 100. Al valorar la heterogeneidad, tanto para la sensibilidad como para la especificidad de los dife-

rentes estudios, la prueba de la χ^2 mostró una p = 0,383 en la sensibilidad y una p = 0,001 para la especificidad (fig. 1). Por tanto, si se considera que hay heterogeneidad cuando p < 0,05, esto implica que, en lo que respecta a la sensibilidad, había homogeneidad global entre estudios y, por tanto, el resultado pudo ser empleado para obtener una estimación global de la sensibilidad. De manera diferente ocurrió en la especificidad, donde obtuvimos una p < 0,05 y, por ello, mostró que había una heterogeneidad entre estudios. Se establecieron unos factores o subgrupos (tipo I/II y A/B), expuestos en el apartado de "Material y métodos", dentro de los artículos analizados, para ver si éstos podían justificar dicha heterogeneidad. Así pues, al rea-

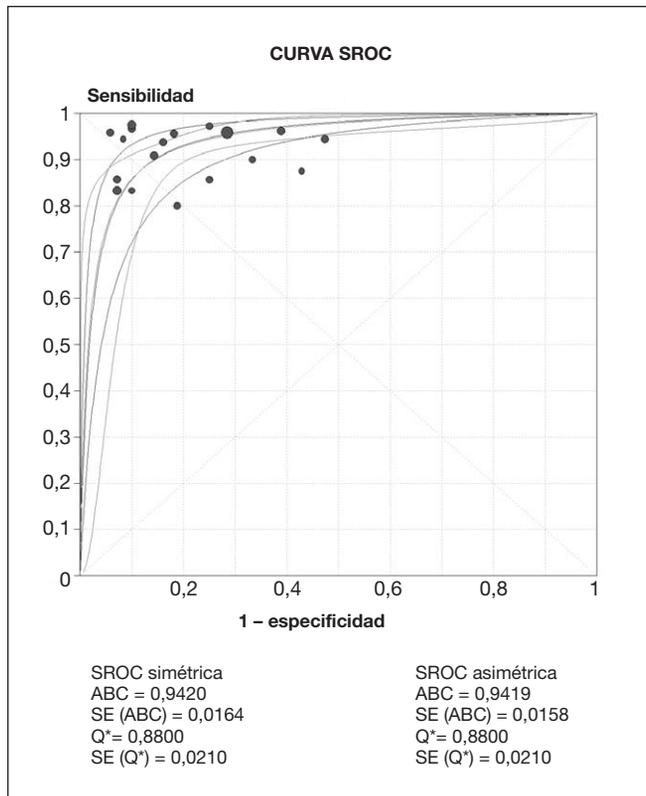


Figura 3. Curva ROC resumen. El umbral empleado en la mayoría de los estudios favorecía la sensibilidad frente a la especificidad, ya que la mayoría de los estudios están situados en la parte superior del diagrama.

lizar la metarregresión (ponderado por el inverso de la variancia), se comprobó que dichos subgrupos no mostraban una diferencia estadísticamente significativa, por lo que no influían en la falta de homogeneidad. Se obtuvo para el subgrupo I/II un valor $p = 0,33$ y de $0,87$ para el subgrupo A/B. En ambos casos, al ser la $p > 0,05$ no mostró significación. Lo mismo ocurrió con el análisis de la calidad metodológica de los estudios, donde resultó que ésta no influía en la heterogeneidad ($p = 0,30$). Se calculó la *odds ratio* diagnóstica como medida del rendimiento de un test diagnóstico. Se obtuvo un valor global de $51,68$ (IC del 95 %, $30,34-88,05$). El CP positivo (o LR positivo) fue $4,25$ (IC del 95 %, $3,17-5,72$), e indica que un resultado positivo de la PET $^{18}\text{F-FDG}$ induce cambios pequeños en la probabilidad pretest (o preprueba). Sin embargo, el CP negativo (o LR negativo) fue $0,11$ (IC del 95 %, $0,07-0,16$), e indica que cuando la PET $^{18}\text{F-FDG}$ fue negativa produjo cambios moderados en la probabilidad pretest (fig. 2). La curva ROC resumen indicó una buena relación entre la sensibilidad y la especificidad; es decir, la variación o la compensación entre sensibilidad y especificidad a medida que se modificaba el umbral de la prueba fue adecuada (fig. 3).

DISCUSIÓN

La homogeneidad observada en la sensibilidad —y en sus IC del 95 %— de los diferentes tipos de estudios hizo posible obtener un efecto global o conjunto²⁸. Sin embargo, en la especificidad sí apareció una heterogeneidad entre estudios ($p = 0,001$), que podría estar justificada por una elevada probabilidad pretest de recurrencia tumoral entre los sujetos estudiados en los artículos seleccionados.

El mejor resumen de los resultados de los estudios es una curva ROC. La curva ROC obtenida en el metaanálisis de la literatura mostró una buena relación entre sensibilidad y especificidad. Un dato estadístico útil cuando se agrupan estudios mediante la curva ROC es el área bajo la curva (ABC), que resume el rendimiento diagnóstico en un solo número. De manera que las pruebas perfectas tienen un ABC cercano a 1 y las inútiles, cercano a 0,5. El ABC en nuestra curva ROC fue de $0,94$, con lo cual vemos que está muy próxima a 1, por lo que se podría establecer el alto rendimiento diagnóstico de la PET $^{18}\text{F-FDG}$ en el diagnóstico de sospecha de recurrencia tumoral en cánceres de cabeza y cuello. La *odds ratio* diagnóstica mostró que existía una correlación positiva significativa entre la recurrencia tumoral y un resultado positivo de la PET $^{18}\text{F-FDG}$, frente a un resultado negativo, por lo que su contribución al diagnóstico fue significativa. El CP positivo (LR positivo) indicó cambios pequeños de la probabilidad pretest a la postest, mientras que el CP negativo (LR negativo) indicó cambios moderados. Los resultados obtenidos apuntan a que la PET $^{18}\text{F-FDG}$ podría ser útil en pacientes con sospecha de recurrencia tumoral por cánceres de cabeza y cuello. La PET $^{18}\text{F-FDG}$ presentó una especificidad intermedia-alta y una sensibilidad alta, lo que indica que hay pocos resultados falsos negativos, un aspecto importante en el manejo de pacientes oncológicos, de modo que podría apuntar a su utilidad en las fases iniciales del proceso diagnóstico.

El metaanálisis de los estudios publicados sobre la PET $^{18}\text{F-FDG}$, en pacientes con sospecha de recurrencia tumoral de cabeza y cuello, proporcionó las siguientes conclusiones:

- En primer lugar, la calidad metodológica fue alta, y las diferencias en la calidad de los estudios no se correlacionaron con diferencias en sus resultados.

- En segundo lugar, hubo homogeneidad, entre los diferentes tipos de estudios detectados, en sensibilidad, cociente de probabilidad positivo, cociente de probabilidad negativo y *odds ratio* diagnóstica, por lo que los resultados pudieron ser puestos en común para obtener una estimación global o conjunta.

- En tercer lugar, la especificidad sí presentó heterogeneidad entre estudios, debida a una alta probabilidad pretest de enfermedad recurrente.

- En cuarto y último lugar, el rendimiento diagnóstico de la PET $^{18}\text{F-FDG}$ en el diagnóstico de sospecha de enfermedad recurrente tumoral de cabeza y cuello, mostrado por la curva ROC resumen, fue elevado, al mostrar un ABC muy próximo a 1.

BIBLIOGRAFÍA

1. Flynn K, Adams E, Anderson D. Positron emission tomography: Descriptive analysis of experience with PET in VA, systematic reviews; FDG-PET as diagnostic test in cancer and Alzheimer's disease. Office of Research and Development. Health Services Research and Development Service Management Decision and Research Center. VA Medical Center. MDRC Technology Assessment Program-PET Report n.º 10, July 1999.
2. Mijnhout GS, Hooft L, Van Tulder MW, Devillé WLJM, Teule GJJ, Hoekstra OS. How to perform a comprehensive search for FDG-PET literature. *Eur J Nucl Med.* 2000;27:91-7.
3. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med.* 1987;107:224-33.
4. Abdel-Dayem HM, Luo JQ, Sadek S. Multifunctional gamma camera coincidence imaging; Current status and evaluation of clinical results. En: Freeman LM, editor. *Nuclear Medicine Annual 2000.* 1.ª ed. Philadelphia: Lippincott Williams & Wilkins; 2000. p. 1-52.
5. Owens DK, Holodniy M, Garber AM, Scott J, Sonnad S, Moses L, et al. Polymerase chain reaction for the diagnosis of HIV infection in adults. A meta-analysis with recommendations for clinical practice and study design. *Ann Intern Med.* 1996;124:803-15.
6. Delgado-Bolton RC, Fernández-Pérez C, González-Maté A, Carreras JL. Meta-Analysis of the performance of ¹⁸F-FDG PET in primary tumor detection in unknown primary tumors. *J Nucl Med.* 2003;44:1301-14.
7. Littenberg B. Aminophylline treatment in severe, acute asthma. A meta-analysis. *JAMA.* 1998;259:1678-84.
8. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med.* 1993;12:1293-316.
9. Rege S, Maass A, Chaiken I, Hoh CK, Choi Y, Lufkin R, et al. Use of positron emission tomography with fluorodeoxyglucose in patients with extracranial head and neck cancers. *Cancer.* 1994;73:3047-58.
10. Anzai Y, Carroll WR, Quint DJ, Bradford CR, Minoshima S, Wolf GT, et al. Recurrence of head and neck cancer after surgery or irradiation: prospective comparison of 2-deoxy-2-[¹⁸F]fluoro-D-glucose PET and MR imaging diagnoses. *Radiology.* 1996;200:135-41.
11. Greven KM, Williams DW, Keyes JW, McQuirt WF, Watson NE Jr, Case LD. Can positron emission tomography distinguish tumor recurrence from irradiation sequelae in patients treated for larynx cancer? *Cancer J.* 1997;3:353-7.
12. Fischbein NJ, AAssar OA, Caputo GR, Kaplan MJ, Singer MI, Price DC, et al. Clinical utility of positron emission tomography with ¹⁸F-Fluorodeoxyglucose in detecting residual/recurrent squamous cell carcinoma of the head and neck. *Am J Neuroradiol.* 1998;19:1189-96.
13. Chia-Hung K, Sheng-Ping C, Poon-Ung C, Ruoh-Fang Y, Tzu-Chen Y. Detection of recurrent or persistent nasopharyngeal carcinomas after radiotherapy with ¹⁸F-Fluoro-2-Deoxyglucose positron emission tomography and comparison with computed tomography. *J Clin Oncol.* 1998;16:3550-5.
14. Farber LA, Bernard F, Machtay M, Smith RJ, Weber RS, Weinstein GS, et al. Detection of recurrent head and neck squamous cell carcinomas after radiation therapy with ²⁻¹⁸F-Fluoro-2-Deoxy-D-Glucose positron emission tomography. *Laryngoscope.* 1999;109:970-5.
15. Hanasono MM, Kunda LD, Segall GM, Ku GH, Terris DJ. Uses and limitations of FDG positron emission tomography in patients with head and neck cancer. *Laryngoscope.* 1999;109:880-5.
16. Nowak B, Di Martino E, Jänicke S, Cremerius U, Adam G, Zimny M, et al. Diagnostic evaluation of malignant head and neck cancer by F-18-FDG PET compared to CT/MRI. *Nuklearmedizin.* 1999;38:312-8.
17. Lowe VJ, Dunphy FR, Varvares M, et al. Evaluation of chemotherapy response in patients with advanced head and neck cancer using PET-FDG. *Head Neck.* 1997;19:666-74.
18. Lonneux M, Lawson G, Ide C, Bausart R, Remacle M, Pauwels S. Positron emission tomography with fluorodeoxyglucose for suspected head and neck tumor recurrence in the symptomatic patient. *Laryngoscope.* 2000;110:1493-7.
19. Di Martino E, Nowak B, Hassan HA, Hausmann R, Adam G, Buell U, et al. Diagnosis and staging of head and neck cancer. A comparison of modern imaging modalities (positron emission tomography, computed tomography, color-coded duplex sonography) with panendoscopic and histopathologic findings. *Arch Otolaryngol Head Neck Surg.* 2000;126:1457-61.
20. Li P, Zhuang H, Mozley PD, Denittis A, Yeh D, Machtay M, et al. Evaluation of recurrent squamous cell carcinoma of the head and neck. With FDG positron emission tomography. *Clin Nucl Med.* 2001;26:131-5.
21. Kresnik E, Mikosch P, Gallowitsch HJ, Kogler D, Wiesser S, Heinisch M, et al. Evaluation of head and neck cancer with ¹⁸F-FDG PET: a comparison with conventional methods. *Eur J Nucl Med.* 2001;28:816-21.
22. Wong Rj, Lin DT, Schöder H, Patel SG, Gonen M, Wolden S, et al. Diagnostic and prognostic value of [¹⁸F] Ffluorodeoxyglucose positron emission tomography for recurrent head and neck squamous cell carcinoma. *J Clin Oncol.* 2002;20:4199-208.
23. Tsai MH, Shiau YC, Kao CH, Shen YY, Lin CC, Lee CC. Detection of recurrent nasopharyngeal carcinomas with positron emission tomography using ¹⁸-fluoro-2-deoxyglucose in patients with indeterminate magnetic resonance imaging findings after radiotherapy. *J Cancer Res Clin Oncol.* 2002;128:279-82.
24. Pasamontes Pingarrón JA, Cabrera Martín MN, Carreras Delgado JL, Scola Yurrita B, Calvo Manuel F, Delgado Bolton RC. Comparación entre la PET ¹⁸F-FDG y las técnicas de imagen convencionales (TAC y RMN) en el diagnóstico de sospecha de recurrencia de tumores de cabeza y cuello. *Acta Otorrinolaringol Esp.* 2006; 57:441-5.
25. Álvarez Pérez RM, Borrego Dorado I, Ruiz Franco-Baux JV, Vázquez Albertino RJ. Evaluación de la eficacia y el impacto clínico de la tomografía de emisión de positrones con ¹⁸F-FDG en pacientes con sospecha de recurrencias locales y metástasis de carcinomas de cabeza y cuello. *Rev Esp Med Nucl.* 2007;26:30-9.
26. Cermik TF, Mavi A, Acikgoz G, Houseni M, Dadparvar S, Alavi A. FDG PET in Detecting Primary and Recurrent Malignant Salivary Gland Tumors. *Clin Nucl Med.* 2007;32:286-91.
27. Huguentobler A, Périé S, Montravers F, et al. A prospective study of the clinical impact of FDG-PET to detect recurrent head and neck cancer during the year following initial curative therapy. *Eur J Nucl Med Mol Imaging.* 2002;225.
28. Lau J, Ionnidis JPA, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med.* 1997;127:820-6.