Colorectal liver metastases: CT, MR imaging, and PET for diagnosis. Meta-analysis

CRD summary
The authors concluded that 18-fluorodeoxyglucose positron emission tomography had the highest sensitivity for detecting patients with liver metastases, but not for detecting individual lesions. Contrast-enhanced magnetic resonance imaging and helical computed tomography using more than 45 g of iodine had the highest sensitivities for determining the number of metastases. This was a well-conducted review and the conclusions are likely to be reliable.

Authors' objectives
To use meta-analyses to estimate the sensitivities of computed tomography (CT), magnetic resonance imaging (MRI) and fluorine-18 fluorodeoxyglucose positron emission tomography (FDG PET) for the detection of colorectal liver metastases, on a per-patient and per-lesion basis.

Searching
MEDLINE, EMBASE, CINAHL, SUMSearch, Web of Science, Cancerlit and the Cochrane Database of Systematic Reviews were searched from January 1990 to December 2003; the search terms were reported. In addition, the reference lists of all retrieved articles were checked. Articles reported in English, German or French were included.

Study selection
Study designs of evaluations included in the review
No inclusion criteria for study design were specified. The included studies were diagnostic accuracy studies.

Specific interventions included in the review
Studies using CT, MRI or FDG PET to identify and characterise colorectal liver metastases were eligible for inclusion. The included studies of CT were of both helical and non-helical CT, while MRI studies used either 1.0 or 1.5 T machines. Further details (e.g. slice thickness, contrast agents) were reported in the text. All 61 included studies presented data on the identification of lesions. Twenty-eight studies characterised lesions as benign or malignant. Three studies provided data on detailed sub-categorisation.

Reference standard test against which the new test was compared
Studies using histopathology (surgery, biopsy or autopsy), intra-operative observation (e.g. manual palpation or intra-operative ultrasound) and/or follow-up ultrasound as the reference standard of diagnosis were eligible for inclusion.

Participants included in the review
Studies of participants with colorectal cancer were eligible for inclusion. The mean age of the study participants was 61 years (range: 12 to 93). Where described, the overall gender distribution was approximately 60% male.

Outcomes assessed in the review
The included studies were required to present sufficient data for the calculation of true-positive and false-negative values, on a per-patient or per-lesion basis, for one or more of the specified imaging techniques. Studies were excluded if the results were presented for a combination of imaging modalities and the performance data for the individual modalities could not be distinguished.

How were decisions on the relevance of primary studies made?
Since one reviewer checked all retrieved articles for inclusion and three reviewers checked a subset of articles (either all CT, all MRI, or all FDG PET studies), each article was checked by two independent reviewers. Any disagreements were resolved by consensus.
Assessment of study quality
The QUADAS tool was used to assess the methodological validity of the included studies. The QUADAS tool was developed and validated specifically for the quality assessment of diagnostic accuracy studies in systematic reviews. Items assessed included: the spectrum and selection of patients; adequacy of the description of index tests and reference standard: time between index test and reference standard (the potential for the disease state to change between tests); application of the reference standard; and blinding of test interpretation.

The quality assessment was conducted using a standardised form. Since one reviewer assessed all included studies and three reviewers assessed a subset of articles (either all CT, all MRI, or all FDG PET studies), each article was assessed by two independent reviewers. Any disagreements were resolved by consultation with a fifth reviewer.

Data extraction
The data were extracted using a standardised form. Since one reviewer extracted data from all included studies and three reviewers extracted data from a subset of articles (all CT, all MRI, or all FDG PET studies), each article were data extracted by two independent reviewers. Any disagreements were resolved by consultation with a fifth reviewer.

Data were extracted on: study characteristics; participant characteristics (including tumour stage and number with liver metastases); details of the index test and reference standard; and results (numbers of true-positive, false-negative and false-positive results on a per-lesion basis, and numbers of true-positive, false-negative, false-positive and true-negative results on a per-patient basis). Data sets for multiple readers, multiple observations per reader and multiple techniques were treated as being separate.

Methods of synthesis
How were the studies combined?
Data were analysed separately for non-helical CT, helical CT, 1.0 T MRI, 1.5 T MRI and FDG PET. The sensitivity of each imaging technique (with 95% confidence interval, CI) was estimated as a proportion of patients with liver metastases (per patient data) or as a proportion of liver metastases (per lesion data) detected. Sensitivity estimates were obtained using a random-effects linear regression model with a mixed-effects approach.

How were differences between studies investigated?
Univariate regression analyses were used to examine the dependence of sensitivity values upon a number of variables, including year of publication, sample size and individual QUADAS criteria. Variables were considered explanatory if the regression coefficient for the variables was significant (p<0.05). A multivariate regression model, including significant variables from the univariate analyses, was then developed using a backward stepwise algorithm; variables with a significance of p<0.1 were retained in the model. This regression model was then used to compare sensitivities between imaging modalities by including a variable for imaging modality; a p-value of less than 0.05 for this factor was considered to indicate a significant difference.

For data on helical CT, subgroup analyses were used to compare the sensitivities given by different section thicknesses, amounts of iodine administered in the contrast agent and numbers of phases. For data on 1.5 T MRI, subgroup analyses were used to compare the sensitivities given by different enhancement techniques (gadolinium or superparamagnetic iron oxide (SPIO)) or no enhancement techniques. Data for helical CT, non-enhanced 1.5 T MRI, gadolinium-enhanced 1.5 T MRI and SPIO-enhanced 1.5 T MRI were further subgrouped by lesion size (less than 1 cm versus 1 cm or more). All subgroup analyses were for per lesion data only.

Results of the review
Sixty-one studies with a total of 3,187 participants were included in the review. Twenty-eight articles provided 58 data sets on non-helical CT, 15 articles provided 53 data sets on helical CT, 5 articles provided 34 data sets on 1.0 T MRI, 12 articles provided 102 data sets on 1.5 T MRI and 21 articles provided 26 data sets on FDG PET.

Study quality.
Most studies (67.2%) were subject to the possibility of disease progression bias (clinically significant delay between the index test and reference standard). The description of the execution of the index test and reference standard in the included studies was generally poor. Blinding of interpreters to the results of other tests was rare (91.2% of studies interpreted the reference standard with knowledge of the index test result), as was the reporting of uninterpretable and/or indeterminate results and accounting for withdrawals.

Sensitivity estimates (per patient).

The sensitivity estimates for non-helical CT, helical CT, 1.5 T MRI and FDG PET were 60.2% (95% CI: 55.7, 64.6), 64.7% (95% CI: 30.4, 88.5), 75.8% (95% CI: 55.9, 88.6) and 94.6 (95% CI: 92.5, 96.1), respectively. FDG PET had significantly higher sensitivity than all other modalities. For non-helical CT and FDG PET, factors relating to the interpretation of the index test and/or reference standard were significant predictors of reported diagnostic performance; full results were tabulated in the article.

Sensitivity estimates (per lesion).

The sensitivity estimates for non-helical CT, helical CT, 1.0 T MRI, 1.5 T MRI and FDG PET were 52.3%, 63.8%, 66.1%, 64.4% and 75.9%, respectively. Non-helical CT had the lowest sensitivity estimate; other imaging modalities were comparable. For non-helical CT, helical CT, 1.0 T MRI and 1.5 T MRI, factors related to the reporting and interpretation of the index test and/or reference standard and to the participant spectrum were significant predictors of reported diagnostic performance; full results were tabulated in the article.

Subgroup analyses (per lesion). Subgroup analyses for aspects of helical CT technique found no significant differences in diagnostic performance. Sensitivity estimates for gadolinium-enhanced MRI (78.2%) and SPIO-enhanced MRI (73.2%) were significantly higher than those for non-enhanced MRI (59.8%) or helical CT with 45 g or less of iodine (61.4%).

Sensitivity estimates for non-helical CT, helical CT, non-enhanced 1.5 T MRI, gadolinium-enhanced 1.5 T MRI and SPIO-enhanced 1.5 T MRI for lesions smaller than 1 cm were 25.3% (95% CI: 15.9, 37.6), 23.1% (95% CI: 7.0, 54.7), 12.6% (95% CI: 8.0, 17.5), 11.6% (95% CI: 9.5, 14.2) and 29.3% (95% CI: 18.2, 43.6), respectively. There were no significant differences between imaging modalities. Sensitivity estimates for non-helical CT, helical CT, non-enhanced 1.5 T MRI, gadolinium-enhanced 1.5 T MRI and SPIO-enhanced 1.5 T MRI for lesions of 1 cm or larger were 74.3% (95% CI: 66.5, 80.9), 73.5% (95% CI: 62.2, 82.4), 65.7% (95% CI: 56.4, 73.9), 68.8% (95% CI: 61.9, 75.0) and 90.2% (95% CI: 87.5, 92.4), respectively. The sensitivity estimate for SPIO-enhanced MRI was significantly higher.

Authors’ conclusions

FDG PET had a higher sensitivity than other imaging modalities for the detection of colorectal liver metastases on a per-patient basis, but not on a per-lesion basis. Contrast-enhanced MRI techniques were more sensitive than non-enhanced MRI or helical CT using 45 g or less of iodine.

CRD commentary

This was a rigorously conducted and clearly reported systematic review that addressed a clearly stated question, which was defined by appropriate inclusion criteria. An extensive literature search was conducted, though the language restrictions applied might have resulted in the loss of some relevant data. Appropriate methods were used to avoid the introduction of error and bias during the review process and these were clearly reported.

The reporting of detailed characteristics of the included studies in the paper was limited. However, further information was reported as being available from the authors on request. A large number of data sets were included in the review, and key aspects of the included studies were discussed in the text and included as variables in the regression modelling. A quality assessment tool appropriate to diagnostic accuracy studies was used to assess the methodological validity of the included studies, and the results of this assessment were incorporated in the analyses; the influence of elements of methodological quality of the included studies upon pooled estimates of diagnostic performance was investigated. The regression model used was generally considered to be the most appropriate for this type of analysis and its application was clearly described. The authors’ conclusions follow from the data presented and are likely to be
reliable.

**Implications of the review for practice and research**

Practice: The authors stated that the choice between contrast enhanced MRI and helical CT using more than 45 g of iodine should depend upon availability and expertise, as well as diagnostic performance data. They further stated that the currently limited role of FDG PET is likely to restrict its use to that of an additional modality for the detection of extra-hepatic disease.

Research: The authors stated that future diagnostic accuracy studies should follow the reporting guidelines given by the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) initiative.

**Funding**

Dutch Order of Medical Specialists.

**Bibliographic details**


**PubMedID**

16100087

**DOI**

10.1148/radiol.2371042060

**Original Paper URL**

http://radiology.rsna.org/

**Indexing Status**

Subject indexing assigned by NLM

**MeSH**

Colorectal Neoplasms /pathology; Fluorodeoxyglucose F18; Gadolinium; Humans; Liver Neoplasms /diagnosis /pathology /secondary; Magnetic Resonance Imaging; Positron-Emission Tomography; Sensitivity and Specificity; Tomography, Spiral Computed; Tomography, X-Ray Computed

**AccessionNumber**

12005001482

**Date bibliographic record published**

30/06/2007

**Date abstract record published**

30/06/2007

**Record Status**

This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.