Retrospective, blinded ROC analysis of diagnosing MRI, PET-CT and CT in patients with colorectal liver metastases: diagnostic performance of three different qualified reviewers

Poster No.: C-2379
Congress: ECR 2013
Type: Scientific Exhibit
Authors: M. Albrecht¹, T. J. Vogl², C. Müller¹, J. L. Wichmann¹, S. Sakthibalan³, W. Bechstein¹, T. Schreckenbach¹, H. Ackermann¹, S. Zangos¹; ¹Frankfurt am Main/DE, ²Frankfurt a. Main/DE, ³London/UK
Keywords: Liver, CT, PET-CT, MR, Staging, Surgery, Localisation, Metastases
DOI: 10.1594/ecr2013/C-2379

Any information contained in this pdf file is automatically generated from digital material submitted to EPOS by third parties in the form of scientific presentations. References to any names, marks, products, or services of third parties or hypertext links to third-party sites or information are provided solely as a convenience to you and do not in any way constitute or imply ECR's endorsement, sponsorship or recommendation of the third party, information, product or service. ECR is not responsible for the content of these pages and does not make any representations regarding the content or accuracy of material in this file.

As per copyright regulations, any unauthorised use of the material or parts thereof as well as commercial reproduction or multiple distribution by any traditional or electronically based reproduction/publication method is strictly prohibited.

You agree to defend, indemnify, and hold ECR harmless from and against any and all claims, damages, costs, and expenses, including attorneys' fees, arising from or related to your use of these pages.

Please note: Links to movies, ppt slideshows and any other multimedia files are not available in the pdf version of presentations.
Purpose

Early detection and correct assessment of the anatomical allocation of colorectal liver metastases (CRLM) is essential for subsequent treatment. The decision regarding the nature of management (surgical, palliative, interventional or a combination) along with planning the extent of resection in case of surgery is mostly made on the basis of imaging. Presently, such imaging is accessible to the majority of physicians involved in patients care. Thus, it is important, that physicians of all radiological knowledge are able to comprehend and reproduce the diagnosis of CRLM.

Many studies were made in the past comparing MRI with CT to evaluate the best imaging modality before the resection of CRLM. The conclusions from prior research suggest that MRI modality significantly provides higher Sensitivity (1, 2) and higher inter-rater agreement (3). Professionals recruited in these studies have always been of a comparable level of radiological knowledge (1, 3). This study uses four reviewers of different levels of radiological experience in an attempt to prove, whether interpreting CT or MRI is greater influenced by experience of the diagnosing physician. The purpose of the study is to explore, whether the coherence of experience and the diagnostic performance of CT and MRI regarding CRLM is comparable and if the advantages of MRI are utilizable for both beginner and experienced radiologists.
Methods and Materials

36 CT and 33 MRI scans of 69 patients were included in this study. We decided to focus on the main imaging modalities MRI and CT and dismissed the separate exploration of PET-CT and we included one additional reviewer, to enlarge the data of the study. This poster represents on-going work as the study is projected to include at least 60 MRI and 60 CT scans by its completion. The included scans were the most recent prior to surgery and were performed from years 2006-2010. All patients had received neoadjuvant chemotherapy before the imaging.

All scans were diagnosed by four blinded reviewers: three were radiologists with respectively over 15 years, over 5 years and over 3 months of experience and one was a sixth year medical student with zero formal years experience. The diagnosing took place on PACS Work Stations, where the reviewers could find 3 cohorts of patients to diagnose. MRI and CT scans were randomly spread among cohorts and were not chronologically arrayed. Criteria in diagnosing were to assess the number of CRLM and to name the affected segments. Reviewers were aware, that every patient had CRLM, but were not aware of previous treatment and outcome.

To create a reference standard regarding number of lesions and anatomical allocation with the highest accuracy, the original radiological reports have been correlated with surgical and histopathological findings. Both imaging and surgery/histopathology were included for the reference standards. This is because discrepancies in the mentioned number of lesions and affected segments between the radiologist and the surgeon/histopathologist were found in the majority of patients, except in those with one single lesion. Furthermore, the data from the surgical and histopathological report often was found to be an insufficient standard of reference. Specifically cases were reported as being similar to the preoperative imaging but without distinct details. The diagnosis of the initial radiologist had therefore firstly to be verified by palpation and intraoperative ultrasound (IOUS), which had been documented by the surgeon and secondly by the histopathologist, who definitively confirmed the diagnosis of CRLM. In case of doubt, the intraoperative and histopathological findings remained the gold standard.

Following this method, we included 84 CRLM for MRI and 110 CRLM for CT in our reference standard. To assess accuracy on anatomical allocation, we included exclusively liver segments diagnosed radiologically, surgically and histopathologically as positive for CRLM in our reference standard group, namely 61 MRI segments and 55 CT segments. An additional group included segments, neither named by the radiologist, nor surgeon/histopathologist as affected to detect false positive diagnosing, totalling 141 MRI segments and 134 CT segments.

The reviewers' results were compared with the reference standard to verify the findings and identify correctness regarding the affected segments. The study dismissed the use
of ROC-Analysis, which was considered to be inappropriate for our study in the final analysis. Instead, percentages of sensitivity and specificity on a per-patient basis were calculated and as well percentages of correctly identified affected and non-affected segments. To evaluate the inter-rater agreement regarding number of lesions, the intra-class coefficients (ICC) were calculated.
Results

Subsequently the percentages are listed in this order: reviewer A (a radiologist with more than 15 years), B (a radiologist with more than 5 years), C (a radiologist with more than 3 months of experience) and reviewer D (a 6th year medical student).

MRI sensitivity was 98.6%, 91.9%, 93.9%, 81.6%, and specificity was 98.8%, 88.4%, 82.1%, 91.4%.

CT sensitivity was 90.7%, 95.9%, 86.7%, 87.5% and specificity was 98.8%, 88.4%, 82.1%, 91.4%. (Please refer to Fig.1 and 2)

Percentages of correctly diagnosed affected segments for MRI were 93.4%, 80%, 67.2%, 62.3% and for CT 98%, 81.4%, 70.9%, 83.5%. (Refer to Figure 4)

Percentages of correctly diagnosed non-affected segments were for MRI 98.6%, 97.4%, 88.7%, 88.7% and for CT 97.8%, 92.2%, 86.1%, 82.4%. (Refer to Fig.5 and 6)

The inter-rater agreement regarding diagnosed numbers of lesions was higher for CT with a intra-class correlation coefficient (ICC) of 0.79 (p=0.000) in comparison to MRI with a ICC of 0.70 (p=0.000). Furthermore, when we compared the ICC for respectively two readers, the coefficients were higher at all time for CT, compared with MRI, as outlined in Fig. 7. (Refer to Fig.7)
Fig. 1: average per-patient sensitivity

The values spread over a larger range regarding MRI than CT. Reviewer A had tendentially the best values, followed in sequence by reviewer B, C and D. However, the more experienced reviewers did not always have better values than the lower experienced reviewers for both modalities.

Fig. 1: sensitivity

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Fig. 2: average per-patient specificity
As an single expection, an inversion of values was found: reviewer D had better specificity values for both CT and MRI than reviewer C.
Fig. 3: sensitivity and specificity overview

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Fig. 4: correctly diagnosed affected segments

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Fig. 5: correctly diagnosed non-affected segments

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Fig. 6: correctly diagnosed affected and non-affected segments overview

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Intra-class correlation coefficients (ICC)

<table>
<thead>
<tr>
<th>ICC of reviewer</th>
<th>CT</th>
<th>MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A + B</td>
<td>0.84</td>
<td>0.71</td>
</tr>
<tr>
<td>A + C</td>
<td>0.66</td>
<td>0.65</td>
</tr>
<tr>
<td>A + D</td>
<td>0.88</td>
<td>0.66</td>
</tr>
<tr>
<td>B + C</td>
<td>0.76</td>
<td>0.65</td>
</tr>
<tr>
<td>B + D</td>
<td>0.84</td>
<td>0.81</td>
</tr>
<tr>
<td>C + D</td>
<td>0.75</td>
<td>0.7</td>
</tr>
<tr>
<td>overall A+B+C+D</td>
<td>0.79</td>
<td>0.7</td>
</tr>
</tbody>
</table>

**Fig. 7:**
Intra-class correlation coefficients (ICC), quantifying the inter-rater agreement regarding the diagnosed number of CRLM. The inter-rater agreement was higher for CT with an ICC of 0.79 in comparison to MRI with an ICC of 0.70. Furthermore, when we compared the ICC for respectively two readers, the coefficients were higher at all time for CT, compared with MRI.

**Fig. 7:** ICC table CT and MRI

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Conclusion

It can be seen in Fig. 3 and 6 that tendentially reviewer A had the best values regarding sensitivity, specificity and correct anatomical allocation. This was in sequence followed by reviewer B, C and D. However, this finding is not absolute, as the more experienced reviewers did not always have better values than the lower experienced reviewers for both modalities. As an exection, a complete order of experience was generated in the course of testing the percentages of correctly diagnosed non-affected segments (see Fig.5). As an other exection, reviewer D had better specificity values for both CT and MRI than reviewer C (see Fig.2).

In detail, reviewer A had the best values for sensitivity and specificity regarding MRI. For CT scans, reviewer A had a 5.2% lower sensitivity than reviewer B, but simultaneously reviewer B had a 9.7% lower specificity. Within the framework of testing the percentages of correctly diagnosed affected and non-affected segments, reviewer A had higher percentages for both CT and MRI.

As seen in Figure 3, the percentages of reviewer C and D spread over a larger range regarding MRI (12.3% regarding sensitivity and 9.3% concerning specificity) compared with CT (0.8% regarding sensitivity and 2.2% concerning specificity). Furthermore, the higher sensitivity of reviewer C compared with D was due to a nearly equally lower specificity. (see Fig.3)

Interestingly it was found, when comparing the values of the reviewers altogether, that both sensitivity and specificity values spread over a larger range regarding MRI than CT. This was most distinct concerning sensitivity. The spread of sensitivity values for CT was 9.2%, while the MRI percentages spread over a range of 17%. This suggests that the values and thus the quality of diagnosing were more comparable for CT.

Additionally, the inter-rater agreement regarding diagnosed numbers of lesions was higher for CT with an ICC of 0.79 in comparison to MRI with an ICC of 0.70 (see Fig.7). This indicates in another regard, that there was a higher consensus and similarity of the diagnosing by CT compared to MRI.

In summary, the coherence of radiological experience and quality of diagnosing does not seem to be comparable between MRI and CT in identifying and assessment of CRLM. We currently interpret that our data indicates MRI to be the imaging modality, whose diagnosing is affected most by experience regarding the diagnosing of CRLM. In our study, the data and values concerning CT were more comparable and similar in comparison with MRI.
Fig. 1: average per-patient sensitivity
The values spread over a larger range regarding MRI than CT. Reviewer A had tendentially the best values, followed in sequence by reviewer B, C and D. However, the more experienced reviewers did not always have better values than the lower experienced reviewers for both modalities.
**Fig. 2:** specificity

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Fig. 3: sensitivity and specificity overview

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Fig. 4: Correctly diagnosed affected segments

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Fig. 5: correctly diagnosed non-affected segments

Fig. 5: percentages of the segments, correctly diagnosed as negative for CRLM. The reference standart were the segments, diagnosed nor radiologically neither surgically/histopathologically in the course of the original diagnosing. Hereby, a complete order of experience was generated.
**Fig. 6:** correctly diagnosed affected and non-affected segments overview

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
Intra-class correlation coefficients (ICC)

<table>
<thead>
<tr>
<th>ICC of reviewer</th>
<th>CT</th>
<th>MRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A + B</td>
<td>0.84</td>
<td>0.71</td>
</tr>
<tr>
<td>A + C</td>
<td>0.66</td>
<td>0.65</td>
</tr>
<tr>
<td>A + D</td>
<td>0.88</td>
<td>0.66</td>
</tr>
<tr>
<td>B + C</td>
<td>0.76</td>
<td>0.65</td>
</tr>
<tr>
<td>B + D</td>
<td>0.84</td>
<td>0.81</td>
</tr>
<tr>
<td>C + D</td>
<td>0.75</td>
<td>0.70</td>
</tr>
<tr>
<td>overall A+B+C+D</td>
<td>0.79</td>
<td>0.70</td>
</tr>
</tbody>
</table>

**Fig. 7:**
Intra-class correlation coefficients (ICC), quantifying the inter-rater agreement regarding the diagnosed number of CRLM. The inter-rater agreement was higher for CT with an ICC of 0.79 in comparison to MRI with an ICC of 0.70. Furthermore, when we compared the ICC for respectively two readers, the coefficients were higher at all time for CT, compared with MRI.

**Fig. 7:** ICC table CT and MRI

© Institute for Diagnostic and Interventional Radiology, J.W. Goethe University Hospital - Frankfurt am Main/DE
References


Personal Information

Moritz Albrecht, MoritzAlbrecht@gmx.net