Beyond diagnostic accuracy. Applying and extending methods for diagnostic test research
Glas, A.S.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Computed Tomography versus Magnetic Resonance Imaging in the reduction of diagnostic uncertainty in patients with lumboradicular syndrome

Afina S. Glas, Nina M. Klemetsö, Jeroen C. van Rijn, Jan Stam, Wilco C. Peul, Charles B.L.M. Mojoie, Gerard J. den Heeten, Patrick M.M. Bossuyt

Abstract

Background - MRI and CT have similar diagnostic accuracy in the diagnosis of patients with lumboradicular symptoms. It is not clear which test to use in daily clinical practice. The aim of this study was to assess the information content of both imaging modalities using the entropy statistic.

Methods - Eligible were patients (18-70 years of age) with radicular pain in whom operative treatment was considered. Findings from the patients' history and physical examination were presented to two physicians in a standardised format. Based on these descriptions both physicians independently assigned the probability of 7 different entities (3 levels, left or right and a category other) on separate visual analogue scales. In two separate sessions, they interpreted the images from CT and MRI. In each session, new probabilities were assigned again. By comparing the pre-test and post-test probabilities for each patient we calculated the information content for each test, in terms of entropy, using a random effects logistic regression model. In addition the physicians were asked after CT and after MRI how certain they were about operating patients and whether additional diagnostic tests were deemed necessary.

Results - The data on 59 patients were available for the analysis. MRI had a significantly higher information content than CT (0.61 bits and 0.73 bits versus 0.43 bits and 0.58 bits, p < 0.01). There was significantly less uncertainty after MRI in the decision whether to operate or not (p=0.003), but not one of the tests on average provided 100% certainty. In up to 49% of cases additional diagnostics were judged to be necessary.

Conclusion - MRI seems to contain significantly more information than CT in the diagnosis of patients with lumboradicular symptoms, leading to higher degrees of certainty in clinical decision making.
9 Computed Tomography versus Magnetic Resonance Imaging in the reduction of diagnostic uncertainty in patients with lumbaradicular syndrome

Introduction
Despite several comparative studies of computed tomography (CT) and magnetic resonance imaging (MRI) in the diagnosis of patients with lumbosacral radicular pain it is still unclear which of the two imaging tests serves best for this indication. The reported sensitivities and specificities range from 62 to 90% for CT versus 60 to 100% for MRI, with specificities between 70 to 87% for CT versus 43 to 97% for MRI.1

It is well possible that both techniques suffice in terms of diagnostic accuracy. Research on the diagnostic and therapeutic impact of these techniques for patients suspected of lumbar disc herniation is scarce for MRI2,3 and even non-existent for CT. Diagnostic impact is the effect a diagnostic test result has on the degree of certainty a physician has regarding a particular diagnosis and future management.4,5 One of the measures to express this level of certainty quantitatively is entropy, a measure of information content derived from conventional information theory.6,8 The difference in before and after test entropy is an indication of the information a test holds and subsequently can be used to compare the diagnostic impact of tests. The aim of this study was to compare CT with MRI in the diagnosis of herniated nucleus pulposus, using the entropy statistic and to explore if any differences in entropy were related to differences in therapeutic decision making. We estimated entropy from probabilities assigned by the physician to every patient, expressing the likelihood of disc herniation at 6 particular locations: HNP on disc space level L3/L4 left and right, L4/L5 left and right, L5/S1 left and right. A seventh category was used for other diagnoses. A similar study has been conducted by Albeck and colleagues, using the pseudo regret function to measure the informativeness of tests.9

Patients and Methods

Patients
Data were obtained as part of a prospective study to evaluate the diagnostic process of patients with low-back pain, performed at a Dutch teaching hospital, between June 1999 and June 2000. Eligible were adult patients, between 18 and 70 years of age, with (incapacitating) radicular pain, in whom conservative treatment for at least 4 weeks had been unsuccessful. In these patients operative treatment was considered and diagnostic imaging is used in the decision making process. These criteria for inclusion reflect the consensus statement on diagnosis and treatment of lumbosacral root entrapment syndromes by the Dutch Society for Neurology.10 Excluded were patients who 1) had undergone previous back
surgery, 2) had already diagnosed lumbar disc herniation 3) could not give informed consent, 4) were pregnant, 5) had a contraindication for MRI investigation. The Institutional Review Board had approved the study protocol. Only informed and consenting patients were included in the study.

**Diagnostic procedure**

This study focused on the information content of the two imaging modalities, CT and MRI, conditional on the information already available through history and clinical examination.

Lumbar CT examinations were performed on a Philips Elscint TWIN CT scanner. Helical CT was made with 1.1 mm collimation and 0.5 mm increments from the level of L3 to the bottom of S1. The gantry angle was angled approximately through the disc space of L4-L5. Reformatted axial sections were made using both the soft tissue and bone windows of all scanned interspaces, precisely parallel to the interspaces to achieve an optimal delineation of the disk. In addition reformatted sagittal sections were made. The distance of the sections was 4.0 mm and the thickness of the sections was for soft tissue window 4.0 mm.

Lumbar MRI examinations were performed with on a 1.5 T Signa Scanner (General Electric Company). The protocol included sagittal T1-weighted (TR 500, TE 14) spin-echo images and proton density T2-weighted (TR 3500, TE 120/20) fast spin-echo images with 4 mm slice thickness and 0.5 mm intersection gap; matrix 200x512; field of view 29 x 29 cm. In addition axial T1 (TR 520, TE 12)- and T2 (TR 4500, TE 120) -weighted fast spin-echo images were obtained from the level of L3 to the bottom of S1 with 4 mm slice thickness and 0.5 mm intersection gap; matrix 200x256; field of view 15 x 15 cm. Axial images were obtained without angulation. Finally, heavily T2- weighted (TR 5000, TE 252) fast spin-echo oblique MR myelography was performed with two slices of 20 mm thickness images; matrix 250 x 220; field of view 16 x 16 cm.

To standardise the clinical information the patients were presented as case descriptions to an experienced neurologist and a neurosurgeon. These experts then assigned seven probabilities on a visual analogue scale (VAS) for seven possible entities: hernia at disc space L3/L4 left or right, L4/L5 left or right, and L5/S1 left or right and 'other'. In separate sessions, the physicians were presented with the same case descriptions, this time accompanied by the matching CT or MRI images with the accompanied report from the radiologist. The experts were asked to assign new probabilities on seven similar VAS scales. The two sessions were separated in time by at least one month. The two physicians were also asked to indicate how certain they were in making a decision on operative treatment and
9 Computed Tomography versus Magnetic Resonance Imaging in the reduction of diagnostic uncertainty in patients with lumboradicu lar syndrome

whether other diagnostic tests were deemed necessary after interpreting the image tests.

**Analysis**

Entropy was used to measure the information content of the CT and MRI. This measure, developed within, information theory, expresses information in bits. For a single condition that is either present or absent and a dichotomous test, an entropy change of 1 bit after the test result has been obtained would mean that the pre-test probability of 0.5 - reflecting maximal uncertainty - is updated to either 1 or 0, reflecting maximal certainty.

To estimate the information content of CT and MRI the difference between the pre-test and post-test entropy was calculated for each patient. The pre-test entropy is estimated as

$$H_{\text{pre}} = - \sum_{i=1}^{7} P(D_i) \log_2 P(D_i)$$

where $D_i$ is one of the 7 entities and the post-test entropy as

$$H_{\text{post}} = - \sum_{i=1}^{7} P(D_i|T) \log_2 P(D_i|T).$$

The probabilities per entity were taken from the scales. As the seven categories are considered to be mutually exclusive, the corresponding probabilities were normalised to sum up to 100%.

We then calculated the mean change in entropy per observer over all patients both for MRI as well as for CT to express the information content of the two techniques.

To test the null hypothesis of similarity in information content for CT and MRI, we used a paired t-test for each observer separately. For the two observers jointly, we used a mixed analysis of variances model. Such a mixed model accounts for the information content of CT and MRI of the two observers being determined in the same patients. A similar analysis was performed on the uncertainty expressions in the decision to operate. Computations were done with the statistical software package SPSS version 11.

**Results**

Of the 64 consecutively included patients, 5 had to be excluded for the following reasons; n=3 claustrophobia and n=2 loss to follow up. The characteristics of the 59 patients available for the analysis can be found in table 1.
Table 1 Clinical and demographic characteristics of the included patients, n=59.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>39 %</td>
</tr>
<tr>
<td>Mean age, years (SD)</td>
<td>44 (12)</td>
</tr>
<tr>
<td>BMI (SD)</td>
<td>25 (3.6)</td>
</tr>
<tr>
<td>History of low back pain/leg pain</td>
<td>74 %</td>
</tr>
<tr>
<td>Median duration symptoms, months (IQR)</td>
<td>3 (2–6.8)</td>
</tr>
</tbody>
</table>

Symptoms

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low back pain</td>
<td>88 %</td>
</tr>
<tr>
<td>Leg pain</td>
<td>98 %</td>
</tr>
<tr>
<td>Pain if ↑ abdominal pressure</td>
<td>78 %</td>
</tr>
<tr>
<td>Paresthesia</td>
<td>30 %</td>
</tr>
</tbody>
</table>

Of these patients 24 (41%) were operated on. The surgical findings were: 1 herniated nucleus pulposus (HNP) on the level lumbar disc 2 and 3 (L2-L3) left, 12 HNP’s L4-L5 (7 right, 5 left sided), 8 HNP’s L5-S1 (4 right, 4 left sided) and 3 patients with root canal stenosis. Of the patients that were assigned to conservative therapy 27 had a HNP on the imaging tests. In these patients the location of the HNP did not match the nature of the complaints or the severity of symptoms had decreased while waiting for surgery.

Figure 1 illustrates, for each patient, the entropy after CT or MRI (Y-axis) relative to the entropy prior to testing for that same patient (X-axis). Deviations from the diagonal reflect a change in entropy, standing for a change in uncertainty. Points above the diagonal reflect a decrease in certainty and points below the diagonal indicate an increase in certainty for the respective patients. Both tests seem to decrease uncertainty for most patients, with only a few points lying above the diagonal.

Table 2 shows the results of the formal analysis. In theory, uncertainty (entropy) is maximal when all possible entities have an equal probability. As we have 7 entities, this would imply seven times a probability of 0.14 (1/7), and the maximal entropy would be 2.8 bits. The pre-test entropy, based on the clinical information only, is close to 1 for both observers, indicating that most uncertainty is eradicated after history and physical examination.
Computed Tomography versus Magnetic Resonance Imaging in the reduction of diagnostic uncertainty in patients with lumbaradicular syndrome

![Scatter plots of the pre-test entropy versus the post-test entropy for both CT and MRI.](image)

The dots above the diagonal indicate patients where the physician's uncertainty of a diagnosis increased after the knowing the test result. The theoretical maximum entropy possible is 2.8 for 7 entities.

From table 2, one can observe that MRI leads to a higher reduction in entropy than CT. The overall difference between the information content of CT and MRI incorporating the results of both observers was significant (p= 0.01).

### Table 2 Results of the information content in bits.

<table>
<thead>
<tr>
<th></th>
<th>Observer S</th>
<th>Observer P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (se)</td>
<td>Mean (se)</td>
</tr>
<tr>
<td>Pre-test entropy after H&amp;A</td>
<td>0.97 (0.044)</td>
<td>1.1 (0.037)</td>
</tr>
<tr>
<td>Post CT</td>
<td>0.54 (0.059)</td>
<td>0.53 (0.070)</td>
</tr>
<tr>
<td>Post MRI</td>
<td>0.35 (0.062)</td>
<td>0.38 (0.054)</td>
</tr>
<tr>
<td>Information CT</td>
<td>0.43 (0.072)</td>
<td>0.58 (0.079)</td>
</tr>
<tr>
<td>Information MRI</td>
<td>0.61 (0.076)</td>
<td>0.73 (0.063)</td>
</tr>
</tbody>
</table>

P=0.008 for the difference between CT and MRI for observer S, and p=0.066 for observer P. (paired t-test)
Table 3 Results of the entropy change in the uncertainty to operate.

<table>
<thead>
<tr>
<th></th>
<th>Observer S Mean (se)</th>
<th>Observer P Mean (se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial entropy after H&amp;A</td>
<td>0.61 (0.030)</td>
<td>0.39 (0.045)</td>
</tr>
<tr>
<td>Post CT</td>
<td>0.31 (0.043)</td>
<td>0.23 (0.040)</td>
</tr>
<tr>
<td>Post MRI</td>
<td>0.15 (0.032)</td>
<td>0.19 (0.035)</td>
</tr>
<tr>
<td>Information CT</td>
<td>0.30 (0.037)</td>
<td>0.15 (0.050)</td>
</tr>
<tr>
<td>Information MRI</td>
<td>0.46 (0.031)</td>
<td>0.20 (0.052)</td>
</tr>
</tbody>
</table>

P=0.01 for the difference between CT and MRI for observer S, and p=0.37 for observer P. (paired t-test)

Table 3 shows the results of the uncertainty in the decision to operate the patient or not. Physicians reported less uncertainty about operative treatment after MRI compared to CT (p=0.003).

We also asked the physicians if they needed additional diagnostic information for the decision to operate. To reach a decision on the need for surgery, observer S needed additional diagnostic information in 47% (28/59, all MRI's) of patients after interpreting CT and in 3.4% (2/59, 1 MRI and 1 CT) of patients after interpreting MRI (p<0.001). These values were 49% (29/59, all MRI's) and 12% (7/59, 4 MRI's, 1 CT, 1 combination (MRI with intravenous contrast and CT) and 1 myelography) respectively for observer P (p<0.001).

Discussion
In this study the information content of CT and MRI was evaluated in the diagnosis of herniated nucleus pulposus in patients with lumbar radicular syndrome. We found MRI to yield more information than CT, leading to a higher reduction in uncertainty on the presence and the location of a herniated nucleus pulposus and a higher level of certainty in the decision to operate. Both observers needed additional diagnostic information in order to make this decision for almost half of the patients after CT. For MRI significant less additional diagnostic tests were needed. This underscores the finding that physicians are more confident after interpreting MRI relative to CT. This might be due to the higher resolution of MRI and as such a better contrast between the dural sac and herniation. The reason that for some patients after MRI a CT scan was needed might be due to uncertainty regarding possible bone abnormalities as a cause of the complaints.

Multiple studies have evaluated the difference in discriminating performance between MRI and CT. Recommendations based on the results from these studies
Computed Tomography versus Magnetic Resonance Imaging in the reduction of diagnostic uncertainty in patients with lumbaradicular syndrome

were equivocal in terms of which one is to be preferred over the other. Albeck and colleagues have previously compared the information content of the MRI, CT and myelography in diagnosing lumbar disc herniation. They concluded that CT produced more certainty relative to MRI, but that this difference was not statistically significant. Myelography yielded significant lower information compared to CT and MRI. Myelography has become obsolete in our hospital, because of its invasiveness and the availability of MRI or CT. In contrast with the paper of Albeck, our study revealed that MRI held significantly more information than CT. This might be due to increased refinements in MRI technique relative to CT from the time Albeck and colleagues collected the data. We do not think our physicians were biased because of a pre-existing preference for MRI. The time interval between CT and MRI interpretation was long enough to prevent such memory bias in favour of MRI. Albeck and colleagues used the pseudo regret function. The interpretation of this measure is similar to the entropy measure. Both reach a maximum at maximal uncertainty, and are zero when uncertainty no longer exists. Since we incorporated the differential diagnosis consisting of 7 different entities, the entropy statistic was more convenient. In addition it was not needed to verify the true diagnosis in the patients.

We are aware of three studies that studied the impact of MRI on the management of patients with possible disc herniation. They all concluded that MRI had considerable impact on confidence in the diagnosis and patient management. One study found that despite this impact no effect was seen on patient outcome. Unfortunately none of these studies compared MRI with CT.

The design of the study did not allow us to evaluate the consequences of extra information content of MRI on patient outcome. It is possible that the difference in entropy is inconsequential, leading to similar health outcomes in the end. The results cannot be unconditionally generalized to other settings, as we relied on the readings of two observers only. A more stable conclusion could be drawn when for more observers a comparable result was found.

The reason this study was performed was to investigate which test to use within a guideline in the workup of patients suspected for herniated nucleus pulposus. Although MRI was superior in the information content we can't state with firm conviction that MRI should become the one and only test to do. Depending on resources like costs or availability, CT is an adequate alternative imaging test in diagnosing the possible cause of patients with lumbaradicular syndrome.
Acknowledgements
We thank Kees Albrecht for his advise on the design of this study and we thank Frank Jan Hulsman for interpreting the images.

References