THE RELIABILITY AND DISTINGUISHABILITY OF ULTRASOUND DIAGNOSIS OF OVARIAN MASSES

ALIREZA AKBARZADEH BAGHEBAN, FARID ZAYERI, FATEMEH BARADARAN ANARAKI', ZAHRA ELAHIPANAH

ABSTRACT

BACKGROUND: For any radiologist, intra-observer agreement in observing and decision making in diagnosis of any disease is of great importance, and so is observing and reading ultrasound pictures of ovarian masses and distinguishing amongst their categories. AIMS: In this study, the reliability and consistency of ultrasound diagnosis of ovarian tumors have been evaluated. SETTINGS AND DESIGN: Two experienced and three less experienced radiologists assessed ultrasounds of 40 patients of Mirza Koochak Khan Hospital in Tehran, Iran, in 2005. MATERIALS AND METHODS: In this prospective observational study, the ultrasounds were performed by an expert radiologist, with a single apparatus. These ultrasounds have been evaluated separately and independently in two periods (with a 1-week interval). STATISTICAL ANALYSIS USED: Weighted kappa was used to calculate intra-observer agreement (reliability), and two statistical models were applied to assess category distinguishability (consistency). SPSS version 10, SAS version 8, and EXCEL 2003 have been used to do an appropriate statistical analysis. RESULTS: Mean of weighted kappa was 0.81, and mean of distinguishability was 0.995 for our experienced radiologists, due to their superior results. Because of weaker results obtained by the less experienced radiologists, mean of weighted kappa and mean of distinguishability were 0.65 and 0.967 respectively. Overall mean of distinguishability for benign and borderline categories was 0.969; and for malignant and borderline categories, it was 0.987. CONCLUSION: Although experienced radiologists functioned better than the less experienced radiologists, all of them showed appropriate distinguishability and intra-observer agreement in diagnosis and categorization of the ovarian masses. Distinguishing benign category from borderline was more difficult than distinguishing malignant category from borderline. In general, experienced radiologists showed better results compared to less experienced radiologists.

Key words: Distinguishability, ovarian mass, reliability, ultrasound

INTRODUCTION

Suppose a radiologist classifies each ultrasound in a sample on an ordinal scale at two different times, so that the first evaluation has no effect on the second one; we could show these two ratings by a contingency table and assess two important issues:

• Intra-observer agreement of the observer at two different times. This actually is the reliability of the observer in decision making. [1]

• Distinguishability by the observer in categorizing the samples. When we have ordinal categories, distinguishability of these categories is of great concern, which could show us the ability of the observer in differentiating different categories from each other. [2]

The majority of ordered categories are subjective definitions, and distinguishability by an observer between two close categories is difficult, even for those who are experts. [3] In general, to assess reliability and consistency, kappa and weighted kappa coefficients were used. [3-5] Utilizing these by themselves has some disadvantages, and the results could show some errors as well; therefore, many researchers have recommended using statistical models, in addition to measuring these coefficients for arriving at more complete conclusions. [2, 6-9] In this study, we have evaluated the first issue by weighted kappa and the second one by statistical models for ovarian mass data.

MATERIALS AND METHODS

This is a prospective observational study. The data were gathered from the radiology department of Mirza Koochak Khan Hospital in Tehran, Iran, in January 2005. After obtaining consent from 40 women whose ultrasounds were performed by an expert radiologist and just with a single apparatus (in order to minimize the performer bias), two experienced radiologists and three less experienced radiologists evaluated these ultrasounds separately and independently and scored them 1 through 3 for benign, borderline, and malignant cases respectively. In a single blind study, each one of these ultrasounds was reevaluated by our observers for a second time after a week. This period (a week) seems reasonable, because our observers would not recall the ultrasounds after a week and we would not encounter loss of quality of ultrasounds in this short period. Cross classification of these observers at two different times provided five different 3×3 tables, and the tables were used as the basis of our analysis.

In this study, intra-observer agreement of the raters has been evaluated by weighted kappa (as index of reliability), and the distinguishability
Table 1: Weighted kappa and distinguishability in different 3×3 tables of the observers

<table>
<thead>
<tr>
<th>Observer</th>
<th>Weighted Kappa</th>
<th>Distinguishability Benign from borderline</th>
<th>Borderline from malignant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less experienced radiologist 1</td>
<td>0.61</td>
<td>0.9423</td>
<td>0.9717</td>
</tr>
<tr>
<td>Less experienced radiologist 2</td>
<td>0.69</td>
<td>0.9601</td>
<td>0.9864</td>
</tr>
<tr>
<td>Less experienced radiologist 3</td>
<td>0.65</td>
<td>0.9640</td>
<td>0.9764</td>
</tr>
<tr>
<td>Experienced radiologist 1</td>
<td>0.75</td>
<td>0.9867</td>
<td>0.9999</td>
</tr>
<tr>
<td>Experienced radiologist 2</td>
<td>0.87</td>
<td>0.9940</td>
<td>0.9999</td>
</tr>
</tbody>
</table>

The less experienced radiologists demonstrated lower distinguishability in categorizing different categories (minimum 0.95 for benign and borderline [1 and 2] and minimum 0.97 for borderline and malignant [2 and 3] entities) [Figure 1]. These raters had an overall distinguishability mean of 0.967, and it was a little lower compared to the experienced radiologists. Mean of weighted kappa for them was 0.65.

The mean of distinguishability for benign and borderline categories was 0.990 for the experienced radiologists and 0.955 for the less experienced radiologists. Besides, the experienced radiologists and the less experienced radiologists had a mean of 0.999 and 0.978 respectively for distinguishing the borderline and malignant cases.

DISCUSSION

To compare distinguishability demonstrated by the observers in categorizing the samples and assessing intra-observer agreement for each one of them, we computed weighted kappa at first. Although there was no complete intra-observer agreement for these observers at two different times, by considering 0.71 for mean of weighted kappa, it can be stated that there was good overall reliability. Besides, minimum and maximum of weighted kappa in our study have been obtained to be 0.61 and 0.86 respectively.

Our findings confirm the results reported by Amer et al.[10] They found 69.4% for the mean intra-observer agreement (kappa = 0.54). One reason for a small difference in reliability index is that they used kappa instead of weighted kappa.

Although the less experienced radiologists demonstrated a lower distinguishability compared to the experienced radiologists, yet this difference was not remarkable; because all the observers had a minimum 0.90 to distinguish between adjacent categories. But for all observers, distinguishability between categories 1 and 2 was lower than that between categories 2 and 3; and experienced radiologists showed better results than the less experienced radiologists.

Generally, for assessing validity and reliability of diagnosing among different categories of ovarian cysts, kappa and weighted kappa coefficients are used.[12] These coefficients show intra-observer agreement generally; and by considering several deficiencies that were reported for them in multiple studies[2,5,7,8] and their inability to show distinguishability by observers, we used statistical models to consider distinguishability demonstrated by them to classify different ordered categories. We could use these results for better future training of raters in big epidemiological studies.

REFERENCES

2. Perkins SM, Becker MP. Assessing rater

Source of Support: Nil
Conflict of Interest: None declared.

UTILITY OF CYSTICERCUS FASCIOLARIS ANTIGEN IN DOT ELISA FOR THE DIAGNOSIS OF NEUROCYSTICERCOSIS

NITIN SHUKLA, NUZHAT HUSAIN, GIRDHAR G AGARWAL, MAZHAR HUSAIN

ABSTRACT

BACKGROUND: Clinical diagnosis of neurocysticercosis (NC) is established by CT scan and MRI. However, absolute diagnosis is not possible in a fair number of cases, and serological assays are used as adjunct. Besides, CT scan and MR imaging are resource-intensive tests and not practical for screening in endemic areas.

AIM: To provide a low-cost, efficient, and reproducible assay for the detection of antibodies against cysticerci. Hence we have attempted to standardize and evaluate the diagnostic utility of the cysticercus fasciolaris antigen in a Dot ELISA assay for diagnosis of NC.

SETTING AND DESIGN: Tertiary hospital-based, case-control series.

MATERIALS AND METHODS: Confirmed cases of NC diagnosed by presence of ring lesions in CT scan or MR imaging with presence of scolex were taken as positive controls (n = 50). Negative controls (n = 50) included subjects with normal CT scan studies (n = 30) and diseased controls with ring lesions in CT scan confirmed to be neurotuberculosis (n = 20). Dot ELISA was standardized and validated with commercially available ELISA (UBI, USA) using sera from the study groups.

STATISTICAL ANALYSIS: The immunodiagnostic performance of the two tests was compared using the chi-square test. A P value <0.05 was considered significant.

RESULTS: The Dot ELISA had a sensitivity of 88% and specificity of 74% with a positive predictive value of 77.19% and negative predictive value of 81.06%. Likelihood ratios for a positive and a negative test were 3.4 and 0.2. The sensitivity and specificity of the commercial ELISA were 92% and 84% respectively.

CONCLUSIONS: Dot ELISA has sensitivity and specificity comparable to ELISA for the diagnosis of NC. The test is simpler, not requiring expertise and instrumentation. Further validation of the test as a screening tool is required.

Key words: Cysticercosis, Dot enzyme-linked immunosorbent assay (ELISA), ELISA

INTRODUCTION

Neurocysticercosis, an infestation caused by lodging of larval stage of Taenia solium in the brain, is the most common parasitic disease of the central nervous system. It is reported...