

The international use of PERFORMS mammographic test sets



Yan Chen, Leng Dong, Hossein Nevisi, Alastair Gale



Applied Vision Research Centre, Loughborough University, UK
y.chen@lboro.ac.uk

Abstract

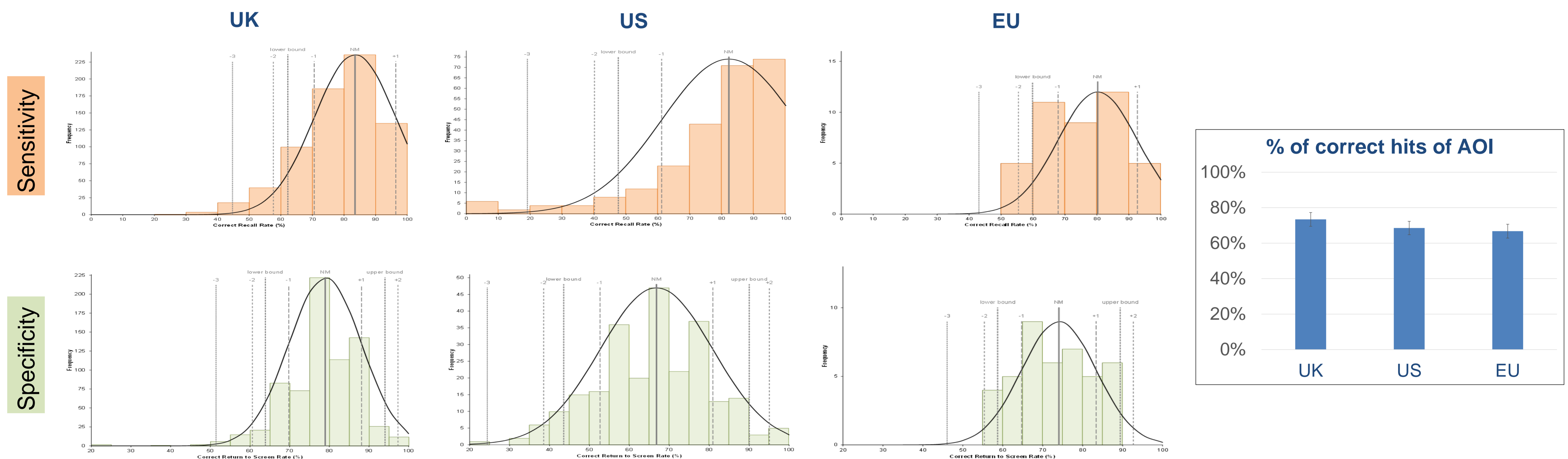
To examine the utility of employing breast screening test sets internationally the data of 1,009 radiologists from the USA, UK and other European countries were examined as they inspected 20 carefully selected difficult recent screening cases. Some 720 UK radiologists, 247 American and 42 European radiologists took part. Whilst similar sensitivity scores between the three groups were found, the main difference was the lower specificity of the American radiologists reflecting their different recall clinical practice. It is argued that using test sets internationally provides participants with useful comparative performance information whilst also providing data on how the same cases are interpreted by radiologists from different countries.

Introduction

There is international interest in using test sets to examine the performance of individual radiologists, and other screeners, from different countries when they examine and report the same set of cases. Such comparative examination can highlight particular training needs of groups or of individual radiologists. Ideally any woman attending for screening, usually by full field digital mammography (FFDM), in any country should expect that their mammograms will be reported in a similar fashion so that any potential early signs of cancer are identified at the earliest treatable stage. The use of such test sets has been nationally used in the UK for over 25 years. Here we report an initial investigation of how radiologists from different countries perform when they use such test sets.

Method

In order to determine how radiologists from different countries respond to the same set of challenging test cases, a set of 20 FFDM cases were selected from the PERFORMS case database which had been reported by 720 UK participants. This set, named PERFORMS mini-lab, was also examined by two other groups: 247 US participants at a Breast Symposium (Orlando, 2015), and 42 European participants at EUSOBI conference (London, 2015). The correct recall of abnormal cases (sensitivity), the correct return to screen (specificity) and the percentage of correct hits of Area of Interest (AOI) were compared and shown as below:



Results

In terms of specificity the UK mean value was 78.94%, US 66.78% and EU 80.24%. In terms of sensitivity the UK mean values was 83.46%, US 82.18% and EU 74.05%. One-way ANOVA showed that there was no significant difference in sensitivity among the three groups ($p=n.s.$). However, there was a significant difference in specificity among those groups. Post hoc showed that there was a significant difference between UK and US ($p<0.001$) and a significant difference between UK and EU ($p<0.05$). There was no significant difference between the US and EU groups ($p=n.s.$). In addition, regarding the correct hits of the correct abnormality, ANOVA showed no significant difference among the three groups ($p=n.s.$).

Discussion

Data showed that it is perfectly feasible for radiologists from different nationalities to read and report the same test cases. The PERFORMS scheme offers this facility by making the clinical cases readily accessible to registered users who download the cases and then report on these using the PERFORMS online App. When the same test cases are examined by radiologists from different countries then they achieve broadly similar sensitivity scores. This is reassuring and reinforces the globality of radiological education. The main difference between the three groups was found for the US specificity scores which we suggest reflects the different clinical management of potential assessment cases as compared to European countries. It is proposed that this work demonstrates that the PERFORMS case sets and reporting App can be used successfully internationally to highlight group or individual performance difficulties and identify training needs for participants. This then can enhance their performance and help them target shortcomings as identified by the test sets.

Acknowledgements

We acknowledge the support of the NHSBSP and Public Health England.