Mammography interpretation: factors influencing the assessment of accuracy and the perception of abnormality

Warren Reed

School of Medical Radiation Sciences, Faculty of Health Sciences
The University of Sydney, Lidcombe, New South Wales, Australia
Corresponding author email: W.Reed@fhs.usyd.edu.au

Abstract This article examines the issues concerning the assessment of observer performance and the process by which abnormalities are perceived and reported upon in screening mammography. The assessment of observer accuracy and the perception of abnormality are discussed to gain some insight into the image interpretation process. It is surmised that, with further research into these areas, it may be possible to inform educators to adopt teaching strategies that can improve the detection of breast cancers in screening mammography.

Introduction

‘There are no facts, only interpretations’. So wrote the German philosopher Frederick Nietzsche. To ensure that clinical facts are accurately interpreted in screening mammography, stringent quality assurance procedures have been put in place so that the accuracy of subsequent diagnosis is paramount. However, the reason why some breast cancers are missed in the interpretation process is an area that warrants further exploration.

The purpose of this article is to examine the factors that influence the process by which observer accuracy is assessed in screening mammography and also how clinical facts are accurately perceived and communicated. This process entails the observer scanning information that is comprised of image-based elements. The observer then has to decide if they perceive this information to be normal or abnormal. Thereby, an interpretation of the image is gleaned from an immense amount of information present. This is conveyed as a summing up of the results of the process in the form of a reasoned mammography report.

Breast cancer carries one of the highest mortality rates for women of all cancers. The best outcomes for effective treatment occur when the breast cancer is detected early. The production of a mammography report is an essential factor in the diagnosis of breast cancer.

Although a mammogram cannot provide an unquestionable conclusion as to whether a woman has breast cancer, it can demonstrate if a breast abnormality is present. This abnormality can then be classified according to predetermined criteria. Additional imaging is very often the result of an abnormal appearance such as magnification views, spot views and ultrasound.

The assessment of observer accuracy

To be able to evaluate the interpretation process, precise tools must be designed to enable observer accuracy to be measured. There are two main factors that largely determine the diagnostic accuracy of a mammography examination. The first centres on the technical quality of the image in its ability to reveal important clinical information. This is most often dependent upon exposure and positioning. Second, the diagnostic accuracy very much depends upon the observer’s proficiency and competence to be able to correctly interpret the clinical information that is shown on the image.

The technical factors are often easier to assess and can be standardised through the use of protocols. However, in terms of observer accuracy, some researchers have examined the variability in the interpretation processes of individual observers. They have identified these differences as in need of more targeted research.

For the observer of mammograms, interpretation of the image is often hampered by the anatomical characteristics of the woman being screened. Accurate interpretation can be affected by the density of the breast tissue to be examined along with the size and shape of any potential lesions. In estimating observer performance in terms of accuracy, it has been shown that studies that use a small sample size (in the form of less examinations) become more variable and therefore more extreme. This is because the statistical models have a reduced scope to adjust for the difference in performance of each observer.

Among the mammography screening population, there is, proportionately, a low incidence of breast cancer that also makes the performance of an observer difficult to assess. As an example, the incidence of invasive breast cancer in the USA is only 3.5 in every 1000 women who are over 49 years old.

A method to attempt to counteract these problems in the estimation of observer accuracy is the use of standardised testing of observers. This alternative method allows each observer to interpret identical images. The observers can also use the same viewing conditions and will have access to the same clinical details. Images from women that lack follow up information can be excluded, consequently, the true disease state of the women involved will be known with a high degree of certainty. In the clinical environment, it is only when the observer detects an abnormality that the woman concerned will have a confirmative procedure in the form of a biopsy. If the observer identifies an abnormal image as normal, the abnormality will only present with obvious symptoms, or as a result of another screening examination. This means that, in the clinical setting, the observer cannot be certain about the woman’s true disease state. The use of standardised test sets enables the inclusion of a greater number of abnormal films than that which would normally occur in the clinical setting. The sensitivity of subsequent film related perfor-
The importance of the observer to interpret the image-based elements accurately into a coherent report is therefore essential. This interpretation can be based on endogenous (based on the observer) and exogenous (based on the image) factors. The endogenous elements incorporate the anatomical and pathological knowledge of the observer, along with how these will be represented in an image. Also included is the use of a systematic search strategy and a sound decision making process that can differentiate between true pathology and inevitable artefact caused by noise or tissue superimposition.

The comparative experience of the observer is also a factor as measured by the number of examinations the observer has previously reported upon. The performance of the observers positively correlated with their respective experience in the field, as measured by the area under the receiver operating characteristic curve. It is surmised that the more experienced observers have a greater relevant knowledge base from which to refer and are often able to detect less obvious, subtler changes in the image representation.

To gain insight into the decision-making process studies have measured what is termed ‘visual dwell’. This is measured by timing how long an observer views a particular area. It has been shown that 70% of missed lesions have a similar visual dwell time to the time spent viewing correct areas of abnormality. This would suggest that these lesions are not missed by the observer but are noticed and then dismissed as a normal variant. Additionally, changes in the position or shape of normal tissue seen in the image that are incorrectly identified as abnormal possess a visual dwell time that is as long as that for correctly reported cases.

The main problem, therefore, is not that there are errors in the basic search strategies of the observers. It has been shown that 70% of unreported abnormalities are errors of interpretation, while only 30% are due to inefficient visual search strategies. Other factors effecting accuracy in mammography

The detection of abnormality in screening mammography

Researchers have estimated that, in the USA, 10–30% of breast cancers go unreported during screening mammography. This may be partly due to the stated low incidence of breast cancers in the screening population. Also, a compromise must be established between not causing unnecessary alarm by over diagnosing and the crucial aspect of identifying and treating breast cancers at an early stage.

A significant retrospective study emphasised the need for further research when it determined that two-thirds of missed breast cancers were apparent upon subsequent review. There are newer technologies available which are designed to increase the observers perception of image based abnormalities. More discrete findings can be made through the use of image displays incorporating increased resolution. However, a study has shown that increasing the amount of information displayed in an image can actually negatively affect the performance of an observer. Also, computer-assisted diagnosis (CAD) acts as a prompting system through the use of algorithms to identify potential abnormalities on digital radiographs. Nevertheless, the impact of such a system on the performance of observers needs to be further researched in the way that it effects the decision making process. High sensitivity and specificity systems have been shown to aid in the detection of subtler breast cancers while low sensitivity and specificity systems adversely affect observer performance.

Another study compared one view (oblique) and two view (oblique and cranio-caudal) mammography in breast cancer screening. These authors concluded that performing two projections of the breast instead of one increases the detection of breast cancer by 24%. Additionally, two projections reduce the number of women recalled by 15% making it a cost effective alternative.

Conclusion

It is clear that some breast cancers go unreported, even though they are deemed visible in subsequent review. These errors may well have serious consequences. The error rate is higher in less experienced observers and also increases with insufficient clinical information. Design of data collection tools need to be planned carefully to assess observer accuracy, both in the ‘clinical’ and ‘standardised’ setting. Observer search strategies appear
adequate, but the incorrect diagnosis of normal is sometimes made in response to what is seen in retrospect to be an abnormal lesion. This may be due to inaccurate training resulting in diagnostic errors. This may also however be due to intrinsic factors associated with the observers themselves.

It has been said that accurate image interpretation is a difficult process involving the capability to merge information derived from visual pattern recognition, anatomical knowledge, knowledge of pathological processes, and patient specific information. Also stated is that there is little known concerning the method by which expert radiologists or students acquire these abilities. Researchers need to investigate the process by which observers gather clinical facts and interpret them. Observer behaviours that increase accuracy can then be analysed to suggest strategies where these attributes can be maximised.

It has been stated that image interpretation is a classic example of clinical problem solving and should be taught in a way that reflects professional practice. Insight into factors associated with success in radiological interpretation can then guide the way in which image interpretation is taught and assessed for accuracy. This would then significantly improve future abnormality detection in breast screening.

References