REVIEW

Errors in Mammography Cannot be Solved Through Technology Alone

Ernest Usang Ekpo*, Maram Alakhras, Patrick Brennan

Abstract

Mammography has been the frontline screening tool for breast cancer for decades. However, high error rates in the form of false negatives (FNs) and false positives (FPs) have persisted despite technological improvements. Radiologists still miss between 10% and 30% of cancers while 80% of woman recalled for additional views have normal outcomes, with 40% of biopsied lesions being benign. Research show that the majority of cancers missed is actually visible and looked at, but either go unnoticed or are deemed to be benign. Causal agents for these errors include human related characteristics resulting in contributory search, perception and decision-making behaviours. Technical, patient and lesion factors are also important relating to positioning, compression, patient size, breast density and presence of breast implants as well as the nature and subtype of the cancer itself, where features such as architectural distortion and triple-negative cancers remain challenging to detect on screening. A better understanding of these causal agents as well as the adoption of technological and educational interventions, which audits reader performance and provide immediate perceptual feedback, should help. This paper reviews the current status of our knowledge around error rates in mammography and explores the factors impacting it. It also presents potential solutions for maximizing diagnostic efficacy thus benefiting the millions of women who undergo this procedure each year.

Keywords: Mammography- radiographic image interpretation- cancer detection- diagnostic imaging- radiological errors

Introduction

Breast cancer is the most prevalent female cancer; it is the most frequent cause of cancer death in females in low and middle-resource countries and the second leading cause of cancer-related deaths in the developed world (Desantis et al., 2013; Ferlay et al., 2013; Jemal et al., 2011; Lauby-Secretan et al., 2015). Survival rates have improved due to improvement in early detection and treatment strategies (Njor et al., 2012). Early detection of the disease can be achieved through breast self-examination and clinical examination; however, some breast lesions are non-palpable and require visual assessment through imaging (Hou et al., 2002; Rosen et al., 1999; Tartter et al., 1999). Screening mammography is the frontline imaging tool for early detection and has been credited with 30% to 40% reduction in mortality from breast cancer (Lauby-Secretan et al., 2015; Njor et al., 2012; Paci, 2012). However, mammography may have some drawbacks such as false positive (FP) and false negative (FN) screening outcomes. Normal breast parenchymal perturbations on mammograms or benign lesions may mimic breast cancer, and may lead to wrong diagnosis of cancer when none is present (FP errors) (Castells et al., 2016). FP errors result in low positive

predictive value (PPV) (10%), psychological effects (Molina et al., 2016), and recall for further assessment (Vernacchia and Pena, 2009). On the other hand, cancer may be present in a mammogram but missed by breast readers and constitute a FN error (Evans et al., 2013). Some of the cancers missed on screening may be clinically and mammographically occult (Buchberger et al., 2000; Tartter et al., 1999) or demonstrate subtle radiographic features that are difficult to perceive (Banik et al., 2011). However, some of the missed cancers are visible on mammograms but are either not identified or are disregarded by breast readers (Evans et al., 2013; Maxwell, 1999). Such FN errors account for 1.3% to 45% of missed cancers, with error rates determined by the subtype and characteristics of the cancer (Banik et al., 2011; Evans et al., 2013; Maxwell, 1999). The literature demonstrates considerable variation in the error levels (Evans et al., 2013; Maxwell, 1999), which result from differences in the physical characteristics of the population being screened (Ekpo et al., 2015; Mandelson et al., 2000), technological and technical factors during image acquisition (Holland et al., 2016; Popli et al., 2014), the subtypes and radiographic features of breast lesions (Bird et al., 1992; Burrell et al., 2001;Suleiman et al., 2016a), and the characteristics of breast imaging readers and reading conditions (Brady et

*Discipline of Medical Radiation Sciences, Faculty of Health Sciences, University of Sydney, Sydney, Australia. *For Correspondence:: ernest.ekpo@sydney.edu.au*

al., 2012; Kerlikowske et al., 1998).

The basis of detection and interpretative errors appears to be a combination of human and technological limitations, however this is not well understood. We urgently need to address this gap in knowledge so that reliable and accurate screening strategies can be developed using more effective technology and education. An understanding of how patient, technical, lesion, and reader-related factors impact upon mammography errors and ways of mitigating them may be the key to improved early cancer detection and further reduction in mortality from breast cancer. Therefore, this paper explores factors responsible for mammography errors including characteristics of the screening population, types and radiographic features of breast lesions, technological and technical factors, characteristics of breast imaging readers, and reading environment. It also examines the types of errors made and potential solutions to mitigate these errors.

1. Factors influencing mammographic performance

This section examines factors affecting the outcome of mammographic image interpretation. These include characteristics of the screening population, lesion characteristics as well as technological, technical and reader related factors and reading conditions.

1.1 Characteristics of the screening population

Patients' physical characteristics such as body habitus, breast density (Ekpo et al., 2015; Freer, 2015), use of hormone replacement therapy (HRT) (Banks et al., 2006; Carney et al., 2003; Kavanagh et al., 2005), breast augmentation and implant, as well as disease prevalence (Fajardo et al., 1995; Handel, 2007; Miglioretti et al., 2004) may impact upon the detection of breast cancer using mammography.

Patients' body habitus such as body mass index (BMI) and breast size have been shown to affect the performance of mammography (Bassett et al., 2000; Elmore et al., 2004; Hunt and Sickles, 2000; Njor et al., 2016; Popli et al., 2014). These two characteristics are inter-related, with overweight (BMI: 25 – 29.9), and obese (BMI: ≥30) women demonstrating larger breasts (Brown et al., 2012). A 3.0% to 38% increase in mammographic sensitivity has been reported for overweight women (BMI: ≥24.9) compared with those who have normal BMI (Banks et al., 2004; Njor et al., 2016), however, no significant differences in recall rates and specificity between these categories of women have been reported (Banks et al., 2004; Elmore et al., 2004; Njor et al., 2016). There are various potential causal agents with regards to breast size which may affect cancer detection; the small-sized imaging plates used in mammography systems has been shown to pose difficulties in positioning large floppy breasts, and often negatively affect compression, compromise image sharpness, and increase blur (Bassett et al., 2000; Popli et al., 2014). Inadequate positioning and compression of large breasts increases tissue superimposition and non-uniformity, and imaging large breasts on small-sized imaging plates may result to incomplete coverage of breast regions such as the posterior portions and pectoral muscles

(Bassett et al., 2000; Popli et al., 2014). Together, these negative confounders may increase the chances of errors during mammography interpretation.

Breast composition also influences the difficulty of cancer detection in mammography. Breast density (the proportion of the breast composed of fibroglandular tissue) and solid breast cancers are mammographically radiodense (Ekpo et al., 2015). Dense tissue produces a masking effect, which reduces the ability to visualize cancer in mammograms (Ekpo et al., 2015; Mandelson et al., 2000; Pisano et al., 2008). Studies have consistently demonstrated lower sensitivity in dense breasts, ranging from 27% to 70.1% compared to 90% in fatty breasts (Carney et al., 2003; Mandelson et al., 2000; Pisano, et al., 2008; Rosenberg et al., 1998). A recent study has reported an association between breast density and recall for additional examination (Ekpo et al. 2016a). It has also been shown to account for 16% higher incidence of interval breast cancer relative to fatty breasts (Boyd et al., 2014). Breast density is inversely related to BMI (Ekpo, 2016a), therefore it is unsurprising that women with low BMI demonstrate 3.0% to 38% reduced sensitivity with mammography as discussed previously (Banks et al., 2004; Njor et al., 2016).

Age is another factor that affects cancer detection in mammography. Mammography is performed with women standing erect, a position that is difficult to maintain by frail elderly women especially with the pain from compression. This can cause inadequate positioning and motion blur, reducing the visibility of image details and reader ability to detect microcalcifications (Abdullah et al., 2017; Popli et al., 2014; Rosen et al., 2002). The composition of the breast changes with age, with women younger than 50 years demonstrating high breast densities compared to older women (McCormack and dos Santos, 2006). This age-related difference in breast composition has been shown to be concomitant with mammography performance (Carney et al., 2003; Rosenberg et al., 1998). The reported sensitivity of mammography for women younger than 50 years varies from 54.0% to 78.0% compared with 78% to 85% in women aged 70 years and older (Carney et al., 2003; Houssami et al., 2003; Keen and Keen, 2008). This is reasonable given that the breast undergoes atrophic changes with ageing and becomes replaced by fat, which appears radiolucent relative to cancer lesions thereby enhancing the visibility of these lesions on mammograms. On the other hand, masking effect produced by dense tissue reduces the visibility of cancer in mammograms of younger women (Boyd et al., 2007). The limitation of mammography in younger women has given rise to recommendations for use of other imaging modalities such as ultrasound for imaging this category of women (Devolli-Disha et al., 2009; Houssami et al., 2003).

Exogenous hormonal agents play a critical role in breast tissue changes as demonstrated by the variations in the mammographic appearance of breast parenchyma between users and non-users of hormonal substances (Buist et al., 2004). HRT use has been shown to lower mammography performance by increasing FP recall (Banks et al., 2006). The literature shows that HRT use is associated with 7%-22% and 12%-50% reductions in

mammographic sensitivity and specificity respectively (Banks et al., 2006; Carney et al., 2003; Kavanagh et al., 2005). Reduced mammography performance has been shown to be particularly pronounced in users of HRT regimens containing estrogen and progesterone (Banks et al., 2006), which can be attributed to their effect on increasing breast density (Buist et al., 2004). This is further supported by the fact that hormonal agents such as tibolone, which do not increase breast density (Ekpo, 2016b), also do not affect mammography performance (Banks et al., 2006). HRT use is also associated with a higher risk of benign breast disease (Rohan and Miller, 1999), a determinant of FP mammography outcome and risk of subsequent cancer (Castells et al., 2013). Thus, the impact of hormonal agents on mammography performance should be considered when referring patients for screening and when interpreting their mammograms in order to minimise FP recall.

Breast implants and augmentation materials such as silicon have been shown to affect mammographic performance and can lead to omission errors, mainly because of their high X-ray attenuation coefficient and opaque appearance on mammograms (Handel, 2007). It has been reported that breast augmentation and implants produce capsular contraction that obscure the visualization of breast tissue by 15% to 88% depending on the size and opacity of the material (Handel et al., 1992; Silverstein et al., 1991). They have been shown to limit the detection of breast cancer in mammograms by about 22% (Handel 2007; Miglioretti et al., 2007). Augmentation mammoplasty is also associated with scarring and distortion of the breast parenchyma (Handel 2007; Hayes et al., 1988), and the opacities created by these scars may mimic malignant calcifications and subtle cancer types such as architectural distortion (AD) mammographically (Handel 2007; Silverstein et al., 1991). Previous breast conservative surgery alters breast architecture (Piroth et al., 2014), and has been shown to reduce mammography sensitivity by 9.1 – 10% (van Breest et al., 2012a; van Breest et al., 2012b). Such surgical procedures are also associated with FP errors (Holli et al., 1998), with FP rates doubling in women who have had post-surgical radiotherapy (Holli et al., 1998). Together, patients' physical characteristics discussed above have the potential to limit adequate visualization of breast parenchyma for features of cancer. They can also cause parenchyma perturbations that may increase the likelihood of FP errors.

The prevalence of disease in the population may also impact upon interpretive performance by affecting reader expectation and influencing their search, perceptual and decision-making behavior and confidence (Evans et al., 2013; Gur et al., 2007; Reed et al., 2011). Studies have shown that low disease prevalence lowers readers' level of concentration, and that normal images tend to attract more scrutiny and fixations at higher disease prevalence (Evans et al., 2013; Gur et al., 2007; Reed et al., 2011). Evidence shows that with increased prevalence, visual search increases and confidence that a normal image is in fact normal decreases, however confidence for abnormals remains unchanged (Fanshawe et al., 2016; Gur et al., 2007; Reed et al., 2011). Therefore, high disease prevalence may increase FP errors and recall rates in a screening scenario.

1.2 Types and radiographic features of breast lesions

The variability in lesion morphology and its effect on the heterogeneity of the breast parenchyma makes cancer detection and characterization challenging (Bird et al., 1992; Popli, 2001). Lesion characteristics and mammographic appearances that have been shown to impact upon radiographic detection and characterization include size, shape, density, margins and subtlety (Bird et al., 1992; Burrell et al., 2001; Mello-Thoms et al., 2014). Lesion location and its impression on adjacent breast tissues is also an important factor affecting cancer detection (Bird et al., 1992; Burrell et al., 2001; Mello-Thoms et al., 2014).

Small-sized lesions have been shown to be more difficult to detect than larger ones (Malich et al., 2003; Mello-Thoms et al., 2014). Even when lesions are clearly visible, their shape, margins and density would determine their classification as benign or malignant (Popli, 2001). Radiologically, round and oval masses with fatty or low-fat content and well-defined borders are associated with benign conditions (Popli, 2001). Isodense masses with lobulated, obscured and ill-defined or indistinct margins are classified as suspicious (James et al., 2010; Lee et al., 2014; Popli, 2001). Highly suspicious lesions are of high density, irregular in shape, with spiculated, ill-defined or indistinct margins (James et al., 2010; Lee et al., 2014; Popli, 2001). Masses demonstrating these characteristics are easily detected (Bird et al., 1992), however, about 10% of malignant lesions demonstrate benign features (round, oval, well-defined), and sometimes spiculations and parenchymal changes induced by malignant masses may be subtle and difficult to perceive (Bird et al., 1992; Roberts-Klein et al., 2011). These scenarios may lead to potentially malignant lesions being overlooked or misinterpreted, with wrong interpretation accounting for 52% of errors in mammography (James et al., 2010; Lee et al., 2014).

Previous works showed that cancer subtype and characteristics influence the difficulty of detection and characterization with mammography, with triple-negative breast cancers and invasive lobar carcinoma more difficult to detect on screening, yet constituting the most common subtype of interval cancer (Caldarella et al., 2013; Domingo et al., 2010; Hoff et al., 2012; Johnson et al., 2015; Lowery et al., 2011; Raposo et al., 2012; Rayson et al., 2011; Sung et al., 2016). Lesion subtlety is also a determinant of cancer detection, and contributes to 43% of mammographically missed cancers (Bird et al., 1992). Subtle masses such as architectural distortion (AD) are particularly difficult to detect (Gaur et al., 2013; Suleiman et al., 2016a) or characterize, which can be due to the plethora of conditions associated with its radiographic features. In mammograms, AD features may be due to malignancies such as invasive lobar carcinoma and ductal carcinoma in-situ or benign conditions such as radial scars, previous surgery, trauma, sclerosing adenosis, infection, and fat necrosis (Gaur et al., 2013). Lesion features such as non-specific or asymmetric densities, isodensity

to fibroglandular breast tissue, indistinct margins, and absence of calcifications and ductal dilatations have also been shown to account for missed cancers (Hoff et al., 2012; Majid et al., 2003). Although the visibility of these features has improved in the digital era, they have low PPV and may be overlooked (Hoff et al., 2012). Thus, cancer subtype and subtlety, as well as the low PPV associated with the features described above may limit perception or reporting of perturbations produced by malignancy. This emphasises the need for human and technological interventions to ensure early detection of these missed signs.

1.3 Impact of technical factors, image quality, and mammography technology on image interpretation

The conspicuity of lesions in their background depends on the technical and radiographic quality of the images produced (Ekpo et al., 2014). Technical factors such as positioning, compression and exposure conditions affect the amount of breast tissues imaged, tissue spread and the visibility of lesion features (Holland et al., 2016). The difficulty of positioning large floppy breasts and older women and their impact on the accuracy of image interpretation have been discussed in section 1.1 (Bassett et al., 2000; Popli et al., 2014). Inadequate positioning is 26% more common in women with BMI ≥30 compared to those having normal BMI (Guertin et al., 2014), and mammograms with inadequate positioning, image quality and compression have been shown to demonstrate 18% lower sensitivity compared to those of good overall quality (Taplin et al., 2002). A study has also reported varying PPV at different degrees of compression, with highest compression demonstrating reduced lesion detection (Holland et al., 2016). All these indicate that technical factors contributing to missed cancers must be addressed.

Advances in imaging technology have led to the transition from screen-film mammography (SFM) to DM. Although SFM has better spatial resolution compared to DM, it has lower contrast resolution (Faridah, 2008). Whilst spatial resolution is relevant to the detection of high spatial frequency lesions such as microcalcifications, the higher contrast resolution of digital systems allows better differentiation of densities as well as normal and diseased tissues on an image (Faridah, 2008). In addition, post-processing capabilities of DM offers opportunities to manipulate image contrast to suit a particular detection task. However, studies comparing the diagnostic performance of SFM and DM have generated conflicting results (Bluekens et al., 2012; Hambly et al., 2009; Pisano et al., 2005; Pisano et al., 2008; Skaane, 2009), with many demonstrating comparable or slightly better cancer detection performance of DM in all breast compositions (Bluekens et al., 2012; Hambly et al., 2009; Pisano et al., 2005; Skaane, 2009), albeit with higher FP recalls (Hambly et al., 2009; Skaane and Skjennald, 2004). Despite important advances in mammography technology, the sensitivity of DM is still below optimal levels and varies between readers (Clauser et al., 2016; Pisano et al., 2005; Pisano et al., 2008). It is clear that improvement in imaging technology is not the only solution for removing detection errors, and instead we must identify and remedy

human errors limiting breast cancer detection, if early breast cancer diagnosis is to be transformed.

1.4 Reader characteristics and interpretative performance

Evidence shows wide inter-reader differences in cancer detection with mammography, suggesting that humans are a major determinant of mammography performance (Evans et al., 2013; Jackson et al., 2015; Maxwell, 1999). It also suggests that reader characteristics such as experience, specialization, number of mammograms read per year, and other factors may influence error rates. Of these, number of years reading mammograms and the number of cases read per year are the most widely studied observer characteristics (Suleiman, 2016b). Experience is quantified using parameters such as specialization in breast radiology, number of years of reading mammograms, and number of mammograms read per year (Rawashdehet al., 2013). These factors considered in isolation have generated contradictory outcomes. For example, Sickles et al (Sickles et al., 2002) reported a two-fold higher cancer detection rate for specialist radiologists compared to general radiologists, but this was not always consistent in other studies comparing radiologists and radiographers (Debono et al., 2015; Torres-Mejía et al., 2015). Whilst number of years of reading mammograms has been shown to improve performance in some studies (Rawashdehet et al., 2013; Suleiman, 2016b), it has demonstrated no significant effect on observer performance in other studies (Jackson et al., 2015; Suleiman et al., 2014a).

The number of mammograms read per year has been reported as a good indicator of observer performance, and a potential optimisation tool for breast screening programs (Rawashdeh et al., 2013; Suleiman, 2016b). However, other studies have reported either no relationship or an inverse relationship between number of cases read per year and performance (Beam et al., 2003; Molins et al., 2008). It has also been shown that beyond a threshold annual volume read, performance stagnates (Rawashdehet et al., 2013) or begins to decline (Kan et al., 2000), suggesting that the impact of volume read per year on performance may be threshold dependant. Evidence suggests that the contradictory evidence for the impact of volume read on performance may be due to differences in other reader characteristics such as experience, hours spent reading mammograms per week, and importantly, the ability to identify normal image features (specificity) (Rawashdeh et al., 2013). Other confounding factors are practice-related and include double reading practices and number of diagnostic and interventional procedures performed (Beam et al., 2003), with radiologists who participate in their own diagnostic mammography and have higher volumes of work-ups demonstrating better outcomes compared to those who do not (Buist et al., 2014). Despite the conflicting evidence for the number of mammogram cases read per year on reader performance, it is used as a criterion for certification in many countries (Rawashdeh et al., 2013). For example, certification for mammography reporting in the USA requires 960 cases read biannually, whilst in Australia and European countries, it is 2000 and 5000 cases read per year respectively (Suleiman, 2016b). However, these

differences do not reflect inter-country variations in cancer detection performance by radiologists (Suleiman et al., 2014b; Suleiman, 2016b). In fact, a recent study showed no difference in performance between Australian and USA radiologists for breast cancer detection (Suleiman et al., 2014a), suggesting that parameters other than volume read per year may also be key determinants of performance.

The findings above demonstrate the complexity of using volume read and number of years of reading mammograms to quantify experience. Although radiologists receive training in breast image interpretation, they are exposed to different levels of mentorship, disease prevalence, and working conditions during the course of practice (Beam et al., 2003). These differences may impact differently on expertise and consequently mammography interpretation performance (Beam et al., 2003). Available evidence emphasise the need for programs and interventions that expose readers to continuous training and mentorship. It also underscores the need for platforms to identify errors, provide immediate feedback, and identify ways of mitigating these errors.

1.5 Impact of reading environment and workload on accuracy of image interpretation

Image interpretation requires both perceptual and cognitive processes, which can be affected by ambient lighting and distractions in the reading environment. Ambient light increases reflection (diffuse and specular) and glare, however no study has reported a significant difference in observer performance between low and moderate ambient light conditions (Chawla and Samei, 2007; Pollard et al., 2009; Pollard et al., 2012). Studies have shown that distractions may lead to attentional deficit and affect radiologist's perceptual and cognitive functions. For example, distractions due to phone call has been reported to account for a 12% discrepancy error among radiology residents (Balint et al., 2014), and negatively affects task completion time (Williams and Drew, 2017).

The number of patients undergoing screening mammography daily has increased exponentially, resulting in increased workload for radiologists. This condition is exacerbated by the increasing volume of radiological data generated with the advent of 3D imaging modalities. The number of hours spent reading these images may cause fatigue and occulomotor strain and reduce image interpretation accuracy (Krupinski et al., 2010a). Studies have reported a significant reduction in image interpretation accuracy after prolonged periods of reporting (Krupinski, 2010a; Krupinski et al., 2010b; Krupinski et al., 2012), for example a recent study reported that fatigue contributed to satisfaction of search (Krupinski et al. 2012). Thus, reading conditions contribute to radiological errors and need to be optimised to mitigate missed cancers.

1.6 Impact of patient's clinical history on accuracy of image interpretation

There are contentions that the availability of clinical history at the time of image interpretation may lead to biases and cognitive heuristics such as anchoring (locking on to salient evidence early in the interpretation

process), availability (making biased judgments based on what frequently comes to mind), confirmation (looking for evidence to confirm a particular disease), representativeness (decision-making based on similarity to mental prototype), and search satisficing (abrupt termination of search after identification of irrelevant features) (Crowley et al., 2013). However to date, studies have reported at least better (Berbaum et al., 1988; Doubilet and Herman, 1981; Leslie et al., 2000; Loy and Irwig, 2004) or at worst no effect on diagnostic performance when good clinical information is provided (Cooperstein et al., 1990; Good et al., 1990).

2. Common human errors limiting mammography interpretation

Radiographic image interpretation involves evaluation and organisation of image information to make a diagnostic decision. These processes can be challenging, particularly with mammography, due to the heterogeneity of the breast parenchyma, anatomical noise arising from dense tissue masking (Ekpo et al., 2015), and the subtlety of some breast cancer types (Gaur et al., 2013). In other situations, the characterization of detected lesions can be difficult and depends on the presenting features of the lesion and the knowledge of the image reader as discussed earlier (Bird et al., 1992; Roberts-Klein et al., 2011). These factors may either cause lesions to be concealed or conspicuous and fixated cancer lesions to be ignored (Bird et al., 1992; Roberts-Klein et al., 2011). The interpretation process involves search, perception and decision-making (Berlin, 2014; Brady et al., 2012; Brunoet al., 2015; Pinto and Brunese, 2010). Faults in any of these processes results in an interpretative error.

2.1 Search errors

Search errors ensue from inadequate scanning of the image, resulting in non-fixation (no visual attention or dwell) on the perturbations produced by cancer (Brady et al., 2012). Search errors may also arise from premature termination of search due to identification of stimuli elicited by another disease condition, which may be irrelevant and unconnected to the disease of concern, a situation referred to as "satisfaction of search (SOS)" (Fleck et al., 2010). Search errors have been estimated to account for 42% of error in DM (Palazzetti et al., 2016; Pinto and Brunese, 2010), and vary with readers' experience, workload and fatigue (Berbaum, 2010; Berlin, 2014; Nodine et al., 1996).

2.2 Perceptual errors

Sometimes malignant lesions briefly (<0.48 seconds) fixated by radiologists go unreported and constitutes a perceptual error (Bruno et al., 2015; Krupinski, 2010c). Perceptual errors account for 31% of errors in DM (Krupinski, 2010; Palazzetti et al., 2016; Samei, 2010.) and may be caused by insufficient stimuli by the lesion(s) and the nature of its background (Mello-Thoms, 2006). Perceptual error rates vary between radiologists and are inversely related to experience and availability of adequate clinical information. A report suggests that perceptual errors may be due to poor pattern recognition skills

(Krupinski, 2010c).

2.3 Decision-making errors

Decision-making errors occur when the region containing the lesion is fixated for a prolonged period (>0.48 seconds), but the lesion is misidentified (Bruno, et al., 2015). Decision-making errors account for 37% of errors in DM (Krupinski, 2010c; Nodine et al., 2002; Samei, 2010.), and may be due to poor reader's knowledge of the radiographic features of the disease and poor reader judgment (Mello-Thoms, 2006; Samei, 2010.). Other causal factors include fatigue, absence of prior images and inadequate clinical history (Krupinski, 2010c; Nodine et al., 2002). Decision-making errors are also influenced by recall rate recommendations, with lower recall rates associated with higher specificity and lower sensitivity values and vice versa (Norsuddin et al., 2016). A recent study showed that non-specific densities and AD lesions classified as malignant at free recall were dismissed at policy-driven fixed (target) recall rates (Norsuddin et al., 2016). It is therefore important that factors contributing to these errors are redressed.

3. Possible strategies to reduce errors in mammography interpretation

Despite advances in mammography technology and image quality improvement in the digital era, limitations remain around diagnostic efficacy. Therefore human factors affecting the accuracy of image interpretation must be remedied to improve the detection of missed cancers. Potential solutions include optimisation of technical and display parameters, adoption of double reading strategy, and technological and educational interventions.

3.1 Optimisation of technical and display parameters

Technical parameters that affect cancer detection such as positioning and compression need to be optimised to ensure adequate visibility of breast tissue. It is important that radiographers adopt correct positioning to ensure nipple alignment and inclusion of pectoral muscles (Taplin et al., 2002). Appropriate compression force should be applied to uniformly spread breast tissue and enhance the visibility of breast parenchyma and subtle cancers (Holland et al., 2016). Display tools and reading environment should be optimised to reduce reflection, glare, and reader fatigue (Krupinski et al., 2010b; Waite et al., 2016). Also, population-based mammography-screening programs must carefully monitor the technical quality of mammograms on a regular basis to ensure that they are adequate for purpose.

3.2 Double reading strategy

Many screening programs including the Dutch Nationwide Breast Cancer Screening Program and BreastScreen Australia employ an independent double reading strategy with arbitration to improve cancer detection rates. This strategy accounts for the differences in human perceptual and cognitive abilities and has been shown to improve cancer detection by between 5.6 – 15% (Duijm et al., 2007; Duijm et al., 2008). Studies

have explored changing the order of reading, where the second reader interprets a batch of mammograms in an order opposite that of the first reader to overcome vigilance diminution, however this has not been shown to improve the effectiveness of double reading strategy (Taylor-Phillips et al., 2016; Taylor-Phillips et al., 2014). It should be acknowledged however that double reading increases financial cost (Posso et al., 2016), has wide inter-reader disagreement, and is time consuming (Ekpo et al., 2016c; Redondo et al., 2012) and therefore to reduce the limitations of cost and time, computer systems have been designed to act as a second reader (Taylor and Potts, 2008).

3.3 Use of computer-aided devices (CAD)

Technological innovations such as computer-aided detection (CAD) have been explored to mitigate perceptual errors. These devices use computerized algorithms to highlight perturbations in the image (CADe) or perform diagnostic assessment (CADx). Whilst CAD has been shown to increases sensitivity (Georgian-Smith et al., 2007; Gromet, 2008; Karssemeijer et al., 2003; Skaane et al., 2007; Taylor and Potts, 2008) there is contrasting evidence for the relative impact of double reading versus single reading with CAD (Georgian-Smith et al., 2007; Gromet, 2008; Karssemeijer et al., 2003; Skaane et al., 2007). The literature generally supports double reading to outperform single reading with CAD (Taylor and Potts, 2008) A major limitation of CAD is its high FP rates (Philpotts, 2009), which increases at lower dose levels (Wittenberg et al., 2011). As a result, some of the malignant lesions marked by CAD may be dismissed by readers. This suggests that such technology is not necessarily the complete solution for removing detection errors, and emphasises the need for educational and practical interventions to improve human perceptual and decision-making skills.

3.4 Audit and immediate feedback mechanisms

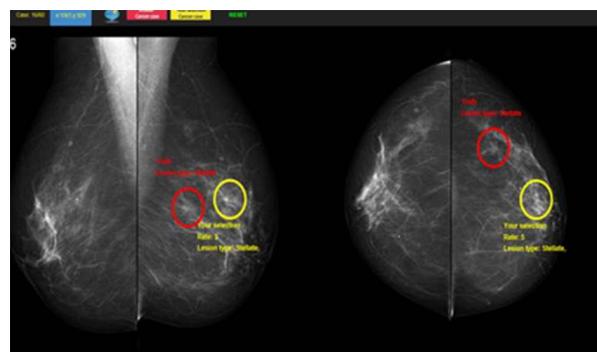Clinical audits are used to evaluate the performance of

Figure 1. Breast Software Display Showing Reader's Mark and Lesion Classification (Yellow) against the Actual Lesion Location and Type (Red). The BREAST tool provides immediate feedback to readers on cancers that present perceptual difficulties and describes their features such as stellate, discrete, architectural distortion and nonspecific density. Radiologists from any part of the world can log unto the system, undertake self-assessment and obtain immediate feedback.

any program and aim to identify errors in order to tailor interventions to improve performance. However, clinical audits often take a longer period to complete, delaying feedback. Also, feedback from audits is often provided to the screening program and not the individual breast reader, making it difficult for individuals to identify their errors and take corrective actions. Perceptual feedback has been shown to improve reader performance (Buist et al., 2011; Donovan et al., 2008; Krupinski et al., 1993). Therefore, platforms that audit and provide immediate perceptual feedback to readers should be explored to enhance interpretative accuracy. Examples of platforms that have been established for this purpose include Breast Reader Assessment Strategy (BREAST) and PERsonal perFORmance in Mammographic Screening (PERFORMS) (Scott and Gale, 2006; Suleimanet al., 2016c). These platforms generate mammographic test-sets containing various cancer types and radiographic features that have been missed by at least one radiologist during clinical reporting. These test-sets are hosted online enabling radiologists to evaluate and receive immediate feedback on location and type (stellate, mass, non-specific density, and AD) of lesion present in the mammogram.

In so doing, they examine reader performance, identify errors, and provide immediate feedback and continuous professional development opportunities. A recent study reported improvement in cancer detection performance of radiologists undertaking BREAST intervention over time regardless of their levels of experience (Suleimanet al., 2016c). However, more studies comparing radiologist test reading performance versus their clinical performance are needed to confirm the clinical impact of these interventions. If successful, such platforms can provide e-Learning opportunities and allow personalized learning and self-assessment, identify common mammography errors, and provide feedback to mitigate these errors.

In conclusion, errors in mammography interpretation arise from patient, lesion, technical, and reader factors as well as other extraneous variables such as distraction and fatigue. Evidence shows that, despite improvement in imaging technology, the accuracy of image interpretation still suffers from intrinsic human limitations. Therefore, technology alone cannot mitigate radiological errors, suggesting that if the benefits of breast screening are to be maximized, human errors limiting early cancer detection need to be identified and remedied. Double reading and immediate feedback loops may facilitate discussions and learning opportunities to improve diagnostic efficacy.

References

Abdullah AK, Thompson JD, Mercer CE, et al (2017). The impact of simulated motion blur on lesion detection performance in full field digital mammography. *Br J Radiol*, **90**, 20160871.

Balint BJ, Steenburg SD, Lin H, et al (2014). Do telephone call interruptions have an impact on radiology resident diagnostic accuracy?. *Acad Radiol*, **21**, 1623-8.

Banik S, Rangayyan RM, Desautels JE (2011). Detection of architectural distortion in prior mammograms. *IEEE Trans Med Imaging*, **30**, 279-94.

Banks E, Reeves G, Beral V, et al (2004). Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. *BMJ*, **329**, 477.

Banks E, Reeves G, Beral V, et al (2006). Hormone replacement therapy and false positive recall in the Million Women Study: patterns of use, hormonal constituents and consistency of effect. *Breast Cancer Res*, **8**, R8.

Bassett LW, Farria DM, Bansal S, et al (2000). Reasons for failure of a mammography unit at clinical image review in the American college of radiology mammography accreditationprogram. *Radiology*, **215**, 698-702.

Beam CA, Conant EF, Sickles EA (2003). Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst*, **95**, 282-90.

Berbaum KS, el-Khoury GY, Franken EA Jr, et al (1988). Impact of clinical history on fracture detection with radiography. *Radiology*, **168**, 507-11.

Berbaum K, Franken EA, Caldwell R, Schartz K (2010). Satisfaction of search in traditional radiographic imaging. In K. E. Samei E (Ed.), The handbook of medical image perception and techniques, Cambridge: Cambridge university press. pp 107–38.

Berlin L (2014). Radiologic errors, past, present and future. *Diagnosis*, **1**, 79.

Bird RE, Wallace TW, Yankaskas BC (1992). Analysis of cancers missed at screening mammography. *Radiology*, **184**, 613-7.

Bluekens AMJ, Holland R, Karssemeijer N, Broeders MJM, Heeten GJD (2012). Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: A multicenter study. *Radiology*, **265**, 707-14.

Boyd NF, Guo H, Martin LJ, et al (2007). Mammographic density and the risk and detection of breast cancer. *N Engl J Med*, **356**, 227-36.

Boyd NF, Huszti E, Melnichouk O, et al (2014). Mammographic features associated with interval breast cancers in screening programs. *Breast Cancer Res*, **16**, 417.

Brady A, Laoide RO, McCarthy P, McDermott R (2012). Discrepancy and error in radiology: concepts, causes and consequences. *Ulster Med J*, **81**, 3-9.

Brown N, White J, Milligan A, et al (2012). The relationship between breast size and anthropometric characteristics. *Am J Hum Biol*, **24**, 158-64.

Bruno MA, Walker EA, Abujudeh HH (2015). Understanding and confronting our mistakes: The epidemiology of error in radiology and strategies for error reduction. *Radio Graphics*, **35**, 1668-76.

Buchberger W, Niehoff A, Obrist P, DeKoekkoek-Doll P, Dunser M (2000). Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. *Semin Ultrasound CT MR*, **21**, 325-36.

Buist DS, Porter PL, Lehman C, Taplin SH, White E (2004). Factors contributing to mammography failure in women aged 40-49 years. *J Natl Cancer Inst*, **96**, 1432-40.

Buist DS, Anderson ML, Haneuse SJ, et al (2011). Influence of annual interpretive volume on screening mammography performance in the United States. *Radiology*, **259**, 72-84.

Buist DS, Anderson ML, Smith RA, et al (2014). Effect of radiologists' diagnostic work-up volume on interpretive performance. *Radiology*, **273**, 351-64.

Burrell HC, Evans AJ, Wilson AR, Pinder SE (2001). False-negative breast screening assessment: what lessons can we learn?. *Clin Radiol*, **56**, 385-8.

Caldarella A, Puliti D, Crocetti E, et al (2013). Biological characteristics of interval cancers: a role for biomarkers in the breast cancer screening. *J Cancer Res Clin Oncol*, **139**, 181-5.

Carney PA, Miglioretti DL, Yankaskas BC, et al (2003). Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med*, **138**, 168-75.

Castells X, Román M, Romero A, et al (2013). Breast cancer detection risk in screening mammography after a f*alse-positive* result. *Cancer Epidemiol*, **37**, 85-90.

Castells X, Tora-Rocamora I, Posso M, et al (2016). Risk of breast cancer in women with false-positive results according to mammographic features. *Radiology*, **280**, 379-86

Chawla AS, Samei E (2007). Ambient illumination revisited: a new adaptation-based approach for optimizing medical imaging reading environments. *Med Phys*, **34**, 81-90.

Clauser P, Nagl G, Helbich TH, et al (2016). Diagnostic performance of digital breast tomosynthesis with a wide scan angle compared to full-field digital mammography for the detection and characterization of microcalcifications. *Eur J Radiol*, **85**, 2161-8.

Cooperstein LA, Good BC, Eelkema EA, et al (1990). The effect of clinical history on chest radiograph interpretations in a PACS environment. *Invest Radiol*, **25**, 670-4.

Crowley RS, Legowski E, Medvedeva O, et al (2013). Automated detection of heuristics and biases among pathologists in a computer-based system. *Adv Health Sci Educ Theory Pract*, **18**, 343-63.

Debono JC, Poulos AE, Houssami N, Turner RM, Boyages J (2015). Evaluation of radiographers' mammography screen-reading accuracy in Australia. *J Med Radiat Sci*, **62**, 15-22.

Desantis C, Ma J, Bryan L, Jemal A (2013). Breast cancer statistics. *CA Cancer J Clin*, **64**, 52-62.

Devolli-Disha E, Manxhuka-Kerliu S, Ymeri H, Kutllovci A (2009). Comparative accuracy of mammography and ultrasound in women with breast symptoms according to age and breast density. *Bosn J Basic Med Sci*, **9**, 131-6.

Domingo L, Sala M, Servitja S, et al (2010). Phenotypic characterization and risk factors for interval breast cancers in a population-based breast cancer screening program in Barcelona, Spain. *Cancer Causes Control*, **21**, 1155-64.

Donovan T, Manning DJ, Crawford T (2008). Performance changes in lung nodule detection following perceptual feedback of eye movements. in Proceedings of SPIE. *Int Soc Opt Engin*, **20**, 6917

Doubilet P, Herman PG (1981). Interpretation of radiographs: effect of clinical history. *Am J Roentgenol*, **137**, 1055-8.

Duijm LE, Groenewoud JH, Fracheboud J, de Koning HJ (2007). Additional double reading of screening mammograms by radiologic technologists: impact on screening performance parameters. *J Natl Cancer Inst*, **99**, 1162-70.

Duijm LE, Groenewoud JH, Fracheboud J, et al (2008). Introduction of additional double reading of mammograms by radiographers: effects on a biennial screening programme outcome. *Eur J Cancer*, **44**, 1223-8.

Ekpo EU, Hoban AC, McEntee MF (2014). Optimisation of direct digital chest radiography using Cu filtration. *Radiography*, **20**, 346-50.

Ekpo EU, Hogg P, Highnam R, McEntee MF (2015). Breast composition: Measurement and clinical use. *Radiography*, **21**, 324-33.

Ekpo EU, Egbe NO, Egom AE, McEntee MF (2016a). Mammographic breast density: Comparison across women with conclusive and inconclusive mammography reports. *J Med Imaging Rad Sci*, **47**, 55-9.

Ekpo EU Brennan PC, Mello-Thoms C, McEntee MF (2016b). Relationship between breast density and selective estrogen-receptor modulators, aromatase inhibitors, physical activity, and diet: A systematic review. *Integr Cancer Ther*, **15**, 127–44.

Ekpo EU, Ujong UP, Mello-Thoms C, McEntee MF (2016c). Assessment of interradiologist agreement regarding mammographic breast density classification using the fifth edition of the BI-RADS atlas. *AJR Am J Roentgenol*, **206**, 1119-23.

Elmore JG, Carney PA, Abraham LA, et al (2004). The association between obesity and screening mammography accuracy. *Arch Intern Med*, **164**, 1140-7.

Evans KK, Birdwell RL, Wolfe JM (2013). If you don't find it often, you often don't find it: why some cancers are missed in breast cancer screening. *PLoS One*, **8**, e64366.

Fajardo LL, Harvey JA, McAleese KA, Roberts CC, Granstrom P (1995). Breast cancer diagnosis in women with subglandular silicone gel-filled augmentation implants. *Radiology*, **194**, 859-62.

Fanshawe TR, Phillips P, Plumb A, et al (2016). Do prevalence expectations affect patterns of visual search and decision-making in interpreting CT colonography endoluminal videos?. *Br J Radiol*, **89**, 20150842.

Faridah Y (2008). Digital versus screen film mammography: a clinical comparison. *Biomed Imag Intervent J*, **4**, e31.

Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, et al (2013). Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*, **49**, 1374-403.

Fleck MS, Samei E, Mitroff SR (2010). Generalized "satisfaction of search": adverse influences on dual-target search accuracy. *J Exp Psychol Appl*, **16**, 60-71

Freer PE (2015). Mammographic breast density: Impact on breast cancer risk and implications for screening. *Radio Graphics*, **35**, 302-15.

Gaur S, Dialani V, Slanetz PJ, Eisenberg RL (2013). Architectural distortion of the breast. *AJR Am J Roentgenol*, **201**, 662-70.

Georgian-Smith D, Moore RH, Halpern E, et al (2007). Blinded comparison of computer-aided detection with human second reading in screening mammography. *Am J Roentgenol*, **189**, 1135-41.

Good BC, Cooperstein LA, DeMarino GB, et al (1990). Does knowledge of the clinical history affect the accuracy of chest radiograph interpretation?. *Am J Roentgenol*, **154**, 709-12.

Gromet M (2008). Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *Am J Roentgenol*, **190**, 854-9.

Guertin MH, Theberge I, Dufresne MP, et al (2014). Clinical image quality in daily practice of breast cancer mammography screening. *Can Assoc Radiol J*, **65**, 199-206.

Gur D, Bandos AI, Fuhrman CR, et al (2007). The prevalence effect in a laboratory environment: Changing the confidence ratings. *Acad Radiol*, **14**, 49-53.

Hambly NM, McNicholas MM, Phelan N, et al (2009). Comparison of digital mammography and screen-film mammography in breast cancer screening: a review in the Irish breast screening program. *Am J Roentgenol*, **193**, 1010-18.

Handel N, Silverstein MJ, Gamagami P, Jensen JA, Collins A (1992). Factors affecting mammographic visualization of the breast after augmentation mammaplasty. *JAMA*, **268**, 1913-7.

Handel N (2007). The effect of silicone implants on the diagnosis, prognosis, and treatment of breast cancer. *Plast Reconstr Surg*, **120**, 81-93.

Hoff SR, Abrahamsen A-L, Samset JH, et al (2012). Breast cancer: Missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography-results from a retrospective review. *Radiology*, **264**, 378-86.

Holland K, Sechopoulos I, den Heeten G, Mann RM, Karssemeijer N (2016). Performance of breast cancer screening depends on mammographic compression. in A. Tingberg, K. Lång and P. Timberg (Eds.), Breast Imaging:

13th International Workshop, IWDM 2016, Malmö, Sweden, June 19-22, 2016, Proceedings, Cham: Springer International Publishing, pp 183-9.

Holli K, Saaristo R, Isola J, Hyöty M, Hakama M (1998). Effect of radiotherapy on the interpretation of routine follow-up mammography after conservative breast surgery: a randomized study. *Br J Cancer*, **78**, 542-5.

Hou MF, Chuang HY, Ou-Yang F, et al (2002). Comparison of breast mammography, sonography and physical examination for screening women at high risk of breast cancer in taiwan. *Ultrasound Med Biol*, **28**, 415-20.

Houssami N, Irwig L, Simpson JM, et al (2003). Sydney breast imaging accuracy study: Comparative sensitivity and specificity of mammography and sonography in young women with symptoms. *Am J Roentgenol*, **180**, 935-40.

Hunt KA, Sickles EA (2000). Effect of obesity on screening mammography: outcomes analysis of 88,346 consecutive examinations. *Am J Roentgenol*, **174**, 1251-5.

Jackson SL, Abraham L, Miglioretti DL, et al (2015). Patient and radiologist characteristics associated with accuracy of two types of diagnostic mammograms. *Am J Roentgenol*, **205**, 456-63.

James JJ, Gilbert FJ, Wallis MG, et al (2010). Mammographic features of breast cancers at single reading with computer-aided detection and at double reading in a large multicenter prospective trial of computer-aided detection: CADET II. *Radiology*, **256**, 379-86.

Jemal A, Bray F, Center MM, et al (2011). Global cancer statistics. *CA Cancer J Clin*, **61**, 69-90.

Johnson K, Sarma D, Hwang ES (2015). Lobular breast cancer series: imaging. *Breast Cancer Res*, **17**, 94.

Kan L, Olivotto IA, Warren Burhenne LJ, Sickles EA, Coldman AJ (2000). Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology*, **215**, 563-7.

Karssemeijer N, Otten JDM, Verbeek ALM, et al (2003). Computer-aided detection versus independent double reading of masses on mammograms. *Radiology*, **227**, 192-200.

Kavanagh AM, Cawson J, Byrnes GB, et al (2005). Hormone replacement therapy, percent mammographic density, and sensitivity of mammography. *Cancer Epidemiol Biomarkers Prev*, **14**, 1060-4.

Keen JD, Keen JE. (2008). How does age affect baseline screening mammography performance measures? A decision model. *BMC Med Inform Decis Mak*, **8**, 40.

Kerlikowske K, Grady D, Barclay J, et al (1998). Variability and accuracy in mammographic interpretation using the American college of radiology breast imaging reporting and data system. *J Natl Cancer Inst*, **90**, 1801-9.

Krupinski EA, Nodine CF, Kundel HL (1993). A perceptually based method for enhancing pulmonary nodule recognition. I*nvest Radiol*, **28**, 289-94.

Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J (2010a). Long radiology workdays reduce detection and accommodation accuracy. *J Am College Radiol*, **7**, 698-704.

Krupinski EA (2010b). Reader fatigue interpreting mammograms. In J. Martí, A. Oliver, J. Freixenet and R. Martí (Eds.), Digital Mammography: 10th International Workshop, IWDM 2010, Girona, Catalonia, Spain, June 16-18, 2010. Proceedings, Berlin, Heidelberg: Springer Berlin Heidelberg, pp 312-8.

Krupinski EA (2010c). Current perspectives in medical image perception. *Atten Percept Psychophys*, **72**, 1205-17.

Krupinski EA, Berbaum KS, Caldwell RT, et al (2012). Do long radiology workdays affect nodule detection in dynamic CT interpretation?. *J Am Coll Radiol*, **9**, 191-8.

Krupinski EA, Berbaum KS, Schartz KM, Caldwell RT, Madsen MT (2017). The impact of fatigue on satisfaction of search in chest radiography. *Acad Radiol*, **24**,1058-63

Lauby-Secretan B, Scoccianti C, Loomis D, et al (2015). Breast-cancer screening viewpoint of the IARC working group. *N Engl J Med*, 372, 2353-8.

Lee SC, Jain PA, Jethwa SC, Tripathy D, Yamashita MW (2014). Radiologist's role in breast cancer staging: providing key information for clinicians. *Radiographics*, **34**, 330-42.

Leslie A, Jones AJ, Goddard PR (2000). The influence of clinical information on the reporting of CT by radiologists. *Br J Radiol*, **73**, 1052-5.

Lowery JT, Byers T, Kittelson J, et al (2011). Differential expression of prognostic biomarkers between interval and screen-detected breast cancers: does age or family history matter?. *Breast Cancer Res Treat*, **129**, 211-9.

Loy CT, Irwig L (2004). Accuracy of diagnostic tests read with and without clinical information: a systematic review. *JAMA*, **292**, 1602-9.

Majid AS, Paredes ESD, Doherty RD, Sharma NR, Salvador X (2003). Missed breast carcinoma: Pitfalls and pearls. *Radio Graphics*, **23**, 881-95.

Malich A, Sauner D, Marx C, et al (2003). Influence of breast lesion size and histologic findings on tumor detection rate of a computer-aided detection system. *Radiology*, **228**, 851-6.

Mandelson MT, Oestreicher N, Porter PL, et al (2000). Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst*, **92**, 1081-7.

Maxwell AJ (1999). Breast cancers missed in the prevalent screening round: effect upon the size distribution of incident round detected cancers. *J Med Screen*, **6**, 28-9.

McCormack VA, dos Santos Silva I (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*, **15**, 1159-69.

Mello-Thoms C (2006). How does the perception of a lesion influence visual search strategy in mammogram reading?. *Acad Radiol*, **13**, 275-88.

Mello-Thoms C, Trieu PD, Rawashdeh MA, et al (2014). Understanding the role of correct lesion assessment in radiologists' reporting of breast cancer. In H. Fujita, T. Hara and C. Muramatsu (Eds.), Breast Imaging: 12th International Workshop, IWDM 2014, Gifu City, Japan, June 29 – July 2, 2014. Proceedings, Cham: Springer International Publishing, pp 341-7.

Miglioretti DL, Rutter CM, Geller BM, et al (2004). Effect of breast augmentation on the accuracy of mammography and cancer characteristics. *JAMA*, **291**, 442-50.

Miglioretti DL, Smith-Bindman R, Abraham L, et al (2007). Radiologist characteristics associated with interpretive performance of diagnostic mammography. *J Natl Cancer Inst*, **99**, 1854-63.

Molina Y, Beresford SA, Thompson B (2016). Psychological outcomes after a false positive mammogram: Preliminary evidence for ethnic differences across time. *J Racial Ethn Health Disparities*, **4**, 123-33

Molins E, Macià F, Ferrer F, Maristany M-T, Castells X (2008). Association between radiologists' experience and accuracy in interpreting screening mammograms. *BMC Health Services Res*, **8**, 91.

Njor S, Nystrom L, Moss S, et al (2012). Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen*, **19**, 33-41.

Njor SH, von Euler-Chelpin M, Tjonneland A, Vejborg I, Lynge E (2016). Body weight and sensitivity of screening mammography. *Eur J Cancer*, **60**, 93-100.

Nodine CF, Kundel HL, Lauver SC, Toto LC (1996). Nature of expertise in searching mammograms for breast masses. *Acad Radiol*, **3**, 1000-6.

Nodine CF, Mello-Thoms C, Kundel HL, Weinstein SP (2002). Time course of perception and decision making during mammographic interpretation. *Am J Roentgenol*, **179**, 917-23.

Norsuddin NM, Mello-Thoms C, Reed W, Brennan PC, Lewis S (2016). Lower recall rates reduced readers' sensitivity in screening mammography. In A. Tingberg, K. Lång and P. Timberg (Eds.), Breast Imaging: 13th International Workshop, IWDM 2016, Malmö, Sweden, June 19-22, 2016, Proceedings, Cham: Springer International Publishing, pp 116-21.

Paci E (2012). Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen*, **19**, 5-13.

Palazzetti V, Guidi F, Ottaviani L, et al (2016). Analysis of mammographic diagnostic errors in breast clinic. *Radiol Med*, **121**, 828-33.

Philpotts LE (2009). Can computer-aided detection be detrimental to mammographic interpretation?. *Radiology*, **253**, 17-22.

Pinto A, Brunese L (2010). Spectrum of diagnostic errors in radiology. *World J Radiol*, **2**, 377-83.

Piroth MD, Fischedick K, Wein B, et al (2014). Fat necrosis and parenchymal scarring after breast-conserving surgery and radiotherapy with an intraoperative electron or fractionated, percutaneous boost: a retrospective comparison. *Breast Cancer*, **21**, 409-14.

Pisano ED, Gatsonis C, Hendrick E, et al (2005). Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*, **353**, 1773-83.

Pisano ED, Hendrick RE, Yaffe MJ, et al (2008). Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. *Radiology*, **246**, 376-83.

Pollard BJ, Samei E, Chawla AS, et al (2009). The influence of increased ambient lighting on mass detection in mammograms. *Acad Radiol*, **16**, 299-304.

Pollard BJ, Samei E, Chawla AS, et al (2012). The effects of ambient lighting in chest radiology reading rooms. *J Digit Imaging*, **25**, 520-6.

Popli M (2001). Pictorial essay: Mammographic features of breast cancer. *Indian J Radiol Imag*, **11**, 175-9.

Popli MB, Teotia R, Narang M, Krishna H (2014). Breast positioning during mammography: Mistakes to be avoided. *Breast Cancer (Auckl)*, **8**, 119-24.

Posso M, Carles M, Rué M, Puig T, Bonfill X (2016). Cost-effectiveness of double reading versus single reading of mammograms in a breast cancer screening programme. *PLoS One*, **11**, e0159806.

Raposo CG, Suarez-Garcia I, Lopez-Gomez M, et al (2012). Comparison of breast cancer subtypes detected by mammography screening and tumors diagnosed with palpable masses. *J Clin Oncol*, **30**, 11028.

Rawashdeh MA, Lee WB, Bourne RM, et al (2013). Markers of good performance in mammography depend on number of annual readings. *Radiology*, **269**, 61-7.

Rayson D, Payne JI, Abdolell M, et al (2011). Comparison of clinical-pathologic characteristics and outcomes of true interval and screen-detected invasive breast cancer among participants of a Canadian breast screening program: a nested case-control study. *Clin Breast Cancer*, **11**, 27-32.

Redondo A, Comas M, Macia F, et al (2012). Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol*, **85**, 1465-70.

Reed WM, Ryan JT, McEntee MF, Evanoff MG, Brennan PC (2011). The effect of abnormality-prevalence expectation on expert observer performance and visual search. *Radiology*, **258**, 938-43.

Roberts-Klein S, Iuanow E, Slanetz PJ (2011). Avoiding pitfalls in mammographic interpretation. *Can Assoc Radiol J*, **62**, 50-9.

Rohan TE, Miller AB (1999). Hormone replacement therapy and risk of benign proliferative epithelial disorders of the breast. *Eur J Cancer Prev*, **8**, 123-30.

Rosen EL, Baker JA, Soo MS (2002). Malignant lesions initially subjected to short-term mammographic follow-up. *Radiology*, **223**, 221-8.

Rosen EL, Sickles E, Keating D (1999). Ability of mammography to reveal nonpalpable breast cancer in women with palpable breast masses. *Am J Roentgenol*, **172**, 309-12.

Samei E, Krupinski E (2010). The handbook of medical image perception and techniques. Cambridge University Pres. The Edinburgh Building, Cambridge CB2 8RU, UK, pp107.

Scott HJ, Gale AG (2006). Breast screening: PERFORMS identifies key mammographic training needs. *Br J Radiol*, **79**, 127-33.

Sickles EA, Wolverton DE, Dee KE (2002). Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology*, **224**, 861-9.

Silverstein MJ, Handel N, Gamagami P (1991). The effect of silicone-gel-filled implants on mammography. *Cancer*, **68**, 1159-63.

Skaane P, Skjennald A (2004). Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program-the Oslo II study. *Radiology*, **232**, 197-204.

Skaane P, Kshirsagar A, Stapleton S, Young K, Castellino RA (2007). Effect of computer-aided detection on independent double reading of paired screen-film and full-field digital screening mammograms. *Am J Roentgenol*, **188**, 377-84.

Skaane P (2009). Studies comparing screen-film mammography and full-field digital mammography in breast cancer screening: updated review. *Acta Radiol*, **50**, 3-14.

Suleiman WI, Lewis SJ, Georgian-Smith D, Evanoff MG, McEntee MF (2014a). Number of mammography cases read per year is a strong predictor of sensitivity. *J Med Imag*, **1**, 015503.

Suleiman WI, Georgian-Smith D, Evanoff MG, Lewis S, McEntee MF (2014b). A comparison of Australian and USA radiologists' performance in detection of breast cancer. in Proceedings of SPIE - The International Society for Optical Engineering; 903714-8, USA

Suleiman WI, McEntee MF, Lewis SJ, et al. (2016a). In the digital era, architectural distortion remains a challenging radiological task. *Clin Radiol*, **71**, 35-40.

Suleman W (2016b). Observer performance factors influencing breast cancer detection in mammography. (Ph.D Thesis), University of Sydney, Sydney.

Suleiman WI, Rawashdeh MA, Lewis SJ, et al (2016c). Impact of breast reader assessment strategy on mammographic radiologists' test reading performance. *J Med Imaging Radiat Oncol*, **60**, 352-8.

Sung JS, Stamler S, Brooks J, et al (2016). Breast cancers detected at screening MR imaging and mammography in patients at high risk: Method of detection reflects tumor histopathologic results. *Radiology*, **280**, 716-22.

Taplin SH, Rutter CM, Finder C, et al (2002). Screening mammography: clinical image quality and the risk of interval breast cancer. *Am J Roentgenol*, **178**, 797-803.

Tartter PI, Weiss S, Ahmed S, et al (1999). Mammographically

occult breast cancers. *Breast J*, 5, 22-5.

Taylor P, Potts HW (2008). Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer*, **44**, 798-807.

Taylor-Phillips S, Wallis MG, Parsons H, et al (2014). Changing case order to optimise patterns of performance in mammography Screening (CO-OPS): study protocol for a randomized controlled trial. *Trials*, **15**, 17.

Taylor-Phillips S, Wallis MG, Jenkinson D, et al (2016). Effect of using the same vs different order for second readings of screening mammograms on rates of breast cancer detection: a randomized clinical trial. *JAMA*, **315**, 1956-65.

Torres-Mejía G, Smith RA, Carranza-Flores M, et al (2015). Radiographers supporting radiologists in the interpretation of screening mammography: a viable strategy to meet the shortage in the number of radiologists. *BMC Cancer*, **15**, 410.

van Breest Smallenburg V, Duijm LEM, Voogd AC, et al (2012a). Lower sensitivity of screening mammography after previous benign breast surgery. *Int J Cancer*, **130**, 122-8.

van Breest Smallenburg V, Duijm LEM, Voogd AC, Jansen FH, Louwman MWJ (2012b). Mammographic changes resulting from benign breast surgery impair breast cancer detection at screening mammography. *Eur J Cancer*, **48**, 2097-103.

Vernacchia FS, Pena ZG (2009). Digital mammography: its impact on recall rates and cancer detection rates in a small community-based radiology practice. *Am J Roentgenol*, **193**, 582-5.

Waite S, Scott J, Gale B, et al (2016). Interpretive error in radiology. *Am J Roentgenol*, **208**, 739-49.

Williams LH, Drew T (2017). Distraction in diagnostic radiology: How is search through volumetric medical images affected by interruptions?. *Cog Res Principl Implicat*, **2**, 12.

Wittenberg R, Peters JF, Sonnemans JJ, et al (2011). Impact of image quality on the performance of computer-aided detection of pulmonary embolism. *Am J Roentgenol*, **196**, 95-101.