Radiography 21 (2015) 334–341

Contents lists available at ScienceDirect

Radiography

journal homepage: www.elsevier.com/locate/radi

Understanding recall rates in screening mammography: A conceptual framework review of the literature

N. Mohd Norsuddin [a, b, *], W. Reed [a], C. Mello-Thoms [a], S.J. Lewis [a]

[a] Medical Imaging Optimisation and Perception Group (MIOPeG), Discipline of Medical Radiation Sciences, Faculty of Health Sciences and Brain Mind Research Institute, The University of Sydney, Lidcombe, NSW, Australia
[b] Diagnostic Imaging & Radiotherapy Programme, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, 50300 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 2 April 2015
Received in revised form
25 May 2015
Accepted 15 June 2015
Available online 17 July 2015

Keywords:
Recall rates
False positive results
Cancer detection rate
Screening mammography
Multidisciplinary
Conceptual framework

ABSTRACT

Recall rates are one of the performance measures used to evaluate the effectiveness of mammography screening programs. There is conflicting evidence regarding the link between recall rates and cancer detection rates and a variety of differing recall rates exist between countries and readers. This variability in recall rates may have important clinical and economic implications such as unnecessary follow-up procedures, additional costs to the health care system and psychological effects for the women themselves associated with false-positive mammograms results. In order to reduce the impact of false positive recall rates in screening mammography, it is essential for all multidisciplinary health care providers, especially those in medical imaging, to fully understand the factors that may contribute and affect recall rates. The multifactorial nature of recall rates is explored in this paper through the construction of a conceptual map based on a review of the current literature.

Introduction

Recall rates are one of the main performance indicators that play a significant role in determining the overall accuracy of a screening mammography programme.[1–6] The performance of screening mammography from a medical imaging perspective is generally measured by indicators such as sensitivity, specificity, positive predictive rates (PPV) and cancer detection rates (CDR). To underpin this paper, the performance measures associated with screening mammography as framed from a medical imaging practice perspective have been defined in Table 1 to assist with understanding the conceptual framework showcasing recall rate variability (Fig. 1).

There are benefits to recalling a percentage of women within a screened population as there is a relationship between recall rates and the early detection of breast cancer.[4] However, it is important to consider what may be an “optimal” rate as false positive recalls have important clinical and economic implications such unnecessary follow-up procedures, costs and adverse psychological effects upon the women recalled.[7–9] Statistics have shown that screening mammography can be successful in reducing breast cancer mortality, with reductions of 21%–26% in women aged 50–69 and a 32% reduction in women aged 70 and above .[10] Inadequate specificity in mammography leads to potential harms such as overdiagnosis, missed cancers and false-positive results.[11–13]

The risk of a false-positive screening result is positively correlated with the recall rate.[4] This rate is initially influenced by the mammographic technologies available at the point of screening, such as screen-film mammography (SFM) and full field digital mammography (FFDM).[14,15] The human physical parameters, that are brought into consideration include the breast reader's expertise and work experiences as well as the woman's presentation (for example clients age, screening history, use of hormone therapy, breast density, previous invasive procedure and familial breast cancer).[5,16,17] Although the technical factors are often easier to control and can be standardized to some extent the human physical parameters of the readers (generally radiologists but also occasionally non-radiologist readers such as breast physicians and radiographers) along with the population of screened women that may present at any given time to a screening mammography programme present a real challenge and contribute to the variability in recall rates.[18]

* Corresponding author. Medical Imaging Optimisation and Perception Group (MIOPeG), Discipline of Medical Radiation Sciences, Faculty of Health Sciences, The University of Sydney, East Street, Lidcombe, NSW 2141, Australia. Tel.: +61 4 50595504.
E-mail address: nmoh5894@uni.sydney.edu.au (N. Mohd Norsuddin).

http://dx.doi.org/10.1016/j.radi.2015.06.003
1078-8174/

Table 1
Definition of terms in screening mammography performance.

Name	Definition
Sensitivity	Measures the percentage or fraction of actual positive cancer cases that are correctly identified. Often described as a decimal. The ability of a test to correctly detect the presence of disease.
Specificity	Measures the percentage or fraction of cancer free cases that are correctly identified. Often described as a decimal. The ability of a test to correctly detect the absence of disease.
Recall rate	The proportion of screened women that are asked to return for further assessment. Often expressed as a percentage.
False positive result	The decision made in error that a case is positive for cancer when the case is actually cancer free.
Positive predictive value (PPV)	The probability of screened women with a positive (malignant) test that do have breast cancer. Often expressed as a percentage.
Negative predictive value (NPV)	The probability of screened women with a negative (normal) test that do have breast cancer. Often expressed as a percentage.
Cancer detection rate (CDR)	The proportion of screened women with breast cancer who test positive for breast cancer. Often expressed as a percentage.

In this review we introduced a novel conceptual mapping of the factors that need to be considered when recall rates are evaluated In particular, we focus on three main areas that may contribute to a woman being recalled for further assessment, including imaging technologies, differences in practices among breast readers and characteristics of the population screened (patient). These three areas were chosen as focal points of discussion in this article through broad thematic analysis of over 400 past published research papers in the area of recall rates in screening mammography.

A conceptual understanding of recall rates

Fig. 1 illustrates the multifactorial nature of recall rates from the literature. The development process of conceptual mapping in this paper began with identification of the wide variation in recall rates in international screening mammography programmes. The search of the literature was conducted in MEDLINE, CINAHL (EbSCOhost), SPIE library, Web of Science, PubMed, Scopus databases and Google Scholar. No specific year of publication was imposed in this search however, we prioritised studies from 2000 onwards which were likely to capture current imaging modalities in screening mammography. Keywords included in this review were recall rates, false positive results, screening mammography, observer performances, performance measures, screening performance, sensitivity, specificity, radiologists and medical imaging. The aetiology of the recall rates were identified and classified into three main sections; imaging technologies such as digital and analogue image acquisition, differences in practice such as the volume of cases read and the experience of the reader and characteristics of screened population including breast density and geographical location. The emboldened factors in the conceptual framework indicate distinct sections that are further discussed in this review.

Despite the fact that breast cancer diagnosis and care is multidisciplinary and inter-professional, much of the past literature has been narrow in scope, seeking to understand recall rates as a single entity remote from other influences. Fig. 1 shows the breath of the variables that affect recall rates, providing health care practitioners with a greater understanding of their role in the context of the larger decision making processes surrounding recalling women in a screening service. This way, each practitioner unit can comprehend holistically the importance of optimising their practice, implementing quality control and an appreciation for individual women's scenarios and why they may be concerned about being recalled for further assessment.

Imaging technologies (screen-film – full field digital mammography)

Over the last decade, the evolution of digital technologies has transformed the technical quality of mammograms through improved image receptor systems that allow for more consistent image quality with higher contrast resolution, fewer artifacts[19,20] and lower radiation dose[21,22] than previous screen-film mammography (SFM). The benefits of post processing capabilities in full field digital mammography (FFDM) have also improved cancer detection, especially in dense breast parenchyma.[14,23] With FFDM, images that have been under-exposed or over-exposed are no longer necessarily repeated, resulting in a lower recall rate among screened women who may have questionable technical images. Although SFM has better spatial resolution than FFDM, which allows better visualization of fine structures that act as biomarkers for breast cancer, such as microcalcification, FFDM has the ability to alter the image contrast and digital information after exposure through magnification, image windowing and panning.

Trials involving comparisons with FFDM and SFM in a screening context have demonstrated conflicting results with regards to recall rates.[15,24,25] A clinical trial by Lewin et al.[15] in the Colorado–Massachusetts Study found no significant differences between FFDM and SFM in cancer detection but with significantly reduced recall rates for women imaged with FFDM. A prospective trials by Skaane et al. concurred with Lewin et al. for results in cancer detection but found higher recall rates for FFDM (Oslo I, 4.6%; Oslo II, 4.2%) when compared to SFM (Oslo I, 3.5%; Oslo II, 2.5%).[24,25] Despite these inconclusive findings, other studies have not replicated such a vast different between SFM and FFDM.[22,26,27] Results from the Digital Mammographic Imaging Screening Trial (DMIST) report there were no differences between SFM and FFDM for the entire population, with the CDR of 0.4% and 0.44% for SFM and FFDM respectively and the recall rate was exactly the same at 8.6% for SFM and FFDM. A specific finding of DMIST was that FFDM demonstrated a significant advantage for specific cohorts of women such as those under the age of 50 years, women with mammographically dense breasts and premenopausal women.[28]

Some limitations of the earlier studies when comparing recall rates in SFM and FFDM included workstation designs with limited post-processing capabilities, emerging detector development and the unfamiliarity of the readers with reading digital cases when transitioning from only reading SFM cases which may affect their study findings.[15] However, with the present wide spread use of FFDM, especially in developed countries, these earlier influencing factors have been improved. Interestingly, readers that have experience in reading SFM perform higher for specificity than those who have only read digital cases. A recent study of 129 radiologists by Rawashdeh et al. (2015) has found that readers who had limited experience with screen-film reading were likely to have lower specificity (0.70 versus 0.83; $p < 0.001$) and hence higher recall rates in comparison to readers that had previous hard copy reading experience, even when there was statistical control for age and experience.[29]

Differences in practice (breast readers)

Reader background

In this section, we explore the characteristics of breast readers, that is, the observer who views the images and arrives at a decision