Detection of pulmonary nodules in chest tomosynthesis

Comparison with chest radiography, evaluation of learning effects and investigation of radiation dose level dependency

Sara Asplund

Department of Radiation Physics
Institute of Clinical Sciences
Sahlgrenska Academy at University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2014

Cover: A pulmonary nodule in the upper part of the lung visible on the tomosynthesis image (middle), and on the reformatted coronal CT image (right), but not on the conventional posteroanterior radiography image (left).

Detection of pulmonary nodules in chest tomosynthesis: Comparison with radiography, evaluation of learning effects and investigation of radiation dose level dependency

sara.asplund@vgregion.se

ISBN 978-91-628-8921-0

E-publication: http://hdl.handle.net/2077/35205

Printed in Gothenburg, Sweden 2014
Aidla Trading AB/Kompendiet

Till mina reskamrater

Nog finns det mål och mening i vår färd -
men det är vägen, som är mödan värd.

Karin Boye

Detection of pulmonary nodules in chest tomosynthesis

Comparison with chest radiography, evaluation of learning effects and investigation of radiation dose level dependency

Sara Asplund

Department of Radiation Physics, Institute of Clinical Sciences at Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

ABSTRACT

Chest tomosynthesis is a relatively recently introduced technique in healthcare, which produces section images of the chest at a lower radiation dose than computed tomography (CT) and with better depth resolution than conventional chest radiography. The primary aims of the studies described in this dissertation were to compare chest tomosynthesis with conventional radiography, to evaluate the effects of clinical experience and learning with feedback on the performance of observers analyzing tomosynthesis images, and to investigate the effect of radiation dose level in tomosynthesis, in the detection of pulmonary nodules. Human observer studies were performed, in which radiologists were instructed to localize and rate pulmonary nodules in patient images. Chest CT was used as reference. The observers' performance regarding the detection of nodules was used as measure of detectability. The results of the studies indicate that the detection of pulmonary nodules is better in chest tomosynthesis than in conventional chest radiography, that experienced thoracic radiologists can quickly adapt to the new technique, that inexperienced observers may perform at a similar level to experienced radiologists after a learning session with feedback, and that a substantial reduction in the effective dose to the patient may be possible.

Keywords: Chest radiology, Chest tomosynthesis, Nodule detection, Observer performance, Free-response receiver operating characteristics.
ISBN: 978-91-628-8921-0
E-publication: http://hdl.handle.net/2077/35205

Detektion av noduler i lungtomosyntes

Jämförelse med lungröntgen, utvärdering av inlärningseffekter och analys av stråldosnivåns inverkan

POPULÄRVETENSKAPLIG SAMMANFATTNING

Vid en traditionell röntgenundersökning (s.k. projektionsröntgen eller slätröntgen) avbildas kroppens 3-dimensionella strukturer genom att de projiceras ner i ett plan. Detta innebär att framför- och bakomliggande strukturer överlappar varandra, vilket kan försvåra detektionen av små strukturer. Datortomografi (DT, eller skiktröntgen) är en röntgenteknik som – till skillnad från traditionell röntgen – ger snittbilder av kroppen, och därmed är problemet med överlagrad anatomi löst. Detta sker dock till kostnaden av en avsevärt högre stråldos till patienten. Tomosyntes är en relativt ny röntgenteknik som – liksom DT – ger snittbilder, men till en stråldos som liknar den för en traditionell röntgenundersökning. Att åstadkomma tillräckligt god bildkvalitet vid så låg stråldos som rimligt är möjligt är viktigt, och det kan framför allt framför allt gynna unga patienter som behöver genomgå många undersökningar med DT, eftersom de då inte utsätts för lika stor risk för att utveckla cancer senare i livet. Tomosyntes innebär också en mindre kostnad och mindre tidsåtgång jämfört med DT. All överlagrad anatomi kan dock inte elimineras helt i tomosyntesbilderna, även om den kan minskas avsevärt jämfört med traditionell röntgen.

I studierna som presenteras i denna avhandling undersöks huruvida lungtomosyntes kan förbättra detektionen av lungnoduler (små tumörmisstänkta strukturer) jämfört med traditionell lungröntgen och om en dossänkning av tomosyntesundersökningen är möjlig utan att detektionen av noduler försämras. Den behandlar också tolkning av tomosyntesbilder vad gäller inlärning och eventuella fallgropar.

Slutsatserna från studierna är att lungtomosyntes är överlägsen traditionell lungröntgen när det gäller detektion av lungnoduler, speciellt då det gäller små noduler, vilka är svårast att detektera i traditionella lungröntgenbilder. När det gäller stråldosen för tomosyntes, så finns det möjligheter att sänka dosen avsevärt utan att minska noduldetektionen nämnvärt. När det gäller inlärning så indikerar resultaten att erfarna thoraxröntgenläkare kan lära sig att granska tomosyntesbilder trots relativt liten erfarenhet av tekniken och att

oerfarna granskare kan komma upp till en liknande nivå med hjälp av inlärning. De fallgropar i tomosyntes som utmärkte sig mest utgjordes främst av svårigheten att avgöra strukturers läge i och nära gränsskikt mellan lunga och kompakt överlagrande anatomi, framför allt längst bak och längst fram i lungorna där revben ofta överlagrar intressanta strukturer. Detta beror på att tomosyntesteknikens begränsade upplösning har störst betydelse i dessa områden. Slutsatserna visar på att tomosyntes har stor potential att förbättra lungröntgendiagnostiken till en försumbar ökning i stråldos samt att tekniken kan utgöra ett värdefullt redskap inom lungdiagnostik i framtiden.

LIST OF PAPERS

This dissertation is based on the following studies, referred to in the text by their Roman numerals.

I. Vikgren J, Zachrisson* S, Svalkvist A, Johnsson Å A, Boijsen M, Flinck A, Kheddache S and Båth M
Comparison of chest tomosynthesis and chest radiography for detection of pulmonary nodules: human observer study of clinical cases
Radiology 2008;249(3):1034-1041

II. Zachrisson* S, Vikgren J, Svalkvist A, Johnsson Å A, Boijsen M, Flinck A, Månsson L G, Kheddache S and Båth M
Effect of clinical experience of chest tomosynthesis on detection of pulmonary nodules
Acta Radiologica 2009;50(8):884-891

III. Asplund S, Johnsson Å A, Vikgren J, Svalkvist A, Boijsen M, Fisichella V A, Flinck A, Wiksell Å, Ivarsson J, Rystedt H, Månsson L G, Kheddache S and Båth M
Learning aspects and potential pitfalls regarding detection of pulmonary nodules in chest tomosynthesis and proposed related quality criteria
Acta Radiologica 2011;52(5):503-512

IV. Asplund S, Johnsson Å A, Vikgren J, Svalkvist A, Flinck A, Boijsen M, Fisichella V A, Månsson L G and Båth M
Effect of radiation dose level on the detectability of pulmonary nodules in chest tomosynthesis
Accepted for publication in European Radiology.
The final publication is available at Springer via http://dx.doi.org/10.1007/s00330-014-3182-1.

*Zachrisson was the author's maiden name until 2010.

Preliminary results have been presented at the following conferences:

Zachrisson S, Vikgren J, Svalkvist A, Johnsson Å A, Boijsen M, Flinck A, Månsson L G, Kheddache S and Båth M
Evaluation of chest tomosynthesis for the detection of pulmonary nodules: effect of clinical experience and comparison with chest radiography
Presented at SPIE Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment, February 7-12, 2009, Orlando, FL, USA

Zachrisson S, Johnsson Å A, Vikgren J, Svalkvist A, Flinck A, Boijsen M, Kheddache S, Månsson L G, and Båth M
Experience of chest tomosynthesis at Sahlgrenska University Hospital
Presented at the Annual Swedish X-ray Conference (*Röntgenveckan*), September 20-24, 2010, Örebro, Sweden

Asplund S, Johnsson Å A, Vikgren J, Svalkvist A, Boijsen M, Fisichella V A, Flinck A, Wiksell Å, Ivarsson J, Rystedt H, Månsson L G, Kheddache S and Båth M
Extended analysis of the effect of learning with feedback on the detectability of pulmonary nodules in chest tomosynthesis
Presented at SPIE Medical Imaging 2011: Image Perception, Observer Performance, and Technology Assessment, February 12-17, 2011, Orlando, FL, USA

Asplund S, Vikgren J, Svalkvist A, Johnsson Å A , Boijsen M, Flinck A, Fisichella V A, Wiksell Å, Ivarsson J, Rystedt H, Månsson L G, Kheddache S and Båth M
Observer performance studies on chest tomosynthesis at Sahlgrenska University Hospital: detectability of pulmonary nodules and observer learning effects
Presented at the Swedish Society for Radiation Physics Conference (*Radiofysikdagarna*), November 14-15, 2011, Ystad, Sweden

CONTENTS

ABBREVIATIONS

AFROC	Alternative free-response receiver operating characteristics
AUC	Area under the curve
CI	Confidence interval
CT	Computed tomography
DICOM	Digital imaging and communications in medicine
DQE	Detective quantum efficiency
FOM	Figure of merit
FPF	False positive fraction
FROC	Free-response receiver operating characteristics
JAFROC	Jackknife alternative free-response receiver operating characteristics
LAT	Lateral
LL	Lesion localization
LLF	Lesion localization fraction
MDCT	Multidetector computed tomography
NL	Non-lesion localization
NLF	Non-lesion localization fraction
NPS	Noise power spectrum
PA	posteroanterior
ROC	Receiver operating characteristics
TPF	True positive fraction

VG	Visual grading
ViewDEX	Viewer for digital evaluation of X-ray images

DEFINITIONS IN SHORT

AFROC curve	The plot of the lesion localization fraction versus the false positive fraction.
FROC curve	The plot of the lesion localization fraction versus the non-lesion localization fraction.
Highest noise rating	The non-lesion localization with the highest rating in a case.
JAFROC (JAFROC2) figure of merit	The area under the AFROC curve, using the highest noise rating in normal cases to calculate the false positive fraction.
JAFROC1 figure of merit	The area under the AFROC curve, using the highest noise rating in normal *and* abnormal cases to calculate the false positive fraction.
ROC curve	The plot of true positive fraction versus the false positive fraction.

1 GENERAL INTRODUCTION

Conventional chest radiography is a radiographic projection technique that has been available in healthcare for more than a century. It is an easily accessible, inexpensive form of examination[1,2], but has the drawback of limited sensitivity, as overlapping anatomy may obscure pathology[3–5]. Computed tomography (CT), which was introduced to healthcare in the 1970s, is a 3-dimensional technique providing parallel sections of the body, and obscuring anatomy can thus be eliminated. Structures of interest may, therefore, be more easily detected than in conventional radiography. The disadvantages usually associated with CT are high effective doses, high cost and lower accessibility than conventional radiography.

Chest tomosynthesis is a rather new technique that has recently been introduced to healthcare[6–11]. In chest tomosynthesis, the same equipment is used as for conventional chest radiography, but the X-ray tube is moved vertically relative to the image detector through a limited angular interval while projection images are acquired. These projection images are then used to reconstruct an arbitrary number of section images, thus reducing the overlapping anatomy. The potential benefits associated with tomosynthesis are low radiation doses, low costs and easy access compared to CT, and enhanced image quality compared to conventional radiography.

When a new imaging technique, such as chest tomosynthesis, is introduced, extensive investigations are required to establish its usefulness and validity in healthcare. For example, it should be tested against already existing standard techniques, optimized and tested for various diagnostic questions. One of the most challenging tasks for the thoracic radiologist is the detection of pulmonary nodules[12], i.e. small rounded structures which may potentially be malignant. Because of the difficulty of the task, but also because of the great clinical importance of pulmonary nodules, the detectability of these lesions is often used as measure of performance. Image quality criteria based on important anatomical landmarks may also be suitable for optimization of this new technique. No such quality criteria are currently available and, therefore, suitable quality criteria need to be developed.

Whenever ionizing radiation is used, the exposure of human beings shall be kept as low as reasonably achievable[13]. In medical imaging, attempts must always be made to obtain a diagnostic image with acceptable quality using the lowest possible radiation dose to the patient. Since chest tomosynthesis is a relatively new technique, there is only limited knowledge on the effects of

dose reduction on the image quality, and the possibility of reducing the dose while ensuring sufficient image quality. Moreover, when a technique has only been in clinical use for a short period of time, there may be a lack of knowledge regarding how to correctly analyze the images of the new modality. Information on the difficulties associated with interpreting the images obtained with the new modality may therefore be valuable.

Chest tomosynthesis was introduced at the Department of Radiology at Sahlgrenska University Hospital in December 2006. In order to study tomosynthesis, a research group was established including both thoracic radiologists and medical physicists. Since the radiologists at the department were among the very first in the world to use chest tomosynthesis clinically, none of them had any experience of the technique at that time. In order to investigate learning effects in chest tomosynthesis, the research group initiated collaboration with another research group at the Institution of Education, Communication and Learning at the University of Gothenburg. The project in which this collaboration was incorporated focuses on how radiologists adapt their methods of interpretation and diagnosis when using new imaging techniques. The project is part of a larger interdisciplinary research collaboration, called the LETStudio (www.letstudio.gu.se), which investigates knowledge, learning, communication and expertise in modern society, particularly through the introduction of new media-based technologies.

At the Department of Radiology at Sahlgrenska University Hospital, chest tomosynthesis has so far primarily been used as an additional mode of examination for the evaluation of suspicious findings in chest radiographs[11,14]. However, chest tomosynthesis may be useful for several other purposes, but thorough investigations are needed to determine situations in which chest tomosynthesis is most suitable. The aim of the studies described in this dissertation was to help elucidate this issue.

2 AIMS

The aims of the research presented in this dissertation were:

- to compare chest tomosynthesis and conventional chest radiography regarding the detection of pulmonary nodules (Paper I),
- to investigate the effect of clinical experience and learning with feedback on observer performance regarding pulmonary nodule detection (Papers II and III),
- to identify potential pitfalls and to formulate image quality criteria for chest tomosynthesis (Paper III), and
- to investigate the effect of radiation dose level on the detectability of pulmonary nodules in chest tomosynthesis (Paper IV).

3 BACKGROUND

3.1 Conventional chest radiography

Conventional chest radiography is one of the most common radiological procedures performed at medical imaging departments. It is a valuable tool for rapidly obtaining information on the status of the heart and lungs, and for identifying various lung diseases, including lung cancer, which is the most common cause of cancer deaths globally[15]. Conventional chest radiography is associated with easy access and low costs. It also has the benefit of low radiation doses to patients. Typically effective doses for a radiography examination, including a posteroanterior (PA) and a lateral (LAT) projection are 0.05-0.1 mSv[16–20]. However, being a projection technique, the overlapping anatomy, which has been shown to be the main factor limiting the detection of many types of lesions in radiographs[21–28], obscures structures of interest and conventional chest radiography has been shown to suffer from low sensitivity of lesions[3–5].

3.2 Computed tomography of the chest

CT provides section images of the chest, and the problem of overlapping anatomy is thus eliminated. The detectability of lesions is therefore superior to that in conventional chest radiography. However, CT also has several drawbacks; patient doses when performing chest CT can be as high as seventy times those in conventional chest radiography[20] and the technique is more expensive and time consuming than conventional radiography. Notwithstanding these disadvantages, the use of CT has increased over the past decade in, for example, the USA and the Nordic countries[29,30]. The Nordic Radiation Protection Co-operation recently expressed concern over the increased use of CT, due to the higher radiation doses associated with this modality. In a press release issued in 2012 they warned against the overuse of CT[30]. Further, in a study on the effect of CT, it has been found that the baseline cancer risk increased with cumulative exposure to radiation from CT[31]. Awareness of radiation risks in medical imaging has recently resulted in efforts to reduce CT doses, resulting, for example, in tube current modulation, adapted selection of tube voltage, iterative reconstruction techniques and low-dose protocols for specific examinations[32–37]. Such efforts have resulted in a reduction in chest CT doses to around or even below 1 mSv[33–37]. Even so,

the effective doses associated with CT in most clinical situations remain up to several mSv[11].

3.3 Tomosynthesis

Tomosynthesis has only recently been introduced into healthcare, despite the fact that it was investigated in conjunction with the development of conventional tomography. During this rather extended time period, many researchers competed in the attempt to develop a section imaging X-ray technique[38]. The technique that was later called tomosynthesis by Grant in 1972[39] was initially described by Ziedses des Plantes in 1932[40], and the first system was constructed by Garrison and coworkers in 1969[41]. However, the lack of fast computers and fast read-out detectors made tomosynthesis unsuitable in healthcare, until recently. CT became the gold standard in medical imaging as it was possible to satisfy its technical demands at an earlier stage.

The same equipment as that used in conventional radiography is used in tomosynthesis, but whereas the configuration is stationary in conventional radiography, the tube is moved in tomosynthesis and images are acquired at various angles in a limited angular interval[6,9,10]. The technique is closely related to conventional tomography, in which one section image is acquired per sweep. The major drawback of conventional tomography is that multiple images require additional sweeps, thus increasing the radiation burden on the patient with every additional image. In tomosynthesis, however, this problem is overcome by fast read-out of the detector, resulting in several projection images being acquired at extremely low doses. These projection images are then used to reconstruct arbitrarily chosen section images of the body. One technique that is often used to describe the reconstruction of the tomosynthesis section images is the shift-and-add method, which is essentially equivalent to unfiltered back projection[6]. According to this method, the projection images acquired at various angles are shifted in relation to each other, resulting in the blurring of structures outside the plane of interest, while the in-plane structures are focused. The concept of the shift-and-add technique is visualized in Figure 1.

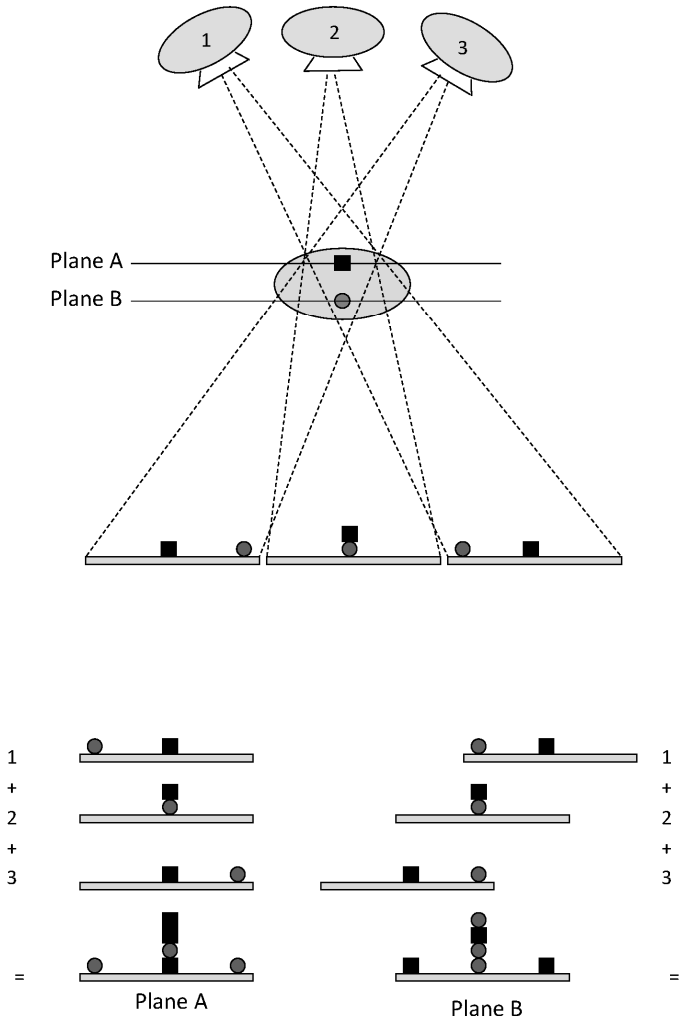


Figure 1. Illustration of the images obtained by moving the X-ray tube relative to the detector in tomosynthesis. The anatomical structures in different planes (A and B) are projected onto different locations in the image receptor at different angles, as shown above. Applying the shift-and-add reconstruction technique brings the structure of interest into focus, while structures in other plains are blurred, as shown in the lower part of the figure.

As only a limited angular interval is used for the acquisition of the projection images in tomosynthesis, parts of the frequency space remain unsampled[6]. This results in poorer depth resolution than CT, in which the whole frequency space is sampled (although the sampling density is higher at the center, and is compensated for by filtering). The depth resolution in tomosynthesis can be increased by increasing the angular interval, but this will either result in a decrease in the projection density, i.e. the number of projections divided by the total angle (if the dose per projection image is unchanged), or an increase in the total radiation dose to the patient (if the projection density is unchanged)[42]. Reducing the number of projection images may lead to artifacts, as the blurring of out-of-plane structures may produce ripple when the number of projection images used for reconstruction is insufficient[42]. Other artifacts are also common in tomosynthesis. One of these is the ghost artifact, which is a result of incomplete blurring of high-contrast objects extending in the sweep direction[42], and is seen as reproduction of the structure in many consecutive section images where it should not be present. Another artifact, which is caused by the limited angle interval used in tomosynthesis, is incomplete cancellation of structures outside the plane of interest[42]. This blurring artifact is most prominent for highly attenuating structures perpendicular to the direction of the tomosynthesis sweep, for example, ribs in the case of chest tomosynthesis. However, although tomosynthesis has poorer depth resolution than CT, the technique results in higher in-plane resolution, since flat-panel detectors with very high resolution is used in tomosynthesis.

The three major applications of tomosynthesis are in breast, chest and orthopedic examinations, and breast tomosynthesis[10,43–45] has been the subject of most interest to date. The use of the technique in clinical imaging has been suggested, but its use in breast cancer screening has also attracted considerable attention[43–49]. Promising results have been reported from screening trials using breast tomosynthesis combined with conventional mammography compared to conventional mammography alone[49,50]. In orthopedic imaging, tomosynthesis has been shown to have the potential to improve radiography in several diagnostic tasks[51–53]. The third application mentioned above, chest tomosynthesis, will be described in more detail in the next section.

3.3.1 Chest tomosynthesis

Chest tomosynthesis is one of the most common applications of tomosynthesis[7–11]. The positioning of the patient is usually identical to the PA positioning in conventional chest radiography, i.e. the patient stands in an upright

position, front of the chest facing the detector. A supine or prone position is also possible if a tabletop system is being used. A LAT chest tomosynthesis examination may be performed, but the effective dose to the patient for a LAT tomosynthesis may be 3-4 times higher than for a PA tomosynthesis (assuming that the relationship between the effective doses for a PA and a LAT projection in conventional chest radiography is also valid for tomosynthesis[17]). In order to avoid breathing artifacts in the tomosynthesis images, the patients are instructed to hold their breath during the tomosynthesis sweep, which takes approximately 5-10 seconds depending on the equipment. Typically reported effective doses to patients undergoing chest tomosynthesis are 0.1-0.2 mSv[16–19,54], but as chest tomosynthesis has not been optimized to the same degree as conventional chest radiography and chest CT, it may be possible to reduce these doses.

Commercial chest tomosynthesis systems are presently supplied by Fujifilm, Shimadzu and GE Healthcare. In all four studies described in this dissertation, the GE Healthcare tomosynthesis system, VolumeRAD (or a beta version of the commercially available product) was used. VolumeRAD has an exposure angular interval of ±15º, and 60 projection images are acquired during the 11 s sweep. In order to determine a suitable tube output, a scout image is acquired at 0º. The tube output used for the scout image is multiplied by a user-adjustable dose ratio, and is then equally distributed between the 60 projection exposures. The reconstruction technique employed is filtered back-projection, and the section images are usually reconstructed at intervals of 5 mm, typically resulting in about 60 coronal images of the chest. However, more images may be reconstructed for larger patients, if necessary.

Apart from the artifacts associated with tomosynthesis in general, motion artifacts may occur in chest tomosynthesis if the patient is unable to stand still, or hold his or her breath during the entire sweep. These may result in lower detectability of lesions[55]. Motion artifacts due to the motion of the beating heart, which are not as severe as breathing artifacts, cannot be avoided.

As mentioned above, chest tomosynthesis is a relatively inexpensive method. The cost of chest tomosynthesis is presented in Paper I. The extra cost involved in purchasing the GE Healthcare tomosynthesis option in addition to the radiography system, is about 25% of the cost of the conventional radiography system. The clinical costs of chest tomosynthesis at Sahlgrenska University Hospital, including reading time, archiving costs etc., adds ~40% to the cost of a conventional chest radiography examination alone, and is ~17% of the cost for a chest CT examination. The radiologist's reading time is

longer for chest tomosynthesis than for conventional chest radiography because of the larger number of images. The reading time for a chest tomosynthesis examination at Sahlgrenska University Hospital was estimated to be 2-5 minutes, while the reading time for a conventional chest radiography examination was 30 seconds - 5 minutes, and that for a chest CT examination 3-10 minutes[10].

Since the tomosynthesis images have higher resolution in the image plane, a tomosynthesis examination may require more storage space than a CT examination, despite the fact that the CT examination consists of a larger number of images. The size of a typical chest tomosynthesis image or a conventional radiograph is ~5-8 megabyte. The size of an entire chest tomosynthesis examination is then ~300-500 megabyte, while the size of a conventional chest radiography examination is ~10-20 megabyte. The typical size of a single chest CT image is ~0.5 megabyte, and an examination might contain up to a thousand images. In such a case the CT examination may reach the storage size of a tomosynthesis examination.

Quaia et al. analyzed the effect on the total cost after the implementation of chest tomosynthesis at the Department of Radiology at the Cattinara Hospital in Trieste, Italy, and found that when chest tomosynthesis was used for follow-up instead of chest CT for patients with suspicious findings on conventional chest radiographs, savings of €8000 and €19 000 were made during one year, compared with the use of unenhanced and contrast-enhanced CT, respectively[56]. Their calculations were based on 271 patients undergoing CT during the year before implementation of tomosynthesis, and 260 patients undergoing conventional radiography, tomosynthesis and CT during the year after implementation.

3.4 Pulmonary nodules

A pulmonary nodule is a well-defined rounded structure, which is restricted to the lung parenchyma and is less than 3 cm in diameter[12]. The detection of such lesions is considered one of the most difficult tasks in thoracic radiology, and the presence of a nodule usually raises the suspicion of malignancy. Because of the clinical importance of nodules and the difficulty in detecting them, the task of detecting nodules is often used for testing the performance of equipment or observers.

The size and growth rate of nodules constitute important information in deciding the follow-up and management of the patient[57]. Larger nodules, i.e. those approaching 30 mm in diameter, are more likely to be malignant, whereas nodules less than 10 mm in diameter are more likely to be benign[58]. The volume doubling time of small nodules may be used as an indication of malignancy, and guidelines for the follow-up and management of nodules have been established by the Fleischner Society[57]. According to these guidelines, nodules are divided into the categories ≤4, >4-≤6, >6-≤8 and >8 mm. For nodules ≤4 mm, there is little risk of malignancy, and they therefore do not require follow-up in low-risk patients. Follow-up is recommended for the other size categories, at shorter intervals the larger the nodule, in order to determine whether the nodule volume doubling time indicates malignancy. For the largest nodules, >8 mm, further diagnostic imaging, biopsy or thoracoscopic resection may be considered in high-risk patients. The categorization of patients as low- and high-risk depends, for example, on age and smoking history; shorter follow-up intervals being recommended in high-risk patients.

Nodule size and volume doubling time are not the only indications of malignancy, and are not directly applicable to all nodules. The characteristics of the nodule must also be taken into account[57]. For example, non-solid or part-solid nodules, which are often associated with malignancy[58], may require longer follow-up periods as they may grow slowly[57]. Calcified nodules are usually benign, but may occasionally be malignant, especially in patients with skeletal cancer, as in such cases they may indicate metastatic disease[59]. The surface of the nodule may also indicate whether it is benign or malignant; smooth surfaces being more often indicative of benignity, while malignant nodules are more likely to be spiculated at the margins[59]. The location of a nodule may also indicate whether it is benign or malignant. It has been observed that malignancies are more likely to be situated centrally, in the upper lobes of the lungs, and more often in the right lung than the left[60,61].

3.5 Image interpretation

Radiological images are complex, and their analysis is often a difficult task. Images containing a high degree of overlapping anatomy, such as conventional radiographs, pose the greatest difficulties since structures of interest are obstructed by anatomical structures, while methods producing many images, such as CT or tomosynthesis, can be demanding due to the large amount of information that has to be processed by the observer[62]. When radiologists

analyze medical images, they seem to compare the image to a "mental library" of previously viewed images of anatomy and pathology; this library being expanded as the experience of the radiologist increases[62]. Therefore, it could be assumed that radiologists with more experience should perform better than inexperienced observers. However, experience may not be the only factor on which the performance of an observer depends. For example, it has been reported that among radiologists screening mammograms, observers who were more recently trained performed better, despite the fact that they were not as experienced[63]. Other factors, such as visual acuity and the quality of feedback may also play an important role[62,63]. Performance may also depend on the conditions under which the observer analyses the images. For example, after many hours of analyzing images observers may suffer from fatigue, which may have negative effects on their performance[64]. Their performance may also depend on the reading environment, and perhaps also on the talent of the observer for the specific task[65].

Observational errors in analyzing medical images can be divided into two major categories: false positive errors and false negative errors. False positive errors (i.e. errors caused by misjudging a structure that is *not* a lesion as a lesion) may often be an effect of overlapping anatomy mimicking a lesion. Decreasing the effects of overlapping anatomical structures could, therefore, reduce the occurrence of such errors. Regarding false negative errors (i.e. lesions not identified as such), three major types of such errors are usually used for error classification in radiography; search errors, recognition errors and decision errors[66]. Search errors are due to the incomplete search of the image, resulting in the observer not focusing on the lesion at all; recognition errors occur because the observer fails to recognize a structure as something that should be reported; while decision errors occur because the observer recognizes the structure but erroneously makes the decision not to report it. One explanation of false negative errors may be that once an observer has already found a lesion, he or she is less likely to find another lesion in the same patient[67]; a phenomenon known as satisfaction of search. It may also be caused by overlapping anatomy hiding the lesions. It is difficult to know how to avoid these errors, although the causes might be known. However, examining the reasons why errors occur, thereby making the observer aware of them, may help the observer to avoid making them in the future.

3.6 Human observer studies

The evaluation of image quality is an important issue in medical imaging. Radiation, as well as social and economic resources, should be used as efficiently as possible[13], and image quality evaluation is of great importance in this context. There are several methods of evaluating image quality, and many aspects must be taken into consideration. For example, physical measures, such as the signal-to-noise ratio or the detective quantum efficiency (DQE), are often used. Physical measures are, however, not sufficient to evaluate a process that includes X-ray transmission through the imaged object, detection, signal sampling, image processing and display, and finally, the human observer[68]. Since the observer's interpretation is crucial for the diagnostic outcome of the patient, the effects of human observers should be included, and this can be achieved by conducting a human observer study. However, human observer studies are laboratory studies, and are therefore limited compared to studies investigating patient care, treatment, outcome and cost-benefit or cost-effectiveness[68]. Human observer studies are, however, easier to perform and are therefore often used to compare image quality in different imaging modalities, although such studies cannot prove the advantage of one modality over another at a higher level, for example, regarding patient outcome.

The terms used in the literature to describe human observer studies vary. In this dissertation, two major branches of human observer studies will be discussed; observer performance studies and visual grading (VG) studies[69]. Observer performance studies will refer to studies in which the observer is analyzing images in order to find pathology, and VG studies will refer to studies in which the observer's judgment concerning the fulfillment of image quality criteria for anatomical structures is analyzed. Observer performance methods are more generally widespread and accepted because they are more objective; i.e. there is a known truth to which the observer's decisions are compared and judged, while the VG task is based on the observer's subjective opinion about the image quality[69]. The disadvantages of observer performance studies are that they are relatively time consuming and they may require large sample sizes in order to attain sufficient statistical power. Visual grading studies are more convenient when evaluating the quality of images from an examination for which there are usually few patients with a specific pathology, as the collection of sufficient data may otherwise be unacceptably long. VG is frequently used in Europe in image quality evaluation studies and it has been indicated that it may be as appropriate for the evaluation of clinical images as observer performance methods[70]. Criticism has been directed towards VG studies as the data are often treated as interval data, when in fact

they are ordinal. Methods for the appropriate analysis of VG have, however, recently been developed[71–73].

3.6.1 Receiver operating characteristics

The most commonly used method for observer performance studies in the healthcare sector is receiver operating characteristics (ROC) analysis[68,74]. The theory behind ROC is signal detection theory, which originates from World War II research on radar, and was introduced into psychophysics in the 1950s[75]. The theory of ROC is based on asking an observer – blinded to the actual truth – to differentiate between images of healthy (normal) and diseased (abnormal) patients using a specified reporting threshold. The proportion of abnormal patients actually reported as abnormal gives the sensitivity of the observer to the task at the specific threshold, and the proportion of the normal patients reported as normal gives the specificity.

If the observer is instructed to use a stricter threshold and read the patient material once again, this will result in fewer normal patients incorrectly being judged as abnormal, but also in fewer abnormal patients correctly being judged as being abnormal. This procedure can be repeated with even stricter thresholds, resulting in new values of the sensitivity and specificity for each new threshold. Instead of repeatedly altering the threshold and re-reading the image material, the same data can be obtained by asking the observer to use a rating scale. Table 1 gives an example of the distribution of ratings and the calculation of the sensitivity, also called the true positive fraction (TPF), and the false positive fraction (FPF), which is defined as 1–specificity. Figure 2 shows a plot of the TPF versus the FPF, which is called the ROC curve. The area under the ROC curve (AUC_{ROC}) provides a measure of the observer performance, and the difference between the AUC_{ROC} for different modalities or settings can be calculated in order to find, for example, the best modality or the optimal settings for a specific system. The point (0,0) is added to the plot in order to represent the strictest possible threshold.

Table 1. Confidence ratings of abnormal and normal cases and the calculated true positive fraction (TPF) and false positive fraction (FPF) at every threshold. High ratings correspond to high confidence levels.

Rating	**4**	**3**	**2**	**1**	**Total**
Abnormal	41	16	8	7	72
Normal	4	9	23	34	70
Threshold	**=4**	**≥3**	**≥2**	**≥1**	
TPF	0.57	0.79	0.90	1.0	
FPF	0.06	0.19	0.51	1.0	

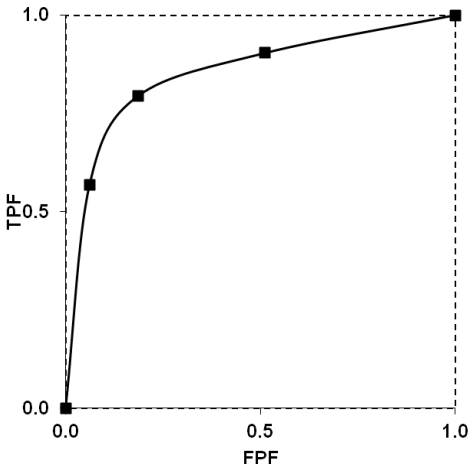


Figure 2. Receiver operating characteristics (ROC) plot based on the values of the true positive fraction (TPF) and false positive fraction (FPF) given in Table 1.

A well-known problem associated with ROC is that the observer's judgment of a case as being abnormal may be based on the classification of a structure that is *not* a lesion as a lesion, while missing the lesion on which the decision should have been based, i.e. is given a TP for the wrong reasons. Another disadvantage of ROC is that the location of the lesion is not included, despite the fact that this information is important, and is reported by the radiologist in the clinical setting. Various strategies have been developed to include lesion location in ROC to improve the method[76]. Only the most commonly used of

these methods, free-response ROC[77], will be described below, as this methodology was used in the studies described in this dissertation.

3.6.2 Free-response receiver operating characteristics

According to the free-response receiver operating characteristics (FROC) paradigm[77] the task of the observer is to detect, mark and rate suspicious lesions. If a mark is made within a predetermined acceptance radius of the lesion, the mark is considered to be a true positive mark, otherwise it is considered to be a false positive mark. In this dissertation, true and false positive *marks* according to the FROC paradigm will usually be referred to as lesion localizations (LLs) and non-lesion localizations (NLs), respectively, in accordance with Chakraborty[78], in order to distinguish them from true and false positive *cases* in traditional ROC. The FROC paradigm has higher statistical power than traditional ROC[79] as the localization of lesions results in more data, and because the observer will not be rewarded for reporting a false lesion while missing a true lesion.

The number of LLs relative to the total number of lesions, the lesion localization fraction (LLF), and the number of NLs per image, the non-lesion localization fraction (NLF), can be plotted against each other to form a FROC curve. The NLF is plotted on the x-axis and the LLF is plotted on the y-axis. The FROC curve can be infinitely extended (at least theoretically) in the x-direction, since there is no limitation on the number of NLs that an observer can make. Since the number of NLs may differ between observers, modalities and tests, it is difficult to determine a consistent figure of merit (FOM) based on FROC curves. The FROC curve is therefore better for visualizing the way in which the confidence levels are used by the observers. Another method, called the alternative FROC (AFROC), has therefore been suggested for the analysis of FROC data[80]. Only the highest rated NL (i.e. highest noise rating) per image is used to calculate the area under the AFROC curve. The AFROC curve is therefore limited to the unit square, ranging from 0 to 1 in both the x- and y-directions, allowing a consistent FOM to be calculated. As all the NLs are not used for plotting the AFROC curve, it contains less of the available information than the FROC curve.

3.6.3 Jackknife alternative free-response receiver operating characteristics

Jackknife alternative FROC (JAFROC) methods are most commonly used for statistical analysis of multi-reader, multi-case FROC data[81]. There are two variants of JAFROC, denoted JAFROC1 and JAFROC2 (the latter is also referred to simply as JAFROC)[82]. Since these JAFROC analysis tools are relatively new, and are still being developed and improved, the recommendations for which of the methods to use have varied with the implementation of new knowledge of the methods.

AUC_{AFROC} is used as the FOM for both JAFROC methods, but in JAFROC2 the NLF is based only on the normal cases, as opposed to JAFROC1, in which abnormal *and* normal cases are used[78]. Since the AUC_{AFROC} is identical to the probability that a LL is given a higher confidence level than the highest rated NL in a case (using only normal cases in JAFROC2, while using abnormal *and* normal cases in JAFROC1), this alternative definition may also be used. The consequence of using only the normal cases to calculate the NLF (as in JAFROC2) is a loss of statistical power of the analysis compared to JAFROC1, and therefore JAFROC1 was recommended for a period[78]. However, JAFROC1 was discovered to be unreliable when the numbers of abnormal and normal cases are approximately equal and, therefore, the use of JAFROC1 is recommended only for datasets including exclusively abnormal cases[83]. The statistical method used in JAFROC analysis includes recalculating the FOM using jackknifing, i.e. recalculating the FOM repeatedly, excluding one case at a time in each calculation[84]. The resulting so-called pseudo value, PV, is obtained from the following equation[84]:

$$PV_{ijk} = n \cdot FOM_{ij} - (n-1) \cdot FOM_{ij(k)}$$

where n is the total number of cases, FOM_{ij} is the figure of merit for modality i and reader j when all cases are included in the calculation, and $FOM_{ij(k)}$ is the FOM for modality i and reader j when case k is excluded from the calculation. In order to perform the statistical analysis, JAFROC uses a mixed model analysis of variance (ANOVA), which calculates a 95% CI for the difference in FOM between the modalities and a p-value for rejecting the null hypothesis that the FOMs of the modalities are equal[84].

3.7 Simulated dose reduction

Patient images obtained at lower dose levels than the original dose level are desirable when investigating the effects of dose reduction on image quality. However, repeatedly examining patients will lead to unnecessary exposure to radiation, and problems may be encountered due to repositioning of the patient, or bias due to motion artifacts. For these reasons, simulated dose reduction can be used to create images that appear as if they had been acquired at lower doses. It is, however, important to use a simulation method that results in dose-reduced images with noise properties similar to those of images actually acquired at the reduced dose level. There are several methods of dose reduction, with varying degrees of sophistication. One simple method uses Gaussian distributed white noise[85], in which noise correlations are not taken into account. Other, more sophisticated, methods account for noise correlations[86,87]. However, one of these methods[86] has the limitation of being based on the assumption that there is little or no noise in the original clinical image, while the other has the has the limitation of using radially symmetric noise power spectrum (NPS)[87].

Another method of simulated dose reduction for conventional radiography, which takes the properties of the clinical image into account without assuming radial symmetry of the NPS, has been developed by Båth et al.[88]. In this method, information about the two-dimensional NPS at the original dose (D_{orig}) and at the desired lower dose (D_{sim}) is used to create a noise image, which, when added to the original image, gives the same NPS as an image actually acquired at the lower dose level. By using flat-field images (images containing only noise) acquired at various doses, a relationship is determined between the NPS and dose. Since it may be difficult to obtain flat-field images at the exact dose levels of a predetermined desired lower dose level and the original clinical setting dose level, the NPS at doses close to these levels, denoted D_1 and D_2, may be used. The NPS of the flat-field images acquired at D_1 and D_2 must, however, be scaled with the dose in order to correspond to the NPS that would have been obtained at D_{sim} and D_{orig}. The noise image can then be created by subtracting the scaled NPS at D_{orig} from the NPS at D_{sim}. The noise image must be corrected for dose variations in the original clinical image by adjusting the pixel values in the noise image according to the pixel values in the original image, giving larger relative variance in areas with lower dose relative to the mean image dose (and vice versa for high-dose areas). Before the noise image can be added to the original clinical image, the pixel values of the original clinical image must be scaled with the dose to correspond to pixel values at D_{sim}. The noise image can then be added to the clinical image, altering its appearance so that it is similar to

an image actually acquired at D_{sim}. The method of Båth et al. was developed for conventional radiography, and is therefore based on the assumption that the DQE does not differ between D_{orig} and D_2, or between D_{sim} and D_1. It is also assumed that the variation in DQE across the image need not be taken into account for the dose ranges used in conventional radiography.

3.7.1 Simulated dose reduction in tomosynthesis

The projection images collected during the tomosynthesis sweep are acquired at extremely low doses. For the detector used in the VolumeRAD system, the DQE decreases rapidly with decreasing dose at these low dose levels[89], and the DQE may vary between D_{sim} and D_1, as well as between D_{orig} and D_2. Also, the DQE may vary across the clinical projection image. In order to solve this problem, a method taking variations in DQE into account, suitable for creating dose-reduced images in tomosynthesis, was developed by Svalkvist and Båth[89]. In this method, the NPS is not only scaled with the dose, but also with the DQE in order to compensate for the differences in DQE across the clinical projection image, between D_{sim} and D_1, and between D_{orig} and D_2. Assuming that the shape of the DQE surface is constant across the dose variations in the clinical projection image, as well as between D_{orig} and D_2, and D_{sim} and D_1, the difference in DQE can be accounted for simply by scaling the NPS with the pixel variance, using the following relationship:

$$NPS(u,v)_{\mathrm{Im}_{noise}} = \left(\frac{\sigma_{D_{sim}}}{\sigma_{D_1}}\right)^2 NPS(u,v)_{D_1} - \left(\frac{D_{sim}}{D_{orig}}\right)^2 \left(\frac{\sigma_{D_{orig}}}{\sigma_{D_2}}\right)^2 NPS(u,v)_{D_2}$$

Where $NPS(u,v)_{\mathrm{Im}_{noise}}$ is the NPS of the noise image, $NPS(u,v)_{D_1}$ and $NPS(u,v)_{D_2}$ are the NPS of the flat-field images acquired at D_1 and D_2, respectively, and σ is the standard deviation in pixel value of the flat-field images acquired at D_{sim}, D_{orig}, D_1 and D_2. A relationship can be determined between mean pixel value and pixel variance using flat-field images acquired at various doses. This relationship is then used to determine the variance at the dose levels D_{sim}, D_{orig}, D_1 and D_2, in turn enabling the calculation of the NPS of the noise image. The dose simulation method is performed on the clinical projection images before the reconstruction of the tomosynthesis section images, and the projection images must, therefore, be available. Evaluation of this simulation method for tomosynthesis has shown that the dose

reduction results in images with very similar noise properties to those of an examination actually performed at a lower dose level[89].

4 MATERIALS AND METHODS

4.1 Overview of the Papers

Paper I describes a study in which conventional chest radiography and chest tomosynthesis were compared regarding the detection of pulmonary nodules. In the study described in Paper II, the effect of clinical experience of experienced thoracic radiologists was investigated by comparing readings of the same 89 chest tomosynthesis cases read in the study presented in Paper I to a new reading conducted after one year, during which chest tomosynthesis had become gradually more established and used at the department.

In the study described in Paper III, the effect of learning with feedback was investigated by asking experienced and inexperienced observers to read images from the same 89 cases again after a learning session. In the learning session, the observers were shown their assessments of a set of 25 new cases and the corresponding multidetector computed tomography (MDCT) images for comparison. The learning session was also used to identify potential pitfalls and to find solutions to these, and to propose image quality criteria suitable for chest tomosynthesis.

In the study described in Paper IV, images from a new group of patients were used to investigate the effect of dose reduction on the detection of pulmonary nodules in chest tomosynthesis. The observers read the original images from 86 cases (i.e. 100% of the standard setting tomosynthesis effective dose) and images simulated at lower doses of 70%, 32% and 12% of the standard setting effective dose.

4.2 Examinations

4.2.1 Conventional chest radiography

The conventional chest radiography imaging system used at the Department of Radiology at Sahlgrenska University Hospital, is a Definium 8000 system (GE Healthcare, Chalfont St Giles, UK), with a CsI flat-panel detector. The standard protocol consisted of a PA and a LAT projection, and the patient was examined in an upright position. The source-to-image distance was

180 cm. Tube voltages of 125 kV and 140 kV were used for the PA and the LAT projections, respectively. The effective dose to a standard-sized patient (70 kg, 170 cm), was approximately 0.05 mSv for the entire standard examination[17]. These conventional chest radiography images were used in Study I, in which conventional chest radiography was compared with chest tomosynthesis.

4.2.2 Chest tomosynthesis

The same equipment was used for the chest tomosynthesis imaging as for the conventional chest radiography examinations, except for the additional software and computer-controlled tube mover enabling the tomosynthesis functionality (VolumeRAD; GE Healthcare, Chalfont St Giles, UK), i.e. the vertical sweeping motion of the X-ray tube and the reconstruction algorithms, used for the acquisition of the tomosynthesis images. The tube movement covered an angle from –17.5 to +17.5 degrees, and exposures were made between –15 and +15 degrees, while the detector was stationary. The required tube current was determined from a scout view (i.e. a PA projection). The tube load for the scout view was multiplied by a factor of 10 and divided equally between the 60 projection images of the tomosynthesis sweep and rounded down to the closest mAs step on the Renard scale. For very thin patients, it was impossible to adapt the tube current correctly, since the tube was unable to produce loads smaller than 0.25 mAs. A tube voltage of 120 kV was used for the tomosynthesis examination, and the patients were examined in the PA projection position and instructed to hold their breath during the sweep. Each examination resulted in approximately 60 section images of the volume examined, with a reconstruction interval of 4 mm (beta version of VolumeRAD) or 5 mm (VolumeRAD) without overlap. The effective dose to a standard-sized patient (70 kg, 170 cm), was approximately 0.13 mSv for the entire tomosynthesis examination (including the scout view)[17]. The beta version of the chest tomosynthesis product was used in Studies I-III, and the final commercially available product was used in Study IV and partly in Study III.

4.2.3 Multidetector computed tomography

MDCT images were acquired using a 16- or a 64-channel multidetector CT system (LightSpeed Pro 16 and LightSpeed VCT; GE Healthcare, Chalfont

St Giles, UK). The patients were examined according to the standard protocol at the Department of Radiology, using tube load modulation and a tube voltage of 140 kV. The original section image thickness was 1.25 mm in the 16-channel CT examinations, and 0.6 mm in the 64-channel CT examinations. Axial, sagittal and coronal images of the cases included in Studies I-III were reconstructed with thicknesses of 5, 4 and 4 mm, respectively; while axial images of the cases used in Study IV were reconstructed with thicknesses of 1.25 and 0.6 mm for the 16- and the 64-channel CT examinations, respectively. The effective dose for a chest MDCT examination was determined using an anthropomorphic phantom (Alderson Lung/Chest Phantom RS-320; Radiology Support Devices, Long Beach, CA, USA), representing an average male patient (73.5 kg, 175 cm), and found to be approximately 4 mSv.

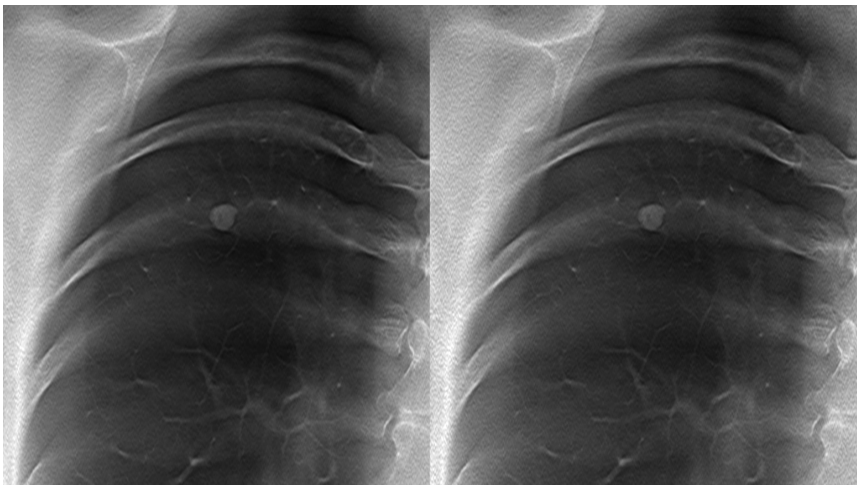
4.3 Data collection

In Studies I-III, 89 consecutive patients referred for CT of the chest were included. The patients were examined with MDCT and conventional chest radiography according to the standard protocol at the Department of Radiology, and also with chest tomosynthesis within a week of the CT examination for study purposes. The conventional chest radiography images were only used in Study I, whereas the MDCT and tomosynthesis images were used in Studies I-III. In Study III, 25 additional patients, examined with both tomosynthesis and MDCT, were included to provide learning material. In Study IV, 86 patients were included, the majority of which had malignant disease. They were examined with MDCT and with tomosynthesis on the same day, for study purposes. The exclusion criteria for all studies were: >20 nodules according to the MDCT images, breathing artifacts in the tomosynthesis images, and technical difficulties in properly displaying the images during reading sessions. In Study IV, additional exclusion criteria were introduced: >20 nodules according to the tomosynthesis images, extensive obscuring pathology or artifacts in the tomosynthesis images, the unavailability of raw tomosynthesis data or thin MDCT reconstructions, and the inability to visually separate nodules from each other in the tomosynthesis images. The Regional Ethical Review Board approved the studies, and all participants gave informed consent.

4.4 Dose reduction

In Study IV, simulated dose reduction was used to produce tomosynthesis images similar to those that would have been acquired at lower doses, using the method developed especially for tomosynthesis by Svalkvist and Båth[89]. The tomosynthesis images acquired at the standard exposure settings resulted in an estimated mean effective dose of 0.12 mSv to the patients included in the study. Images were simulated at doses of 70%, 32% and 12% of this effective dose. The 32% and the 12% dose levels corresponded to the effective dose used for the LAT and the PA projections in conventional chest radiography, respectively[17]. The 70% dose level was selected as an intermediate level. Flat-field images were acquired at various doses for all the projection angles used in a chest tomosynthesis examination using the clinically used tube voltage of 120 kV.

The ratio between the mean pixel value and the pixel variance in a ROI in each flat-field image was plotted against the pixel mean of the ROI, and a quartic polynomial was fitted to the data in order to obtain a relationship between the integrated DQE and dose, as described in Section 3.7.1. The NPS was determined at doses close to the standard dose level and close to the lower dose level, scaled with the dose, and also adjusted for variations in the DQE using the relationship between pixel variance and the mean pixel value. A noise image was created for each original tomosynthesis projection image, using the difference in NPS between the standard setting dose level and the lower dose level. The pixel values of the original clinical projection image were scaled with the dose, so that they corresponded to pixel values obtained at the lower dose level. The variation in DQE across the clinical projection image at various doses was taken into account by adjusting the pixel variances in the created noise image using the polynomial describing the relation between pixel mean and pixel variance obtained from the flat-field images. The noise image was then added to the scaled original image, and after all projection images for a patient had been simulated to represent the lower dose, the tomosynthesis images were reconstructed. This procedure was repeated for all three dose levels for every patient. Examples of the original and simulated images are shown in Figure 3.



a) b)

c) d)

Figure 3. Example of a tomosynthesis section image containing a 9 mm nodule at the original dose (100%) (a) and simulated images at 70% (b), 32% (c) and 12% (d) of the original dose.

4.5 Truth consensus panel

The truth consensus panel consisted of two experienced thoracic radiologists with 11 and 14 years of experience of thoracic radiology at the start of Study I. They used MDCT images of the patients to determine the ground truth regarding the existence of pulmonary nodules. They first read the MDCT images individually, before a joint session in which they reached consensus. In Study IV, computer-aided detection was used as a third observer, and one of the observers used computer-aided detection in her individual reading of the images before the joint session. The largest nodule dimension in axial reconstructions was used as a measure of size to categorize the nodules in all studies, as this is the standard procedure at the Department of Radiology.

4.6 The observers

The observers participating in the studies had varying degrees of experience regarding chest radiology and chest tomosynthesis. Table 2 provides information on the experience of the observers at the start of the first study in which they participated, together with information on the studies in which they participated as observers. In the individual papers, the observers are called Observers 1, 2, 3 etc., regardless of the notation used in previous papers. In order to avoid confusion, they are referred to in this dissertation as Observers A-G.

The four observers who participated in Study I (Observers A-D) were senior consultant radiologists with 11 to 20 years of experience in chest radiology at the start of the first study. One of these radiologists (Observer D) also participated in the truth consensus panel and was therefore included as an observer only in the first study, in which she participated in the detection study before analyzing the CT images as part of the consensus panel. The remaining three observers in Study I (Observers A-C) also participated in Studies II and III, and three inexperienced observers (Observers E-G) were also included in Study III. The reason for including the inexperienced observers was to enhance the discussions during the learning session included in Study III, the purpose being to encourage the experienced observers to express themselves in more detail, preventing them from leaving out important information that could have been disregarded if only experienced radiologists had been present. Another reason for including the inexperienced

observers was for studying the effects of learning with feedback on the performance of these observers. Observers A, B and E participated in Study IV.

Table 2. Observer experience in chest radiology and chest tomosynthesis at the start of the first study in which they participated, and their participation in each of the four studies.

Observer	Position	Clinical experience		Study			
		Chest radiology	Chest tomosynthesis	I	II	III	IV
A	Senior consultant thoracic radiologist	20 years	~6 months	x	x	x	x
B	Senior consultant thoracic radiologist	20 years	~6 months	x	x	x	x
C	Senior consultant thoracic radiologist	20 years	~6 months	x	x	x	-
D	Senior consultant thoracic radiologist	11 years	~6 months	x	-	-	-
E	Consultant radiologist	~1 year	None	-	-	x	x
F	Radiology resident	~3 months	Limited (<3 months)	-	-	x	-
G	Medical physicist	-	None	-	-	x	-

4.7 Detection studies

All detection studies were conducted under the FROC paradigm[77]. The observers were shown images (PA and LAT projections in the case of conventional chest radiography) or image stacks (in the case of chest tomosynthesis) and were asked to mark suspected pulmonary nodules and rate their confidence of the presence of a nodule on a 4-point rating scale, 4 representing the most strict criterion (definitely a nodule), 3 and 2 intermediate criteria (probably and possibly a nodule, respectively) and 1 the most lax criterion

(probably *not* a nodule). In Study I, the software DicomWorks[90] was used to record the coordinates and ratings, and a standard random number generator was used to create an individual reading order for each observer. In the other studies, the software Viewer for Digital Evaluation of X-ray images (ViewDEX)[91–94] was used to show the images and to record the marks and ratings associated with them. In ViewDEX, the cases are shown in a unique random order to each observer, and the coordinates of marks and their related ratings are recorded automatically. In all detection studies, the observers were instructed to mark the nodules at the nodule center and only in the image in which it was most prominently visible. In Study IV, the observers were also instructed to measure the longest dimension of the nodules in the image plane in order to enable categorization according to size. Before each study, the observers were instructed to go through a demonstration set, containing cases of varying degrees of difficulty included in the up-coming detection study, in order to adjust their confidence levels to the task. A learning set containing additional cases with marked example nodules of various sizes simulated at the lower dose level of the following session was also displayed in Study IV together with the same cases at the 100% dose level. The observers were allowed to use pan and zoom functions, and to alter the window width and window level settings in all studies. The background light was kept at a low and constant level during all the studies. Medical-grade monitors, calibrated according to DICOM 3.0 part 14[95], were used. Figure 4 shows an example of a tomosynthesis image at the 100% dose level in ViewDEX together with the options available in Study IV.

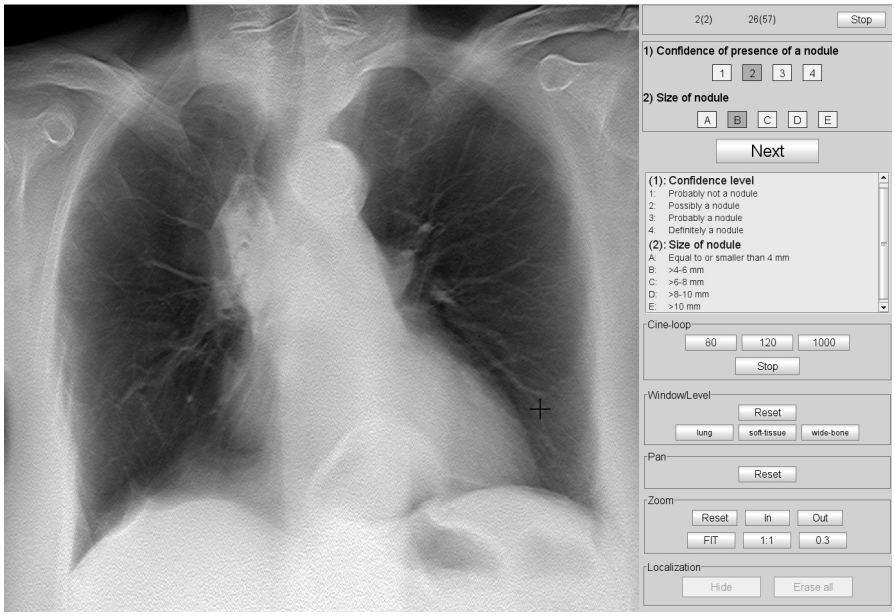


Figure 4. A tomosynthesis image at the 100% dose level shown in ViewDEX containing a measured, marked and rated suspected pulmonary nodule together with the options available in Study IV.

In Study I, in which chest tomosynthesis was compared to conventional chest radiography, the detection study was divided into two reading sessions separated by at least one week in order to avoid recall bias. The cases from both modalities (conventional radiography and tomosynthesis) were divided into two groups; each group consisting of half the conventional radiography examinations and half the tomosynthesis examinations, the same patient being present only once in each group. Two of the observers read one group first, and the other two observers read the other group first.

It was found that the NLF was much higher for tomosynthesis than for conventional radiography (see Section 5.1). One explanation of this was considered to be the lack of observer experience in chest tomosynthesis. Therefore, Study II was performed to investigate the effect of clinical experience on the observers' performance. In that study, two readings of the 89 cases were compared, one before and one after an additional year of clinical experience of tomosynthesis. At the start of the first reading, chest tomosynthesis had recently been introduced at the Department of Radiology and at that time the observers had very limited experience of the technique, as they were among the very first in the world to use chest tomosynthesis clinically. During the

year that followed, the situation gradually changed as chest tomosynthesis became established at the department, and more publications on chest tomosynthesis had begun to appear[7,96].

Study II showed that the additional clinical experience of tomosynthesis did not improve the performance of the observers (see Section 5.2). The reasons suggested for this were that the observers had not been given feedback on their analyses of the tomosynthesis cases, and that a reference truth (such as CT) had not been consistently available during their clinical work. The effect of learning with feedback was therefore investigated in Study III (see Section 4.8). In Studies II and III, only tomosynthesis cases were included, and these were therefore not divided into separate sessions. The images from the additional 25 cases used in Study III for the learning with feedback were read separately. In Study IV, in which the effect of radiation dose level on the detectability of pulmonary nodules in chest tomosynthesis was investigated, the images from the 86 cases included were shown in increasing order of radiation dose level. This study set up would probably generate less bias than a study randomizing the cases across dose levels, as it can be assumed that a higher dose level results in higher visibility of nodules. Thus, the 12% dose level images were shown in one reading session, followed by the 32%, the 70% and the 100% doses level sessions. The reading sessions were separated by at least two weeks.

4.8 Learning with feedback

In order to study the effect of learning with feedback, two readings of the images of the 89 tomosynthesis cases read in Studies I and II were compared. The readings were separated by a learning session, in which the observers were given feedback on their assessment of 25 additional cases. Before the learning session, the observers read the images of the 89 tomosynthesis cases to detect pulmonary nodules individually. For the observers that had participated in Study II, the data of the second reading in Study II were used as the data of the reading before the learning session in Study III, while for the remaining observers participating in Study III, a new reading was conducted. The observers then read the 25 additional tomosynthesis cases individually, and the marks made by all observers in this reading were displayed to them at an individual preparation session. They were then shown the tomosynthesis images in conjunction with the corresponding MDCT images for comparison in a joint learning session. The tomosynthesis images and CT images for each case were shown side by side on large screens. The observers were shown

every true nodule and every false positive mark together with the ratings of all observers, as shown in Figure 5. They were encouraged to explain the reasons for making false positive decisions and for low ratings of true nodules, in order to learn from their misinterpretations and to formulate potential pitfalls in chest tomosynthesis. Further, image quality criteria that had been suggested prior to the learning session by two experienced radiologists, based on the existing criteria for conventional chest radiography[97] and chest CT[98] given by the European commission, were developed in consensus using normal patients included in the 25 cases of the learning material as reference. After the learning session, the observers read the 89 tomosynthesis cases again, using the list of potential pitfalls as support, and the difference in detectability before and after the learning session was calculated for each observer individually.

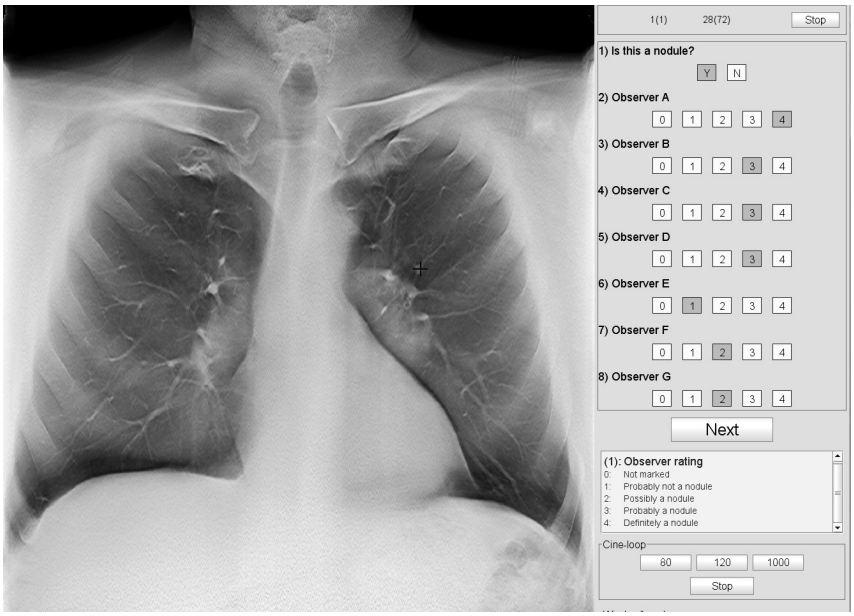


Figure 5. A tomosynthesis image including a true nodule, as shown in ViewDEX at the learning session, together with the consensus panel assessment (option 1) and the ratings of all observers.

4.9 Detectability measures and statistics

JAFROC was used as the principal method of statistical analysis in all studies. In Studies I, III and IV JAFROC2 was used, whereas in Study II JAFROC1 was used. JAFROC versions 2.3a, 1.0, 3g and 4.2 were used in Studies I, II, III and IV, respectively. The latest JAFROC version can be downloaded from the FROC home page[99]. The JAFROC FOM was calculated together with the 95% CI and the differences in FOM between modalities/readings/dose levels were tested for significance at the 5% level. In Studies I and II, FROC curves were calculated for each reading and each observer, and used to visualize differences between modalities or differences in interpretation between groups of tomosynthesis images. In Study I, the LLF was calculated for the four size categories of nodules according to the Fleischner Society guidelines[57] (≤4, >4-6, >6-8 and >8 mm).

5 RESULTS

5.1 Comparison between chest tomosynthesis and conventional chest radiography

The results of Study I, in which nodule detection in chest tomosynthesis and conventional chest radiography was compared, showed that tomosynthesis was superior to conventional radiography. The JAFROC2 FOMs for tomosynthesis and conventional radiography for the four observers in Study I are shown in Figure 6. The difference between the modalities was 0.24 (95% CI: 0.16, 0.33 and $p<0.0001$) in favor of tomosynthesis for the observer-averaged FOM.

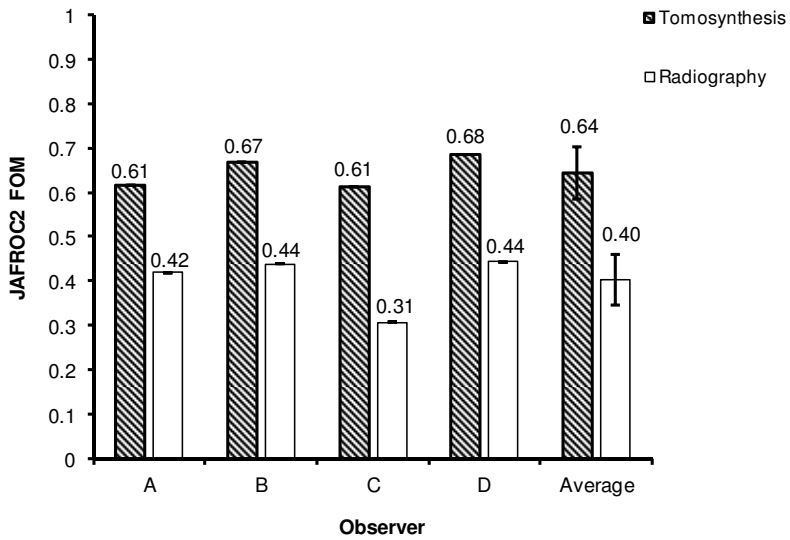


Figure 6. The JAFROC2 FOMs for chest tomosynthesis and conventional chest radiography for the four observers in Study I. The error bars represent 95% CIs.

The analysis according to size category showed that 39% of the smallest nodules (≤4 mm) were detected in chest tomosynthesis, and that the percentage increased with nodule size category to 83% for the largest nodules (>8 mm) (in this analysis, the most lax confidence level was used, and thus all marked nodules were included.) In total, the percentage of nodules detected using tomosynthesis was 56%, while for conventional radiography it was 16%. For nodules ≤8 mm the result for tomosynthesis was even more superior, as can be seen in Figure 7, since the fraction of nodules detected in chest radiography was very small for this size category.

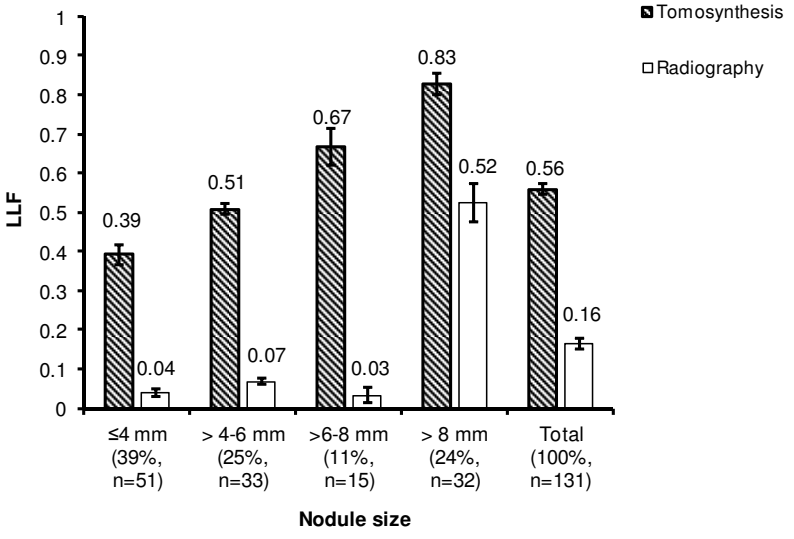


Figure 7. The observer-averaged LLF for nodules in each size category for chest tomosynthesis and conventional chest radiography in Study I, using the most lax confidence level. The error bars represent ± 1 standard error of the mean calculated from the LLF for each of the four observers.

The FROC curves for conventional chest radiography and chest tomosynthesis showed a substantial difference between the two modalities for each of the four observers (Figure 8). This is consistent with the results of the JAFROC

analysis, since at the same NLF, the LLF was substantially higher for tomosynthesis. The NLF was higher in tomosynthesis images than in conventional radiographic images for all observers and all thresholds, meaning that the observers made more NLs (false positive marks) at a given confidence rating in tomosynthesis.

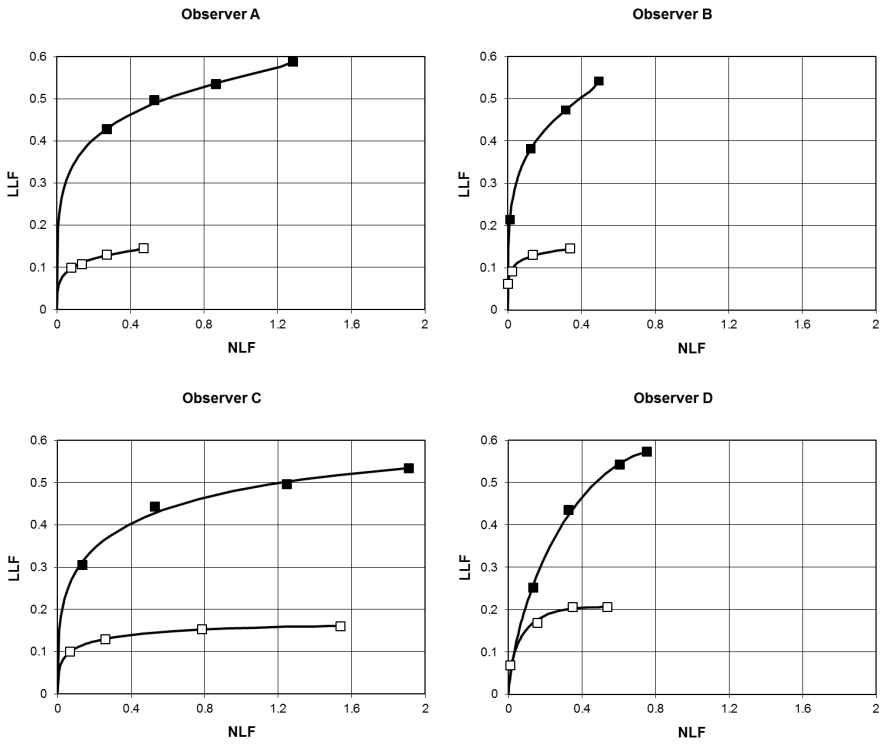


Figure 8. FROC curves for chest tomosynthesis (filled squares) and conventional chest radiography (open squares) for the four observers in Study I.

5.2 Learning effects in chest tomosynthesis

The investigation of learning effects after additional clinical experience of chest tomosynthesis in Study II showed no statistically significant differences between the two readings for the three experienced observers (p=0.91). The

JAFROC1 FOMs of the reading of the 89 chest tomosynthesis cases at baseline (when tomosynthesis had only very recently been implemented) and the reading after a year (when the observers had additional clinical experience of the technique and tomosynthesis had been established in clinical use) are shown in Figure 9.

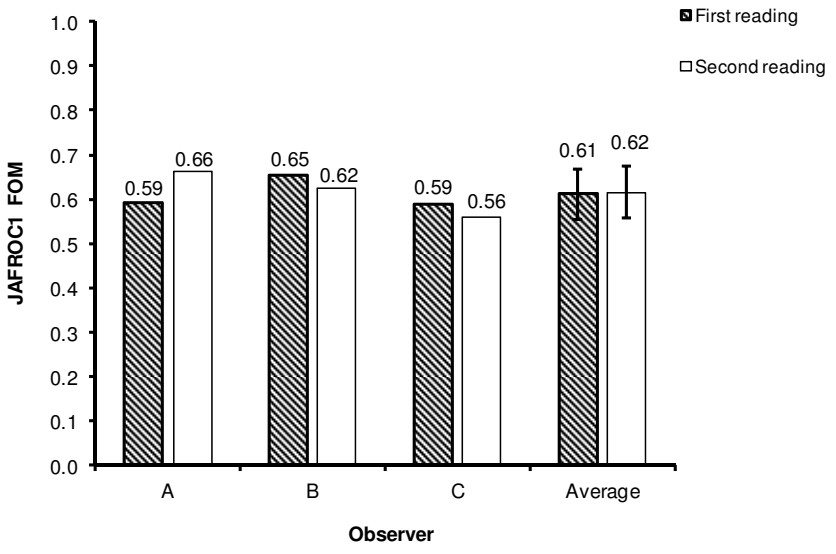


Figure 9. The JAFROC1 FOMs for the first and the second readings of the 89 chest tomosynthesis cases for the three experienced observers in Study II. The error bars represent 95% CIs.

The FROC curves for the first and the second readings of the 89 chest tomosynthesis cases for the three observers showed that there was a tendency for the observers to alter their confidence levels between the readings (Figure 10). The observers were more cautious during the second reading, leading to fewer NLs and fewer LLs. Although the extension of the curves to the right differed between the readings, the curves coincide reasonably well, showing that the observers operated along similar curves for both readings. Their performance therefore remained unchanged after clinical experience, although the number of NLs and LLs decreased.

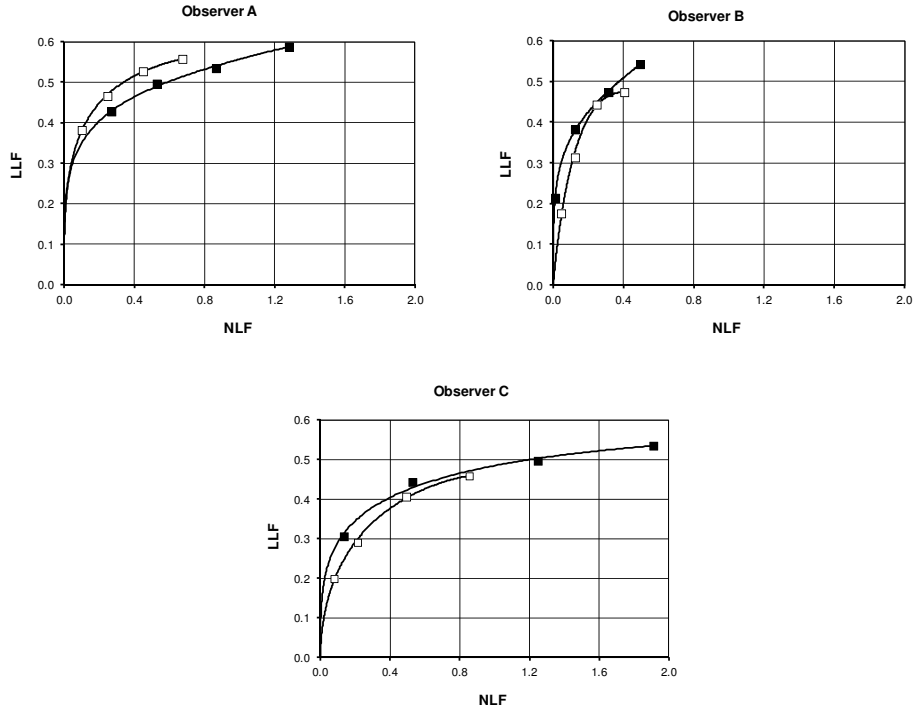


Figure 10. FROC curves for the first (filled squares) and the second reading (open squares) of the 89 chest tomosynthesis cases for the three experienced observers in Study II.

In the study on the effects of learning with feedback on the detectability of pulmonary nodules (Study III), no statistically significant differences were found between the reading before and the reading after the learning session, for the four observers most experienced in chest tomosynthesis. However, statistically significant differences were found between the two readings of the two observers with least experience of tomosynthesis, i.e., the consultant radiologist (Observer E) and the medical physicist (Observer G). The differences between the readings were 0.08 ($p<0.01$) and 0.14 ($p<0.01$) for Observer E and Observer G, respectively (Figure 11).

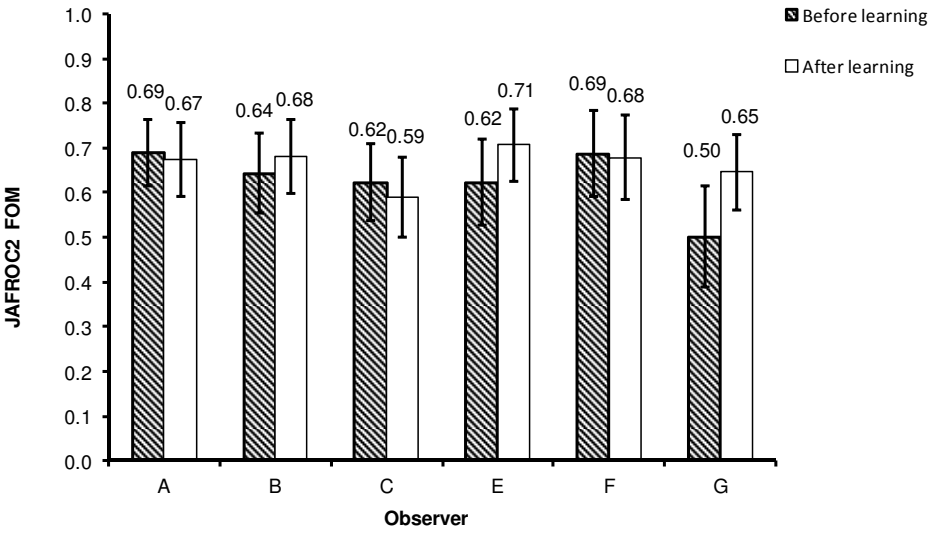


Figure 11. The JAFROC2 FOM for the six observers before and after the learning session in Study III. The error bars represent 95% CIs.

5.3 Image quality criteria and potential pitfalls in chest tomosynthesis

The image quality criteria developed during the learning-with-feedback session in Study III are presented in Table 3. The image criteria related to the positioning of the patient are based on the European quality criteria for conventional chest radiography[97], and the criteria related to the visibility of anatomic structures are based on the European quality criteria for chest CT[98]. Note that the criterion "reproduction of the whole rib cage" refers not only to the positioning of the patient, as is the case in conventional chest radiography, but also includes the demand for reconstruction of the entire volume in chest tomosynthesis.

Table 3. Proposed positioning and image quality criteria for chest tomosynthesis.

Positioning criteria	Image quality criteria
1. Performed at full inspiration as assessed by the position of the diaphragm	1. Clear reproduction of the trachea, carina and main bronchi
2. Performed with suspended respiration as assessed by clear reproduction of the diaphragm	2. Clear reproduction of the lobar bronchi
3. Symmetrical reproduction of the thorax as assessed by adequate position of the spinous processes, carina and sternoclavicular joints	3. Clear reproduction of the large and medium-sized vessels
4. Medial borders of the scapulae positioned outside the lung fields	4. Clear reproduction of the small-sized vessels as seen 3 cm from the costopleural border
5. Reproduction of the whole rib cage	5. Clear reproduction of the interlobar fissures
	6. Reproduction of the paratracheal tissue
	7. Reproduction of the thoracic aorta

The potential pitfalls in chest tomosynthesis identified and compiled during the learning-with-feedback session in Study III are given in Table 4. The most common pitfalls were related to the difficulty in differentiating pulmonary nodules from pleural and subpleural structures close to the chest wall, especially anteriorly or posteriorly. These regions were recognized as difficult due to highly attenuating anatomy, such as ribs, which may overlap over- and underlying structures. The reason for the difficulties at these locations was thus the limited depth resolution of tomosynthesis.

Table 4. Suggestions for avoiding potential pitfalls regarding nodules in chest tomosynthesis.

False positives	False negatives
Subpleural and pleural changes may often be misinterpreted as nodules because of their proximity to pleural borders, where skeletal structures overlap anatomy and pathology. This may possibly be prevented by relating the location where the ribs are in focus to the position of the suspicious finding.	Nodules situated close to the pleural border may often be misinterpreted as pleural or subpleural changes, because skeletal structures may overlap nodules at such locations. This may possibly be prevented by relating the location where the ribs are in focus to the position of the suspicious finding.
Lymph nodes may sometimes be misinterpreted as nodules close to hilar and mediastinal node stations. Even though the probability is high that the structures are lymph nodes, it is not possible to characterize them.	Nodules located closely to vessels, especially at branching points, may appear as part of the vessel itself. These nodules are usually too small (<5 mm) to properly be distinguished from the vessel that they are close to.
Skeletal changes, including costochondral calfications, may be misinterpreted as nodules, especially those located posteriorly and anteriorly. This may possibly be prevented by relating the location where the skeletal structure is in focus to the position of the suspicious finding.	Very small nodules (2-3 mm): sometimes discharged by radiologists as unspecific findings. It is important to bear in mind that small nodules may be very well depicted with tomosynthesis.

5.4 Effect of dose reduction in chest tomosynthesis

The results of the effect of dose level on nodule detection in chest tomosynthesis investigated in Study IV, showed a JAFROC2 FOM of 0.45, 0.54, 0.55, 0.54, and for the 12%, 32%, 70%, and 100% dose levels, respectively. The differences in the performance of the observers between the 32%, 70% and 100% dose levels and the 12% dose level were 0.087 ($p<0.01$), 0.099 ($p<0.01$), and 0.093 ($p<0.01$), respectively. No significant differences were found between the other dose levels. There was, thus, no statistically significant difference in detectability when the standard setting radiation dose was decreased to a dose similar to that of a lateral projection radiograph (0.04 mSv), corresponding to the 32% dose level. The observer-averaged JAFROC2 FOMs for the three observers are presented in Figure 12.

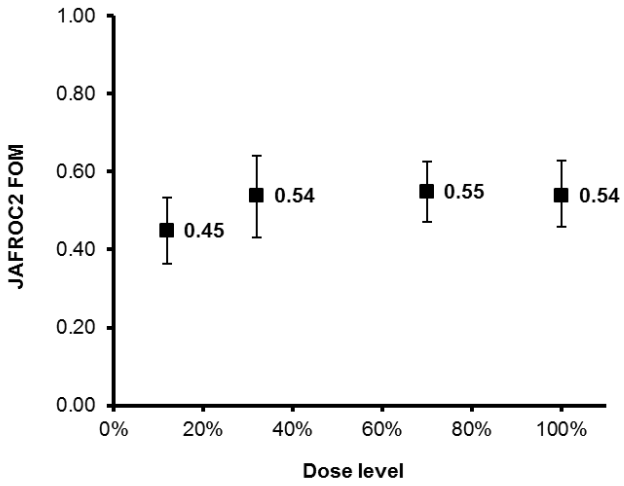


Figure 12. The observer-averaged JAFROC2 FOM for the original (100%) and the 70%, 32% and 12% dose levels in Study IV. The error bars represent 95% CIs.

6 DISCUSSION

6.1 Comparison between chest tomosynthesis and conventional chest radiography

In Study I, chest tomosynthesis and conventional chest radiography were compared regarding the detectability of pulmonary nodules. It was found that tomosynthesis was superior; the JAFROC FOM being 60% higher than that for conventional radiography. When studying nodules of various sizes, an especially large difference in LLF was found between the modalities for nodules ≤8 mm, although the LLF was higher for tomosynthesis for all size categories. Later studies have confirmed that tomosynthesis is superior to conventional radiography regarding the detectability of pulmonary nodules[18,55]. Yamada et al.[18] reported an increase in the JAFROC FOM of almost 40% compared to conventional radiography, while Kim et al.[55] reported a 16% increase in AUC_{ROC} for tomosynthesis performed at a reduced dose compared to the PA projection in conventional radiography for 4-10 mm nodules. Further, Dobbins et al[96] found that the visibility of already known nodules was approximately three times higher in tomosynthesis images than in the PA projection in conventional radiography, and a similar proportion of nodules was also visible according to the consensus panel in Study I. Other recent publications have also shown the potential of chest tomosynthesis compared to conventional radiography in the cases of mycobacterial disease[100], metastasis resulting from colorectal cancer[54], pulmonary lesions[101], pleuropulmonary disease[102], pulmonary emphysema[19], and cystic fibrosis[103,104].

As can be seen in the FROC curves in Figure 8, the NLF for chest tomosynthesis was higher than for conventional chest radiography for all observers, meaning that the observers erroneously marked a larger number of normal structures in tomosynthesis. It may thus appear that the performance of the observers in tomosynthesis was poorer than in conventional radiography. However, in order to correctly assess the detectability of a modality as inferior or superior to another, both the LLs and the NLs must be taken into account. The observers in Study I actually made a higher number of LLs in the tomosynthesis images, and at the same value of LLF the value of NLF was actually lower for tomosynthesis than for conventional radiography. The higher NLF for tomosynthesis is thus not a consequence of poorer performance of the observers, but may be a result of the observers using the confi-

dence levels differently in tomosynthesis and conventional radiography. Also, it is possible that tomosynthesis render a larger number of both LLs and NLs because the examination results in more images and reveals more suspicious structures.

6.2 Learning effects in chest tomosynthesis

The effects of learning on the detectability of pulmonary nodules in chest tomosynthesis were investigated in Studies II and III. No statistically significant differences were found among the experienced thoracic radiologists investigated in Study II between the two readings of the same 89 cases before and after the additional year of experience, during which chest tomosynthesis had become clinically established at the department. A possible explanation of this may be that the observers had already reached a level within the initial six months of using tomosynthesis from which they could not improve further. This assumption was strengthened by the results of Study III, where no statistically significant differences could be found for these observers after learning with feedback. However, the FROC curves shown in Figure 10 differed between the readings before and after additional clinical experience. Since the curves overlap each other, this difference is not due to a change in performance, but is rather an effect of the observers using the confidence levels differently in the two readings. This may be partly explained by the fact that the observers were aware of the large number of NLs in tomosynthesis in Study I. Thus, attempts to decrease the number of NLs in the second reading resulted in a corresponding decrease in the number of LLs. This demonstrates the importance of analyzing the detectability of lesions, in which sensitivity and specificity are related to each other, rather than using these fractions separately; the detectability being more difficult to alter, while the fractions might more easily change with the attitude of the observer although the performance remains unchanged.

For the observers with least experience in chest tomosynthesis in Study III (Observers E and G), statistically significant differences were found between their performance before and after learning with feedback, enabling these observers to reach a similar level of performance to the experienced observers. It may seem surprising that the performance of Observer F, the resident, did not improve after learning with feedback, since this observer had less experience in general radiology than Observer E, the consultant radiologist. However, Observer F had some experience in chest tomosynthesis during her

3 months' experience in thoracic radiology, while Observer E had no experience at all of the technique. It is thus plausible that observers, in terms of detecting pulmonary nodules, may reach a high level of performance by either a relatively short period of assessing chest tomosynthesis images, or by learning with feedback.

The learning-with-feedback session was conducted in close collaboration with researchers from the Department of Education, Communication and Learning at the University of Gothenburg, Sweden, who video recorded the learning session. In order to investigate how experienced radiologists analyze images obtained with the new modality, and how the session design affects the development of skills of the inexperienced observers, the conversations and gestures of the observers during the session were analyzed. It was seen that the experienced radiologists were able to quickly draw conclusions from the images, based on their previous knowledge in the field of medical imaging[105]. In the case of the inexperienced observers, it was concluded that the public display of the individual assessments of the chest tomosynthesis images, the diversity of experience among the participants and the fact that CT images and tomosynthesis images were displayed side by side for comparison might have been important factors leading to improvement[106]. These observations agree well with the conclusions of Papers II and III, stating that experienced radiologists may easily adapt to the analysis of tomosynthesis images, and that the learning session might be useful for inexperienced radiologists to help them learn how to analyze tomosynthesis images.

The most experienced observers analyzed the same 89 chest tomosynthesis cases three times, and it may be anticipated that this might have led to recall bias. However, since the readings were separated by at least one month, and since the observers did not know which nodules were true according to the consensus panel, it could be argued that the use of these cases several times did not lead to any bias. Also, there was no improvement in performance after several readings of the cases, which should have been the case, had the observers benefited from repeatedly analyzing the images.

6.3 Image quality criteria and potential pitfalls in chest tomosynthesis

At the learning session described in Paper III, potential pitfalls in chest tomosynthesis were compiled and image quality criteria were developed. The most common pitfall was found to be the result of the limited depth resolution in tomosynthesis. False positive marks and false negatives were often made near the pleural border, especially posteriorly and anteriorly when highly attenuating structures such as the ribs were perpendicular to the direction of the radiation field. Knowing the location of a structure close to the pleural border is of the uttermost importance, as the clinical significance of a structure located in the parenchyma is higher than that of a structure situated in the pleura. Although it is difficult to distinguish between a parenchymal and subpleural structure in chest tomosynthesis, mistakes could be avoided by relating the location where a suspected nodule is in focus to the location where structures outside the parenchyma, closest to the nodule (for example ribs), are in focus. The suggestions for avoiding pitfalls reported in Paper III were primarily intended for radiologists with little or no experience of chest tomosynthesis, such as Observer E, who might have benefited from these instructions when analyzing the 89 cases after learning with feedback.

The regions close to the pleura have also been reported to cause the greatest problems in chest tomosynthesis by others. Yamada et al. noted that the majority of the missed nodules in tomosynthesis were located in the subpleural region[18], and in a series of studies Quaia et al. found that lesions that were misinterpreted in tomosynthesis images were often located anteriorly or posteriorly, close to the thoracic wall[56,101,107,108].

The opinion of the radiologists participating in Study III was that the proportion of pleural and subpleural changes in the 25 cases used for learning with feedback was higher than normal. As a consequence, the problem of location close to the pleura may have been overestimated in the formulation of the pitfalls, as these were based on the analysis of these 25 cases. However, this pitfall has been identified as common also by others[18,56,101,107,108], as mentioned above, thus confirming the conclusion in Paper III. As the number of cases was small, only a limited number of pathologies may have been included, and all the potential pitfalls in tomosynthesis may not have been identified. The list of pitfalls may, therefore, be extended in the future.

The process of drawing up the list of potential pitfalls was analyzed by Lymer et al.[109] using the video recordings of the learning session. The con-

versations between the participants, when condensing the difficulties into different categories of pitfalls, were analyzed. This analysis revealed that the observers gradually became familiar with the various types of pitfalls, finally recognizing them very quickly and referring to them as something obvious. The discussions seem to have rendered a new awareness of pitfalls, although this was not manifested as an improvement in the performance of the experienced radiologists. It is therefore plausible that these observers, although they had already reached their peak performance regarding the detection of nodules in tomosynthesis, nevertheless lacked experience in communicating their interpretations of the new modality, and thus the learning session could, in this sense, also have been beneficial for the experienced observers.

During the learning session, suggested quality criteria for chest tomosynthesis, based on the European quality criteria for conventional chest radiography[97] and thoracic CT[98], were discussed and evaluated using normal tomosynthesis cases. The criteria were not subject to such thorough investigations as in the development of the European quality criteria[97,98], and they may therefore be further improved if necessary. Nonetheless, they should still be useful in their current form for the optimization of tomosynthesis examinations, as has also been pointed out by Chou et al.[110].

6.4 Dose reduction in chest tomosynthesis

Paper IV describes the study in which the effect of dose level on the detectability of pulmonary nodules was investigated. The results indicate that a substantial dose reduction from the standard clinical setting may be possible. No statistically significant difference was found between the 32% dose level (corresponding to an effective dose of 0.04 mSv for a standard-sized patient) and the original standard setting giving an effective dose of 0.12 mSv. However, a statistically significant difference was found between the 12% dose level and the original dose.

It should be emphasized that a lack of evidence of a significant difference is not evidence of a lack of difference. Even so, the potential differences in FOM between the higher dose levels are probably small considering the relatively large sample size. Similar results have been reported by Hwang et al.[111], who found no statistically significant differences in the detectability of nodules >4 mm in a phantom between the original dose of 0.14 mSv and chest tomosynthesis performed at a reduced dose of 0.06 mSv. Also, Kim et

al.[100] reported that chest tomosynthesis performed at an effective dose of 0.05 mSv was statistically significantly superior to the PA projection in conventional radiography for the detection of lesions resulting from pulmonary mycobacterial disease.

Implementation of the dose reduction in the VolumeRAD system, as suggested by the results of Study IV, may be difficult without changing the beam quality. The automatic exposure control cannot adapt the dose correctly to thin patients even at the standard dose setting due to the limitation on the minimum output of 0.25 mAs per projection. However, a reduction in dose may be possible if the tube voltage is lowered in combination with additional copper filtration, as conducted by Hwang et al.[111] and Kim et al.[100].

Because the DQE decreases with the dose for the extremely low doses used in the VolumeRAD tomosynthesis system[89], the dose reduced images will appear as being acquired at an even lower dose than that chosen for the dose reduction simulation. For the 32% dose level, for example, the noise level will correspond to an examination acquired at ~25% of the original dose had the DQE remained constant at low doses. It may, therefore, be possible to lower the dose even further than indicated by the results of the dose reduction study if the DQE could be increased for low doses.

Before each detection study performed in study IV, the observers analyzed a set of demonstration cases for each dose level in order to adjust their confidence levels. As a consequence, the meaning of the confidence levels for each dose level is unique, and the NLFs and the LLFs at different dose levels cannot be compared separately. However, it is appropriate to compare the JAFROC FOMs at different dose levels, since this measure includes both the NLF and the LLF, and should thus be independent of the meaning of the confidence levels at the various dose levels.

6.5 General discussion

In all four studies, chest CT was used as the reference for determining the truth. However, although it is superior to chest tomosynthesis and conventional chest radiography regarding the detection of lesions, CT as an imaging modality does not provide definitive information on the true nature of lesions. Invasive procedures, such as biopsy, are an option, but such methods are also associated with uncertainties, and a certain risk to the patient. Since the stud-

ies described here were based on comparisons between modalities, readings or dose levels, it is, however, reasonable to believe that the results were unaffected by the lack of an ideal truth.

The reference method differed between the studies, as the truth in Study IV was determined using sub-mm CT reconstructions and computer-aided detection as a third observer, while in Study I-III thicker CT reconstructions were used. This might explain the finding that the 86 cases used in Study IV were deemed to have a higher fraction of small nodules, according to the consensus panel, than the 89 cases used in the other studies. As a consequence, the JAFROC FOMs reported in Paper IV were lower than those reported in the previous papers, even for the 100% dose level, despite the fact that only the nodules visible in the tomosynthesis images according to the consensus panel were included in the analysis in Study IV. Because of this, the results of Study IV cannot be compared with those from the other studies. Further, the JAFROC FOM was low in all studies. However, this does not reflect the actual difference in detectability between tomosynthesis and CT, since the observers did not analyze the CT images. If they had, the FOM for CT would most probably have been <1. Also, the low FOM should be viewed in light of the fact that the JAFROC FOM ranges from 0 to 1, in contrast to AUC_{ROC}, which ranges from 0.5 to 1.

In the studies presented here, FROC was used to obtain the observer performance data. The gold standard in observer performance studies is, however, still the ROC methodology. Both paradigms have been accused of drawbacks[112–114], for example, of not sufficiently resembling the clinical situation: FROC methodology because it does not allow classification in terms of which patients need further work-up, and ROC since it does not include information on the location of the lesions. FROC has also been criticized for being dependent on a lesion location acceptance radius, and that markings may have different distributions in normal and abnormal cases. Regarding the acceptance radius, care should be taken to minimize the risk of erroneously classifying a mark as true or false. In the studies presented here, all dubious marks were therefore visually compared with the marks made by the consensus panel in order to ensure correct classification. In order to prevent the potential problem of differences in distributions between markings in normal and abnormal cases, JAFROC2, which uses only NL in normal images, has recently been recommended instead of JAFROC1. Regarding ROC, one of the most common arguments against the methodology is that no information on the location of the lesion is included, as this might lead to an observer reporting an abnormal case as abnormal based on the classification of a non-

lesion as a lesion, while in fact missing the true lesion. Another drawback is that it has lower statistical power than FROC, and the number of observers and/or cases may be small in medical imaging studies. The advantages of the ROC methodology are that it has been widely used, and therefore more rigorously tested, is more intuitively understood, and that the statistical issues associated with it are less problematic. However, the FROC methodology is gradually gaining greater acceptance, although it has been subject to repeated changes and may therefore be more difficult to grasp. The choice of method should depend on the purpose of the study. In the case of a disease that cannot be exactly located, ROC may be the appropriate choice, while for the detection of more delimited lesions, for example, pulmonary nodules, FROC may serve the purpose better.

Chest tomosynthesis is a relatively new modality and has not yet been thoroughly evaluated. Therefore, it is probably not yet used according to its true potentials. So far, it has primarily been used as a problem-solving tool when suspicious findings on chest radiographs have been difficult to assess[11,14]. The advantage of tomosynthesis when used as a problem solver is the potential to reduce the number of patients requiring CT. Since tomosynthesis can be performed in direct conjunction with the conventional radiography examination, rescheduling of patients for CT may be avoided, and consequently the radiation dose, the time required for examination and the financial cost of diagnosis can be reduced. Another advantage, of the utmost importance for the patient, is that of an early diagnosis. The clinical evaluations of tomosynthesis reported thus far support the opinion that the technique may be valuable as an addition to conventional radiography, as it has been shown that the use of tomosynthesis increases the diagnostic accuracy of conventional radiography and reduces the number of patients requiring CT after suspicious findings in conventional radiography[101,107,108].

6.6 Future perspectives

Potential scenarios for how tomosynthesis may be used in the future include the evaluation of suspicious lesions seen on conventional chest radiographs, as an additional or replacement form of examination in all or some patients undergoing conventional chest radiography, and for the follow-up of known nodules (i.e. replacing CT)[8]. So far, chest tomosynthesis has mainly been used as a problem-solving tool for suspicious findings in conventional chest radiography examinations[11,14], and since this use of the technique has been

recognized as beneficial, it is probable that chest tomosynthesis will continue to be used in this way, at least in the near future.

Although it has been discussed, the use of chest tomosynthesis as a replacement or additional mode of examination for all patients undergoing conventional chest radiography has not been implemented. As indicated in Paper IV, the dose used for chest tomosynthesis might at present be higher than that required to obtain images of diagnostic quality. Also according to the results of other studies, a reduction of the dose to a level in the vicinity of that used for a conventional LAT radiography examination may be possible[100,111]. The question of whether chest tomosynthesis can replace the lateral projection, and thus the whole conventional chest radiography examination (since the scout image is identical to the PA projection) is therefore an interesting one, but must be thoroughly evaluated, not least because tomosynthesis is more expensive and time consuming than conventional radiography. Also, and most importantly, it remains to be demonstrated that tomosynthesis at a reduced dose can provide information of the same, or better, diagnostic value as the lateral view in conventional radiography. Finally, the risk of higher numbers of false positive findings in tomosynthesis than in conventional radiography must be taken into consideration.

The use of chest tomosynthesis as a replacement for, or in addition to, conventional chest radiography may be appropriate in some patient groups, for example, high-risk patients. One such category is patients at risk of developing lung cancer. The use of conventional radiography in lung cancer screening has not been found to reduce lung cancer mortality[115]. However, CT at a reduced dose has been found to lead to a 20% reduction in lung cancer mortality compared to screening with conventional radiography[116]. Since it has been found that the detectability of nodules is higher in tomosynthesis than in conventional radiography (Study I and other studies[18,55]), and that tomosynthesis may be an alternative to CT for detection of artificial nodules[117] and metastases from colorectal cancer[54], it is possible that tomosynthesis may be an option for lung cancer screening, resulting in advantages associated with costs, workflow and radiation dose compared to CT. However, the benefit to the patient of tomosynthesis screening must be thoroughly investigated. Further, replacing conventional chest radiography by chest tomosynthesis has been recognized as suitable for staging of cystic fibrosis in children and young adults, and has therefore been suggested as replacement for conventional radiography for this patient category[118].

The use of chest tomosynthesis instead of CT may also be appropriate for selected patients. Tomosynthesis may be an interesting option for the follow-up of known nodules, since many lesions found by CT can be identified ret-

rospectively in tomosynthesis, according to e.g. Paper 1 and Dobbins et al.[96]. Recent studies have shown that the nodule size measurement in tomosynthesis images corresponds well with the values obtained from CT images[119,120], supporting the potential use of tomosynthesis for this purpose. However, as Dobbins and McAdams pointed out[8], there may be a risk of missing new small nodules in patients followed up with tomosynthesis instead of CT.

As always when a new imaging technique is implemented, it is necessary to optimize the examination protocols. For chest tomosynthesis, optimization may include finding the optimal tube voltage and filtering. The use of a lower tube voltage in the VolumeRAD system, in combination with additional filtering, may be beneficial as higher tube outputs could be generated at a lower tube voltage (eliminating the problem of the tube output limitation of 0.25 mAs for thin patients), without increasing the effective dose. This approach may also be useful if a dose reduction is to be implemented. An interesting topic for future studies would thus be the dependency of image quality and/or the detectability of lesions on tube voltage and filtering. The angular range and number of projections are other important parameters in the optimization of tomosynthesis examinations. One option for dose reduction may be to perform tomosynthesis in a narrower angle interval, or to acquire fewer projection images. However, this may have a negative influence on image quality because artifacts may arise when tomosynthesis is performed using an insufficient angle interval or number of projection images, or too large an interval between projections[42]. Nevertheless, the above mentioned factors must be investigated to find the parameters resulting in the optimal performance of the tomosynthesis system.

As tomosynthesis lies between CT and radiography, regarding image quality, radiation dose and cost, it has some of the advantages and disadvantages of both techniques. Therefore, the addition of chest tomosynthesis leaves room for many options. The most important issue regarding the future use of chest tomosynthesis is, however, whether the patients examined with the technique will actually benefit from tomosynthesis instead of, or in addition to, the standard procedures. This question must be answered before the optimal use of chest tomosynthesis can be established. The work presented in this dissertation has hopefully contributed to some aspects of this quest.

7 CONCLUSIONS

The general findings presented in this dissertation indicate that chest tomosynthesis is a promising technique for improving chest radiography with only small increases in the radiation dose to the patient and the cost. Chest tomosynthesis is superior to conventional chest radiography for the detection of pulmonary nodules, but its limited depth resolution in comparison to CT may be problematic. Although chest tomosynthesis is a new technique, it appears that radiologists can easily learn to interpret the images. Further, chest tomosynthesis is already a low-dose examination, but it may be possible to reduce the dose further, making the technique even more interesting in thoracic imaging in the future.

The specific conclusions drawn from the studies presented in this dissertation are that:

- chest tomosynthesis is superior to conventional chest radiography regarding the detection of pulmonary nodules, especially nodules ≤8 mm;

- experienced thoracic radiologists with a short initial period of experience of chest tomosynthesis may benefit from the new modality, and inexperienced observers may reach a level of performance regarding the detection of pulmonary nodules similar to that of experienced observers after learning with feedback;

- the most common pitfalls in chest tomosynthesis are related to the limited depth resolution of the technique, which is especially troublesome in areas close to the thoracic wall anteriorly and posteriorly; and

- a substantial reduction from the default setting dose in the chest tomosynthesis VolumeRAD system may be possible without a considerable decrease the detectability of pulmonary nodules.

ACKNOWLEDGEMENTS

Först och främst vill jag tacka min huvudhandledare **Magnus Båth** för att du ha gett mig en så bra forskarutbildning. Tack för ditt stora engagemang, din noggrannhet samt ständiga tillgänglighet för frågor och bollande av idéer. Sist, men inte minst vill jag också tacka dig för att du, i mina stunder av uppgivenhet, visade dina allra bästa sidor genom att muntra upp mig och få mig att må bättre.

Jag vill också tacka mina bihandledare **Lars Gunnar Månsson** och **Åse Allansdotter Johnsson**. **Lars Gunnar**, för att du alltid lyckats komma med nyttiga infallsvinklar och för att du delat med dig av den stora klokhet som en gedigen erfarenhet för med sig. **Åse**, formellt blev du min handledare sent under doktorandutbildningen, men du var hela tiden som en handledare för mig. Tack för att du har gett mig radiologens perspektiv och kämpat igenom många facitsessioner tillsammans med mig med bibehållet gott humör.

Till alla medförfattare vill jag rikta ett varmt tack. **Jenny Vikgren**, tillsammans har vi genomgått den ibland mödosamma processen att göra facit, men det har ändå känts mer som ett nöje. **Susanne Kheddache, Marianne Boijsen, Agneta Flinck**, **Valeria Fisichella** och **Åsa Wiksell**: ni har varit fantastiska på att orka granska tunga material (med eller utan godis) och jag är glad att vårt samarbete har fungerat så bra. Ett stort tack till våra samarbetspartners i LET-studion, speciellt **Hans Rystedt, Jonas Ivarsson** och **Gustav Lymer.** Jag är tacksam för att ni (genom en lyckosam slump) kom in i vår grupp och bidrog med er expertis.

Till mina rumskamrater på Sahlgrenska, **Angelica Svalkvist** och **Kerstin Ledenius**: ni har bidragit till att göra det vardagliga doktorandlivet roligt och det har varit en lyx att alltid ha era klipska hjärnor och inte minst er vänskap nära till hands. Tack för att ni delat detta med mig.

Sune Svensson och **Markus Håkansson**, tack för att ni har varit till så stor hjälp när det gäller ViewDEX. Det hade varit mycket svårt att klara sig utan er.

Thank you to **Dev Chakraborty**, for always responding quickly to my emails asking for advice on the JAFROC analysis, statistics or study setup, even though you were on the other side of the world. Thank you also to **Gerhard Brunst**, for kindly answering my questions about the VolumeRAD system.

Till alla på diagnostisk strålningsfysik: tack för att ni har varit så trevliga och roliga att umgås med. Och för den hjälp ni alltid har erbjudit. Tack **Christina Söderman**, för att du såg till min granskningsstudie när jag var mammaledig och för att du underlättade mitt arbete på distans i Trollhättan. Tack **Jonny Hansson**, för att du aldrig har varit främmande för att prata om allt som rör granskningsstudier när det har behövts. Tack också **Patrik Sund**, för att du ofta släppte det du hade för händer för att hjälpa till med datorproblem.

Tack alla trevliga människor på Avdelningen för Radiofysik. Doktorandkompisar, jag saknar er redan. Tack för allt roligt vi haft tillsammans. Särskilt tack också till **Gunilla Adielsson** och **Eva Forsell-Aronsson** för effektiv hjälp med administrativa frågor.

Jag vill också tacka mina nya kollegor i NU-sjukvården **Johnny Kallin, Joel Larsson** och **Lea Sillfors Elverby,** samt **Timo Melakari,** för att ni varit så tillmötesgående och underlättat för mig att slutföra mina doktorandstudier. Ett extra tack till mina nya rumskamrater **Joel** och **Johnny**, för att ni ofta ställt upp som bollplank och, det får erkännas, ibland också som klagomur.

Och sist men inte minst: stort tack till familj och vänner. Ni har bidragit så mycket genom att förgylla livet och stöttat mig när allt inte varit lika roligt. Särskilt tack till mina svärföräldrar **Ingrid** och **Anders** och till min mamma **Lena** för att ni kommit så fort vi behövt hjälp på hemmafronten. Till sist vill jag uttrycka min tacksamhet mot min älskade **Örjan** och mina fantastiska pojkar, käraste **Manfred** och **Alfons**, för att ni fått mig att tänka på annat och för att ni alltid finns vid min sida. Att få vara med er är den största gåvan av alla!

This work was supported by grants from the Health and Medical Care Committee of the Region Västra Götaland, the King Gustav V Jubilee Clinic Cancer Research Foundation, the Swedish Federal Government under the LUA/ALF agreement, the Swedish Radiation Safety Authority, and the Swedish Research Council.

REFERENCES

1. Geitung, J. T., Skjærstad, L. M. & Göthlin, J. H. Clinical utility of chest roentgenograms. *Eur Radiol* **9,** 721–723 (1999).

2. Speets, A. M., van der Graaf, Y., Hoes, A. W., Kalmijn, S., Sachs, A. P., Rutten, M. J., Gratama, J. W., Montauban van Swijndregt, A. D. & Mali, W. P. Chest radiography in general practice: indications, diagnostic yield and consequences for patient management. *Br J Gen Pract* **56,** 574–578 (2006).

3. Schueller, G., Matzek, W., Kalhs, P. & Schaefer-Prokop, C. Pulmonary infections in the late period after allogeneic bone marrow transplantation: chest radiography versus computed tomography. *Eur J Radiol* **53,** 489–494 (2005).

4. Blanchon, T., Bréchot, J.-M., Grenier, P. A., Ferretti, G. R., Lemarié, E., Milleron, B., Chagué, D., Laurent, F., Martinet, Y., Beigelman-Aubry, C., Blanchon, F., Revel, M.-P., Friard, S., Rémy-Jardin, M., Vasile, M., Santelmo, N., Lecalier, A., Lefébure, P., Moro-Sibilot, D., Breton, J.-L., Carette, M.-F., Brambilla, C., Fournel, F., Kieffer, A., Frija, G. & Flahault, A. Baseline results of the Depiscan study: a French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR). *Lung cancer* **58,** 50–58 (2007).

5. Elmali, M., Baydin, A., Nural, M. S., Arslan, B., Ceyhan, M. & Gürmen, N. Lung parenchymal injury and its frequency in blunt thoracic trauma: the diagnostic value of chest radiography and thoracic CT. *Diagn Interv Radiol* **13,** 179–182 (2007).

6. Dobbins III, J. T. & Godfrey, D. J. Digital x-ray tomosynthesis: current state of the art and clinical potential. *Phys Med Biol* **48,** R65–R106 (2003).

7. Dobbins III, J. T., McAdams, H. P., Godfrey, D. J. & Li, C. M. Digital tomosynthesis of the chest. *J Thorac Imaging* **23,** 86–92 (2008).

8. Dobbins III, J. T. & McAdams, H. P. Chest tomosynthesis: technical principles and clinical update. *Eur J Radiol* **72,** 244–251 (2009).

9. Dobbins III, J. T. Tomosynthesis imaging: at a translational crossroads. *Med Phys* **36,** 1956–1967 (2009).

10. Tingberg, A. X-ray tomosynthesis: a review of its use for breast and chest imaging. *Radiat Prot Dosimetry* **139,** 100–107 (2010).

11. Johnsson, Å. A., Vikgren, J. & Båth, M. Chest tomosynthesis: technical and clinical perspectives. *Semin Respir Crit Care Med* **35,** 17–26 (2014).

12. Hansell, D. M., Bankier, A. A., McLoud, T. C., Müller, N. L. & Remy, J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* **246,** 697–722 (2008).

13. The 2007 Recommendations of the International Commission on Radiological Protection. ICRP publication 103. *Ann ICRP* **37,** 1–332 (2007).

14. Johnsson, Å. A., Vikgren, J., Svalkvist, A., Zachrisson, S., Flinck, A., Boijsen, M., Kheddache, S., Månsson, L. G. & Båth, M. Overview of two years of clinical experience of chest tomosynthesis at Sahlgrenska University Hospital. *Radiat Prot Dosimetry* **139,** 124–129 (2010).

15. Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D. & Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available at: http://globocan.iarc.fr. Accessed: December 16, 2013.

16. Sabol, J. M. A Monte Carlo estimation of effective dose in chest tomosynthesis. *Med Phys* **36,** 5480–5487 (2009).

17. Båth, M., Svalkvist, A., von Wrangel, A., Rismyhr-Olsson, H. & Cederblad, Å. Effective dose to patients from chest examinations with tomosynthesis. *Radiat Prot Dosimetry* **139,** 153–158 (2010).

18. Yamada, Y., Jinzaki, M., Hasegawa, I., Shiomi, E., Sugiura, H., Abe, T., Sato, Y., Kuribayashi, S. & Ogawa, K. Fast scanning tomosynthesis for the detection of pulmonary nodules: diagnostic performance compared with chest radiography, using multidetector-

row computed tomography as the reference. *Invest Radiol* **46,** 471–477 (2011).

19. Yamada, Y., Jinzaki, M., Hashimoto, M., Shiomi, E., Abe, T., Kuribayashi, S. & Ogawa, K. Tomosynthesis for the early detection of pulmonary emphysema: diagnostic performance compared with chest radiography, using multidetector computed tomography as reference. *Eur Radiol* **23,** 2118–2126 (2013).

20. Mettler, F. A., Huda, W., Yoshizumi, T. T. & Mahesh, M. Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology* **248,** 254–263 (2008).

21. Samei, E., Flynn, M. J. & Eyler, W. R. Detection of subtle lung nodules: relative influence of quantum and anatomic noise on chest radiographs. *Radiology* **213,** 727–734 (1999).

22. Samei, E., Flynn, M. J., Peterson, E. & Eyler, W. R. Subtle lung nodules: influence of local anatomic variations on detection. *Radiology* **228,** 76–84 (2003).

23. Båth, M., Håkansson, M., Börjesson, S., Kheddache, S., Grahn, A., Ruschin, M., Tingberg, A., Mattsson, S. & Månsson, L. G. Nodule detection in digital chest radiography: introduction to the RADIUS chest trial. *Radiat Prot Dosimetry* **114,** 85–91 (2005).

24. Håkansson, M., Båth, M., Börjesson, S., Kheddache, S., Flinck, A., Ullman, G. & Månsson, L. G. Nodule detection in digital chest radiography: effect of nodule location. *Radiat Prot Dosimetry* **114,** 92–96 (2005).

25. Håkansson, M., Båth, M., Börjesson, S., Kheddache, S., Johnsson, Å. A. & Månsson, L. G. Nodule detection in digital chest radiography: Effect of system noise. *Radiat Prot Dosimetry* **114,** 97–101 (2005).

26. Båth, M., Håkansson, M., Börjesson, S., Kheddache, S., Grahn, A., Bouchud, F. O., Verdun, F. R. & Månsson, L. G. Nodule detection in digital chest radiography: part of image background acting as pure noise. *Radiat Prot Dosimetry* **114,** 102–108 (2005).

27. Båth, M., Håkansson, M., Börjesson, S., Hoeschen, C., Tischenko, O., Kheddache, S., Vikgren, J. & Månsson, L. G. Nodule detection in

digital chest radiography: effect of anatomical noise. *Radiat Prot Dosimetry* **114,** 109–113 (2005).

28. Håkansson, M., Båth, M., Börjesson, S., Kheddache, S., Grahn, A., Ruschin, M., Tingberg, A., Mattsson, S. & Månsson, L. G. Nodule detection in digital chest radiography: summary of the RADIUS chest trial. *Radiat Prot Dosimetry* **114,** 114–120 (2005).

29. Medicare Payment Advisory Commission. A Data Book: Healthcare Spending and the Medicare Program. June 2012. Available at: http://www.medpac.gov/documents/Jun12DataBookEntireReport.pdf. Accessed January 7, 2014.

30. The Nordic Radiation Protection Co-operation, "statement concerning the increased use of computed tomography in the Nordic countries" available at: http://www.stralsakerhetsmyndigheten.se/Global/Pressmeddelanden/2012/justification_statement_nordic_2012.pdf.

31. Sodickson, A., Baeyens, P. F., Andriole, K. P., Prevedello, L. M., Nawfel, R. D., Hanson, R. & Khorasani, R. Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. *Radiology* **251,** 175–184 (2009).

32. McCollough, C. H., Bruesewitz, M. R. & Kofler, J. M. CT dose reduction and dose management tools: overview of available options. *Radiographics* **26,** 503–513 (2006).

33. Pontana, F., Pagniez, J., Duhamel, A., Flohr, T., Faivre, J.-B., Murphy, C., Remy, J. & Remy-Jardin, M. Reduced-dose low-voltage chest CT angiography with sinogram-affirmed iterative reconstruction versus standard-dose filtered back projection. *Radiology* **267,** 609–618 (2013).

34. Katsura, M., Matsuda, I., Akahane, M., Yasaka, K., Hanaoka, S., Akai, H., Sato, J., Kunimatsu, A. & Ohtomo, K. Model-based iterative reconstruction technique for ultralow-dose chest CT comparison of pulmonary nodule detectability with the adaptive statistical iterative reconstruction technique. *Invest Radiol* **48,** 206–212 (2013).

35. Baumueller, S., Alkadh, H., Stolzmann, P., Frauenfelder, T., Goetti, R., Schertler, T., Plass, A., Falk, V., Feuchtner, G., Scheffel, H.,

Desbiolles, L. & Leschka, S. Computed tomography of the lung in the high-pitch mode: is breath holding still required? *Invest radiol* **46,** 240–245 (2011).

36. Bendaoud, S., Remy-Jardin, M., Wallaert, B., Deken, V., Duhamel, A., Faivre, J.-B. & Remy, J. Sequential versus volumetric computed tomography in the follow-up of chronic bronchopulmonary diseases: comparison of diagnostic information and radiation dose in 63 adults. *J Thorac Imaging* **26,** 190–195 (2011).

37. Baumueller, S., Winklehner, A., Karlo, C., Goetti, R., Flohr, T., Russi, E. W., Frauenfelder, T. & Alkadhi, H. Low-dose CT of the lung: potential value of iterative reconstructions. *Eur Radiol* **22,** 2597–2606 (2012).

38. Webb S., From the watching of shadows – the origins of radiological tomography, ed. A.J. Meadows. Bristol and New York: Adam Hilger (1990).

39. Grant, D. G. Tomosynthesis: a three-dimensional radiographic imaging technique. *IEEE Trans Biomed Eng* **19,** 20–28 (1972).

40. Ziedses des Plantes, B. G. Eine neue Methode zur Differenzierung in Röntgenographie (Planigraphies). *Acta Radiol* **13,** 182–192 (1932).

41. Garrison, B., Gant, D. G., Guier, W. H. & Johns, R. J. Three dimensional roentgenography. *Am J Roentgenol Radium Ther Nucl Med* **105,** 903–908 (1969).

42. Machida, H., Yuhara, T., Mori, T., Ueno, E., Moribe, Y. & Sabol, J. M. Optimizing parameters for flat-panel detector digital tomosynthesis. *Radiographics* 549–562 (2010).

43. Tingberg, A. & Zackrisson, S. Digital mammography and tomosynthesis for breast cancer diagnosis. *Expert Opin Med Diagn* **5,** 517–526 (2011).

44. Houssami, N. & Skaane, P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast* **22,** 101–108 (2013).

45. Kopans, D. B. Digital breast tomosynthesis from concept to clinical care. *AJR Am J Roentgenol* **202,** 299–308 (2014).

46. Skaane, P. Tomosynthesis in X-ray: proven additional value? *Eur J Radiol* **81S1,** S156–S157 (2012).

47. Houssami, N. & Zackrisson, S. Digital breast tomosynthesis: the future of mammography screening or much ado about nothing? *Expert Rev Med Devices* **10,** 583–585 (2013).

48. Tingberg, A., Förnvik, D., Mattsson, S., Svahn, T., Timberg, P. & Zackrisson, S. Breast cancer screening with tomosynthesis - initial experiences. *Radiat Prot Dosimetry* **147,** 180–183 (2011).

49. Ciatto, S., Houssami, N., Bernardi, D., Caumo, F., Pellegrini, M., Brunelli, S., Tuttobene, P., Bricolo, P., Fantò, C., Valentini, M., Montemezzi, S. & Macaskill, P. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet oncol* **14,** 583–589 (2013).

50. Skaane, P., Bandos, A. I., Gullien, R., Eben, E. B., Ekseth, U. & Haakenaasen, U. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* **267,** 47–56 (2013).

51. Canella, C., Philippe, P., Pansini, V., Salleron, J., Flipo, R.-M. & Cotten, A. Use of tomosynthesis for erosion evaluation in rheumatoid arthritic hands and wrists. *Radiology* **258,** 199–205 (2011).

52. Hayashi, D., Xu, L., Roemer, F. W., Hunter, D. J., Li, L., Katur, A. M. & Guermazi, A. Detection of osteophytes and subchondral cysts in the knee with use of tomosynthesis. *Radiology* **263,** 206–215 (2012).

53. Geijer, M., Börjesson, A. M. & Göthlin, J. H. Clinical utility of tomosynthesis in suspected scaphoid fracture. A pilot study. *Skeletal Radiol* **40,** 863–867 (2011).

54. Jung, H. N., Chung, M. J., Koo, J. H., Kim, H. C. & Lee, K. S. Digital tomosynthesis of the chest: utility for detection of lung metastasis in patients with colorectal cancer. *Clin Radiol* **67,** 232–238 (2012).

55. Kim, S. M., Chung, M. J., Lee, K. S., Kang, H., Song, I.-Y., Lee, E. J. & Hwang, H. S. Digital tomosynthesis of the thorax: the influence of

respiratory motion artifacts on lung nodule detection. *Acta Radiol* **54,** 634–639 (2013).

56. Quaia, E., Grisi, G., Baratella, E., Cuttin, R., Poillucci, G., Kus, S. & Cova, M. A. Diagnostic imaging costs before and after digital tomosynthesis implementation in patient management after detection of suspected thoracic lesions on chest radiography. *Insights Imaging* **5,** 147–155 (2014).

57. MacMahon, H., Austin, J. H., Gamsu, G., Herold, C. J., Jett, J. R., Naidich, D. P., Patz, E. F. & Swensen, S. J. Radiology guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology* **237,** 395–400 (2005).

58. Winer-Muram, H. T. The solitary pulmonary nodule. *Radiology* **239,** 34–49 (2006).

59. Erasmus, J. J., Connolly, J. E., McAdams, H. P. & Roggli, V. L. Solitary pulmonary nodules: Part I . Morphologic evaluation for differentiation of benign and malignant lesions. *Radiographics* **20,** 43–58 (2000).

60. Parkash, O. Lung cancer. A statistical study based on autopsy data from 1928 to 1972. *Respiration* **34,** 295–304 (1977).

61. Byers, T. E., Vena, J. E. & Rzepka, T. F. Predilection of lung cancer for the upper lobes: an epidemiologic inquiry. *J Natl Cancer Inst* **6,** 1271–1275 (1984).

62. Krupinski, E. A. The role of perception in imaging: past and future. *Semin Nucl Med* **41,** 392–400 (2011).

63. Beam, C. A., Conant, E. F. & Sickles, E. A. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. *J Natl Cancer Inst* **95,** 282–290 (2003).

64. Krupinski, E. A., Berbaum, K. S., Caldwell, R. T., Schartz, K. M. & Kim, J. Long radiology workdays reduce detection and accommodation accuracy. *J Am Coll Radiol* **7,** 698–704 (2010).

65. Krupinski, E.A., Samei, E. The handbook of medical image perception and techniques, Chapter 7; perceptual factors in reading medical images, Cambridge University press, New York (2010).

66. Kundel, H. L. & La Follette Jr, P. S. Visual search patterns and experience with radiological images. *Radiology* **103,** 523–528 (1972).

67. Berbaum, K. S., Franken Jr, E. A., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., Behlke, F. M., Sato, Y., Lu, C. H., El-Khoury, G. Y., Flickinger, F. W. & Montgomery, W. J. Satisfaction of search in diagnostic radiology. *Invest Radiol* **25,** 133–140 (1990).

68. International Commission on Radiation Units and Measurements (ICRU), Journal of the ICRU, Report 79, Vol 8, No 1, Oxford University Press (2008).

69. Månsson, L. G. Methods for the evaluation of image quality: a review. *Radiat Prot Dosimetry* **90,** 89–99 (2000).

70. Tingberg, A., Herrmann, C., Lanhede, B., Almén, A., Besjakov, J., Mattsson, S., Sund, P., Kheddache, S. & Månsson, L. G. Comparison of two methods for evaluation of the image quality of lumbar spine radiographs. *Radiat Prot Dosimetry* **90,** 165–168 (2000).

71. Svensson, E. Application of a rank-invariant method to evaluate reliability of ordered categorical assessment. *J Epidem Biostat* **3,** 403–409 (1998).

72. Båth, M. & Månsson, L. G. Visual grading characteristics (VGC) analysis: a non-parametric rank-invariant statistical method for image quality evaluation. *Br J Radiol* **80,** 169–176 (2007).

73. Smedby, Ö. & Fredrikson, M. Visual grading regression: analysing data from visual grading experiments with regression models. *Br J Radiol* **83,** 767–775 (2010).

74. Metz, C. E. Basic principles of ROC analysis. *Semin Nucl Med* **8,** 283–298 (1978).

75. Smelser, N. J. & Baltes, P. B. International Encyclopedia of Social & Behavioral Sciences, 1st Edition, p: 14078, Pergamon (2001).

76. Chakraborty, D. P. New developments in observer performance methodology in medical imaging. *Semin Nucl Med* **41,** 401–418 (2011).

77. Bunch, P. C., Hamilton, J. F., Sanderson, G. K. & Simmons, A. H. A free-response approach to the measurement and characterization of radiographic-observer performance. *J Appl Photogr Eng* **4,** 156–171 (1978).

78. Chakraborty, D. P. Validation and statistical power comparison of methods for analyzing free-response observer performance studies. *Acad Radiol* **15,** 1554–1566 (2008).

79. Chakraborty, D. P. Statistical power in observer-performance studies: comparison of the receiver operating characteristic and free-response methods in tasks involving localization. *Acad Radiol* **9,** 147–156 (2002).

80. Chakraborty, D. P. Maximum likelihood analysis of freeresponse receiver operating characteristic (FROC) data. *Med Phys* **16,** 561–568 (1989).

81. Chakraborty, D. P. A brief history of free-response receiver operating characteristic paradigm data analysis. *Acad Radiol* **20,** 915–919 (2013).

82. Chakraborty, D. P. & Yoon, H. JAFROC analysis revisited : figure-of-merit considerations for human observer studies. *Proc SPIE* **7263,** 72630T (2009).

83. Chakraborty, D. Choice of figure of merit in JAFROC. Private communication. November 12, 2013

84. Chakraborty, D. P. A status report on free-response analysis. *Radiat Prot Dosimetry* **139,** 20–25 (2010).

85. Kroft, L. J., Veldkamp, W. J., Mertens, B. J., van Delft, J. P. & Geleijns, J. Detection of simulated nodules on clinical radiographs: dose reduction at digital posteroanterior chest radiography. *Radiology* **241,** 392–398 (2006).

86. Saunders Jr, R. S. & Samei, E. A method for modifying the image quality parameters of digital radiographic images. *Med Phys* 3006 – 3017 (2003).

87. Li, C. M. & Dobbins III, J. T. Methodology for determining dose reduction for chest tomosynthesis. *Proc SPIE* **6510,** 65102D (2007).

88. Båth, M., Håkansson, M., Tingberg, A. & Månsson, L. G. Method of simulating dose reduction for digital radiographic systems. *Radiat Prot Dosimetry* **114,** 253–259 (2005).

89. Svalkvist, A. & Båth, M. Simulation of dose reduction in tomosynthesis. *Med Phys* **37,** 258–269 (2010).

90. Peuch, P. & Boussel, L. DicomWorks 1.3.5b. Available from: http://www.dicomworks.com/.Published February 16, 2003; Accessed June 15, 2007.

91. Börjesson, S., Håkansson, M., Båth, M., Kheddache, S., Svensson, S., Tingberg, A., Grahn, A., Ruschin, M., Hemdal, B., Mattsson, S. & Månsson, L. G. A software tool for increased efficiency in observer performance studies in radiology. *Radiat Prot Dosimetry* **114,** 45–52 (2005).

92. Håkansson, M., Svensson, S., Båth, M. & Månsson, G. L. ViewDEX – a Java-based software for presentation and evaluation of medical images in observer performance studies. *Proc SPIE* **6509,** 65091R (2007).

93. Håkansson, M., Svensson, S., Zachrisson, S., Svalkvist, A., Båth, M. & Månsson, L. G. ViewDEX 2 .0: a Java-based DICOM-compatible software for observer performance studies. *Proc SPIE* **7263,** 72631G (2009).

94. Håkansson, M., Svensson, S., Zachrisson, S., Svalkvist, A., Båth, M. & Månsson, L. G. ViewDEX: an efficient and easy-to-use software for observer performance studies. *Radiat Prot Dosimetry* **139,** 42–51 (2010).

95. National Electrical Manufacturers Association (NEMA), Digital imaging and communications in medicine (DICOM) part 14: grayscale standard display function, Virginia, USA, (2006).

96. Dobbins III, J. T., McAdams, H. P., Song, J.-W., Li, C. M., Godfrey, D. J., Delong, D. M., Paik, S.-H. & Martinez-Jimenez, S. Digital tomosynthesis of the chest for lung nodule detection: interim sensitivity results from an ongoing NIH-sponsored trial. *Med Phys* **35,** 2554–2557 (2008).

97. Commission of the European Communities. European guidelines on quality criteria for diagnostic radiographic images. Report EUR 16260 EN. Luxembourg: Office for Official Publications of the European Communities (1996).

98. Commission of the European Communities. European guidelines on quality criteria for computed tomography. Report EUR 16262 EN. Luxembourg: Office for Official Publications of the European Communities (1999).

99. Chakraborty, D. P. Dev Chakraborty's FROC web site. Available at: http://www.devchakraborty.com/. Updated June 10, 2013; Accessed February 4, 2014.

100. Kim, E. Y., Chung, M. J., Lee, H. Y., Koh, W.-J., Jung, H. N. & Lee, K. S. Pulmonary mycobacterial disease: diagnostic performance of low-dose digital tomosynthesis as compared with chest radiography. *Radiology* **257,** 269–277 (2010).

101. Quaia, E., Baratella, E., Cioffi, V., Bregant, P., Cernic, S., Cuttin, R. & Cova, M. A. The value of digital tomosynthesis in the diagnosis of suspected pulmonary lesions on chest radiography: analysis of diagnostic accuracy and confidence. *Acad Radiol* **17,** 1267–1274 (2010).

102. Lee, G., Jeong, Y. J., Kim, K. I., Song, J. W., Kang, D. M., Kim, Y. D. & Lee, J. W. Comparison of chest digital tomosynthesis and chest radiography for detection of asbestos-related pleuropulmonary disease. *Clin Radiol* **68,** 376–382 (2013).

103. Vult von Steyern, K., Björkman-Burtscher, I. M., Höglund, P., Bozovic, G., Wiklund, M. & Geijer, M. Description and validation of a scoring system for tomosynthesis in pulmonary cystic fibrosis. *Eur Radiol* **22,** 2718–2728 (2012).

104. Vult von Steyern, K., Björkman-Burtscher, I. & Geijer, M. Tomosynthesis in pulmonary cystic fibrosis with comparison to radiography and computed tomography: a pictorial review. *Insights Imaging* **3,** 81–89 (2012).

105. Rystedt, H., Ivarsson, J., Asplund, S., Johnsson, Å. A. & Båth, M. Rediscovering radiology: New technologies and remedial action at the worksite. *Soc Stud Sci* **41,** 867–891 (2011).

106. Ivarsson, J., Rystedt, H., Asplund, S., Johnsson, Å. A. & Båth, M. Facilitating professional modes of reasoning in radiology through Technology enhanced learning sessions. *Manuscript* (2014).

107. Quaia, E., Baratella, E., Cernic, S., Lorusso, A., Casagrande, F., Cioffi, V. & Cova, M. A. Analysis of the impact of digital tomosynthesis on the radiological investigation of patients with suspected pulmonary lesions on chest radiography. *Eur Radiol* **22,** 1912–1922 (2012).

108. Quaia, E., Baratella, E., Poillucci, G., Kus, S., Cioffi, V. & Cova, M. A. Digital tomosynthesis as a problem-solving imaging technique to confirm or exclude potential thoracic lesions based on chest X-ray radiography. *Acad Radiol* **20,** 546–553 (2013).

109. Lymer, G., Ivarsson, J., Rystedt, H., Johnsson, Å. A., Asplund, S. & Båth, M. Situated abstraction: from the particular to the general in second-order diagnostic work. *Discourse Stud* **16,** 185–215 (2014).

110. Chou, S. S., Kicska, G. A., Pipavath, S. N. & Reddy, G. P. Digital tomosynthesis of the chest: current and emerging applications. *Radiographics* **34,** 359–372 (2014).

111. Hwang, H. S., Chung, M. J. & Lee, K. S. Digital tomosynthesis of the chest: comparison of patient exposure dose and image quality between standard default setting and low dose setting. *Korean J Radiol* **14,** 525–531 (2013).

112. Gur, D. & Rockette, H. E. Performance assessments of diagnostic systems under the FROC paradigm: experimental, analytical, and results interpretation issues. *Acad Radiol* **15,** 1312–1315 (2008).

113. Chakraborty, D. P. Counterpoint to "Performance assessment of diagnostic systems under the FROC paradigm" by Gur and Rockette. *Acad Radiol* **16,** 507–510 (2009).

114. Gur, D. & Rockette, H. E. Performance assessments of diagnostic systems under the FROC paradigm: experimental, analytical, and results interpretation issues. *Academic radiology* **16,** 770–771 (2009).

115. Oken, M. M., Hocking, W. G., Kvale, P. A., Andriole, G. L., Buys, S. S., Church, T. R., Crawford, E. D., Fouad, M. N., Isaacs, C., Weissfeld, J. L., Brien, B. O., Ragard, L. R., Rathmell, J. M., Riley, T. L., Wright, P., Izmirlian, G., Pinsky, P. F., Prorok, P. C., Kramer, B. S., Miller, A. B., Gohagan, J. K. & Berg, C. D. Screening by chest radiograph and lung cancer mortality: the prostate, lung, colorectal, and ovarian (PLCO) randomized trial. *JAMA* **306,** 1865–1873 (2011).

116. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365,** 395–409 (2011).

117. Gomi, T., Nakajima, M., Fujiwara, H., Takeda, T., Saito, K., Umeda, T. & Sakaguchi, K. Comparison between chest digital tomosynthesis and CT as a screening method to detect artificial pulmonary nodules: a phantom study. *Br J Radiol* **85,** e622–e629 (2012).

118. Vult von Steyern, K. Tomosynthesis in pulmonary cystic fibrosis (doctoral dissertation). Lund: Media-Tryck, (2013) Available at: http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=4145732&fileOId=4145821.

119. Johnsson, Å. A., Svalkvist, A., Vikgren, J., Boijsen, M., Flinck, A., Kheddache, S. & Båth, M. A phantom study of nodule size evaluation with chest tomosynthesis and computed tomography. *Radiat Prot Dosimetry* **139,** 140–143 (2010).

120. Johnsson, Å. A., Fagman, E., Vikgren, J., Fisichella, V. A., Boijsen, M., Flinck, A., Kheddache, S., Svalkvist, A. & Båth, M. Pulmonary nodule size evaluation with chest tomosynthesis. *Radiology* **265,** 273–282 (2012).