BMJ Open

Digital breast tomosynthesis plus synthesised images versus standard full-field digital mammography in population-based screening (TOSYMA): protocol of a randomised controlled trial

Stefanie Weigel,[1] Joachim Gerss,[2] Hans-Werner Hense,[3] Miriam Krischke,[4] Alexander Sommer,[1] Jörg Czwoydzinski,[1] Horst Lenzen,[1] Laura Kerschke,[2] Karin Spieker,[4] Stefanie Dickmaenken,[4] Sonja Baier,[4] Marc Urban,[4] Gerold Hecht,[5] Oliver Heidinger,[6] Joachim Kieschke,[7] Walter Heindel[1]

To cite: Weigel S, Gerss J, Hense H-W, *et al*. Digital breast tomosynthesis plus synthesised images versus standard full-field digital mammography in population-based screening (TOSYMA): protocol of a randomised controlled trial. *BMJ Open* 2018;**8**:e020475. doi:10.1136/bmjopen-2017-020475

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-020475).

Received 6 November 2017
Revised 14 March 2018
Accepted 27 March 2018

For numbered affiliations see end of article.

Correspondence to
Prof. Walter Heindel;
heindel@uni-muenster.de

ABSTRACT

Introduction Development of digital breast tomosynthesis (DBT) provides a technology that generates three-dimensional data sets, thus reducing the pitfalls of overlapping breast tissue. Observational studies suggest that the combination of two-dimensional (2D) digital mammography and DBT increases diagnostic accuracy. However, because of duplicate exposure, this comes at the cost of an augmented radiation dose. This undesired adverse impact can be avoided by using synthesised 2D images reconstructed from the DBT data (s2D). We designed a diagnostic superiority trial on a high level of evidence with the aim of providing a comparison of screening efficacy parameters resulting from DBT+s2D versus the current screening standard 2D full-field digital mammography (FFDM) in a multicentre and multivendor setting on the basis of the quality-controlled, population-based, biennial mammography screening programme in Germany.

Methods and analysis 80 000 women in the eligible age 50–69 years attending the routine mammography screening programme and willing to participate in the TOSYMA trial will be assigned by 1:1 randomisation to either the intervention arm (DBT+s2D) or the control arm (FFDM) during a 12-month recruitment period in screening units of North Rhine-Westphalia and Lower Saxony. State cancer registries will provide the follow-up of interval cancers. Primary endpoints are the detection rate of invasive breast cancers at screening examination and the cumulative incidence of interval cancers in the 2 years after a negative examination. Secondary endpoints are the detection rate of ductal carcinoma in situ and of tumour size T1, the recall rate for assessment, the positive predictive value of recall and the cumulative 12-month incidence of interval cancers. An adaptive statistical design with one interim analysis provides the option to modify the design.

Ethics and dissemination This protocol has been approved by the local medical ethical committee (2016-132-f-S). Results will be submitted to international peer-reviewed journals.

Trial registration NCT03377036; Pre-results.

Strengths and limitations of this trial

- This trial is conducted as a large multicentric randomised controlled pragmatic study in the setting of the German routine mammography screening programme.
- The German mammography programme has a high level of quality due to central standardised quality assurance procedures regarding imaging, technology and diagnostic work-up.
- Assessment of the performance of tomosynthesis as a screening tool as compared with standard two-dimensional digital mammography screening is accomplished without duplicate radiation exposure.
- As the trial phase comprises the first-time use of digital breast tomosynthesis (DBT) as a screening tool, results for subsequent rounds of DBT screening may be different.

INTRODUCTION

Screening for breast cancer

With more than 70 000 new diagnoses per year in Germany, breast cancer accounts for one in three new female cancers, making it by far the most common form of cancer among women. Moreover, it is the leading fatal cancer in women, and about one in six annual cancer fatalities in women is attributable to breast cancer.[1]

There is consistent evidence from randomised studies that organised population-based mammography screening programmes are able to reduce breast cancer mortality by around 20% in women invited to screening as compared with an uninvited

group.[2] Similarly, a European network of population-based routine mammography screening programmes found evidence from various observational studies for effective breast cancer screening.[3] However, the expected benefit in terms of mortality reduction and improved quality of life due to early detection and less aggressive therapies needs to be balanced against the potential harms, in particular, overdiagnosis.[2–5]

In Germany, a systematic, quality-controlled, population-based mammography screening programme was started in 2005 and fully implemented in 2010. It offers biennial mammography screening examinations—in accordance with the European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis[6]—for all women in the age range 50 to 69 years. The programme includes novel digital imaging techniques. Of note, the population-based mammography screening programme needs to be clearly distinguished from multimodal prevention programmes, which are specifically dedicated to the surveillance of women at high genetic risk for breast cancer.[7 8]

In standard two-dimensional digital mammography (2D-DM), overlapping breast tissues may result in concealment of features of malignancy. This may cause false-negative findings and a delayed diagnosis of breast cancer in the subsequent screening interval. In addition, in 2D imaging, superposition of tissue structures may also lead to false-positive findings.

Digital breast tomosynthesis (DBT) is a novel imaging technology generating three-dimensional data sets of the breast, thus potentially reducing the pitfalls of overlapping tissue. To date, only few studies have evaluated the use of DBT in the setting of routine service screening.

The Italian group of Ciatto *et al*[9] found that DBT, added sequentially to standard 2D-DM, significantly increased the invasive breast cancer detection rate by 48% (4.8 per 1000 examinations for 2D-DM alone and 7.1 per 1000 examinations for 2D-DM plus DBT); furthermore, recalls were reduced by 17% without missing any case of breast cancer.

In Oslo, Skaane *et al*[10] performed a prospective screening study in which women had both standard 2D-DM and DBT; combined imaging significantly increased detection of breast cancer and reduced the false-positive rate before arbitration compared with 2D-DM alone. Invasive cancer detection rates were 4.4 per 1000 examinations for 2D-DM alone and 6.4 per 1000 examinations for 2D-DM plus DBT (P<0.001), confirming the results of the Italian study.[9] Of note, DBT plus 2D-DM did not lead to a significant increase of the detection of ductal carcinoma in situ (DCIS).[10]

In support of these results, a multicentre screening study from the USA by Friedewald *et al* reported that 2D-DM plus DBT was associated with an increase in breast cancer detection and a decrease in recall rate.[11]

The major advantage of combining 2D-DM with DBT is a substantial improvement in the diagnostic accuracy. However, because of the duplicate exposures, this comes at the cost of a radiation dose that is roughly doubled compared with the standard examination.

This undesired adverse effect may be avoided by using the innovative technique of synthesised 2D (s2D) images reconstructed without any additional radiation exposures, directly from the DBT data sets (DBT+s2D). Data regarding the diagnostic performance of single use of DBT+s2D in the screening setting are scarce.

Since start of its use, s2D image reconstruction has improved with new technical developments. For example, Skaane *et al* reported for DBT+s2D an overall performance level (cancer detection rates, false-positive scores) comparable with DBT plus standard 2D-DM.[12] However, this observational study was restricted to only a single institution and a single device vendor.[10 12]

In the TOMMY trial, the diagnostic performance of DBT as assessment tool was compared in a recent retrospective multicentre reading study; 2D-DM alone was compared with DBT combined with 2D-DM or synthetic 2D mammogram in women recalled for further assessment after routine 2D breast screening.[13 14] The addition of DBT was associated with a 34% increase in the odds of depicting cancer; a significant increase in sensitivity was based on the radiological feature of masses. Synthetic 2D mammography appeared to have a diagnostic accuracy similar to that of 2D mammography when used in conjunction with DBT. In detailed analyses, synthetic 2D mammography was inferior to 2D mammography alone or combined with DBT for detecting small sizes of DCIS (11–20 mm).

American researchers evaluated the early implementation of synthesised 2D mammography in a population screened entirely with DBT and s2D and compared the recall rates and cancer detection rates with historic results of DM combined with DBT screening. They found screening with DBT+s2D in a large urban practice resulted in similar outcomes compared with 2D-DM plus DBT imaging.[15]

Other American researchers found that screening with DBT+s2D mammography in a large community-based practice improved recall rate and positive predictive values without loss of cancer detection rate when compared with DBT plus FFDM and FFDM alone.[16]

Likewise, the Italian group found that both 2D-DM plus tomosynthesis (cancer detection rate 8.5/1000 women screened) and synthetic mammography plus tomosynthesis (8.8/1000) had significantly higher rates of breast cancer detection than 2D mammography alone (6.3/1000) while the cancer detection rate between the two DBT groups did not differ significantly.[17]

According to a recent Spanish observational, retrospective, single-centre, multireader blinded study, the performance of stand-alone synthetic image mammography was not inferior for lesion visibility and for lesion BIRADS categorisation as compared with full-field DM.[18]

In agreement, a prospective study in the Verona screening programme found for DBT plus synthetic 2D compared with a cohort of women screened with FFDM in

the previous year an increased cancer detection rate with recall rates comparable with those of FFDM. Increased cancer detection rates were present among women classified as having low breast density or high breast density.[19]

Consequently, a recent review concludes that new studies of DBT should preferentially use DBT with s2D instead of DBT plus 2D-DM.[20]

To assess whether the findings of increased cancer detection with DBT can in fact be translated into improved screening efficacy, the incidence of interval cancers needs to be investigated. Interval cancers are DCIS and invasive breast cancers that occur before the next scheduled screening after a negative screening examination. The monitoring of interval cancers is an important part of the evaluation of a population-based screening programme since their reduction provides evidence that additional cancer detection does not primarily represent overdiagnosis. In Germany, the requirements for a comprehensive evaluation of interval cancers have not yet been fully implemented. Recently published articles reported interval cancer rates for a 2-year screening interval of 23.2 and 25.0 per 10 000 negative screening examinations in North Rhine-Westphalia and Lower Saxony,[21 22] respectively, including categorisation of interval cancers.[22] Furthermore, there is only little evidence on the impact of DBT on interval cancers. For the participants of the Italian study on DBT, added sequentially to standard 2D-DM,[9] an interval cancer rate of 12.4 per 10 000 (9/7235) negative screens has been reported, whereas in a concurrent group of women receiving 2D-DM, a rate of 16.1 per 10 000 (40/24 922) negative screens has been observed.[23]

Ultimately missing is a diagnostic superiority trial on a high level of evidence that provides a statistically sound comparison of DBT+s2D versus the standard 2D full-field DM in routine screening. The clinical study protocol describes the planned concept of a randomised, controlled, multicentre and multivendor pragmatic clinical trial that is embedded in the routine population-based German mammography screening programme.

METHODS AND ANALYSIS

Study setting and design

Screening-eligible women in Germany, aged 50 to 69 years, receive with their routine screening invitation a brochure about benefits, risks and limitations of mammography screening according to the established process of informed decision.[24] Women living in the catchment area of screening units that participate in the TOSYMA trial will receive additional information material about TOSYMA together with their regular screening invitations. The mailings are sent out by two central offices for the state North Rhine-Westphalia and by a third for the state Lower Saxony. Screening units located in the two federal states have confirmed their intention of participating in the trial as study centres.

A study hotline for personal information via telephone will be set up. In addition to the present right of invited women to be personally informed about the advantages and disadvantages of the German mammography screening programme, women can make an individual appointment with a screening physician to clarify specific questions relating to TOSYMA. Women willing to participate in the trial need to bring the informed consent documents with them to the screening examination at the local mammography screening unit, which acts as a study centre. Study centres will register women who consent to participation in a trial-specific software component that has been embedded into the routine and certified screening documentation software MaSc. All non-consenters will receive standard screening 2D-FFDM.

The study will be conducted as a multicentric two-arm parallel, randomised, controlled diagnostic superiority trial. Women are assigned by a study-specific randomisation tool of the software to the test arm, digital breast tomosynthesis plus synthetic 2D mammography (DBT+s2D), or to the control arm, full-field digital mammography (2D-FFDM). The selection of DBT devices used in this multivendor trial is entirely and only in the responsibility of each participating study centre as long as these comply with the specific requirements outlined below under intervention protocol.

According to routine screening standards, masked double readings of screening images will be performed in the test and control arm as laid down in the European Guidelines. If at least one of both readers marks any suspicious abnormality, the case will be discussed together with a third reader, the arbitrator, to decide if further diagnostic work-up of the lesion will be necessary. If required, women will be called back for an assessment to verify or rule out the presence of breast cancer. The established additional routine diagnostic procedures may include clinical examination, additional 2D mammographic views, tomosynthesis, ultrasound examinations, MRI or needle biopsies, dependent on the specific suspicious breast abnormality. In rare cases (<1% of all screened women), breast abnormalities may need another diagnostic assessment, typically after 6 months. Procedures are similar to the primary assessment, including imaging and if necessary image-guided biopsies. Data will be entered on site by the study centres using MaSc.

Participation eligibility

Inclusion criteria are identical with that of the target population of the national mammography screening programme, that is, women aged 50–69 years.[6 24 25] Women attending the routine screening mammography at a participating screening unit who have given written informed consent for participation in the trial will be included. Exclusion criteria of the trial will be, as generally applied, a breast cancer diagnosis up to 5 years prior to screening invitation and mammography within the preceding 12 months. Additionally, breast implants are a trial-specific exclusion criterion.

First-round participation is not defined as an exclusion criterion since the number of first-round participants of the German Mammography Screening Programme is expected to be stable and the estimated effects thus reflect the benefits that can be expected in the target population for screening.

Intervention protocol and controls

Participants allocated to the test intervention will be screened by two-view DBT (cranial–caudal and medio-latero-oblique). Synthesised 2D mammograms will be reconstructed of each view. The control intervention will be the standard two-view full-field DM (cranial–caudal and medio-latero-oblique) of each breast. The procedures will be similar apart from a few seconds longer breast compression time for DBT compared with 2D-DM. Regular quality assurance standards of the national mammography screening programme and additional study-specific quality assurance measurements will be implemented for both study arms. That is, all participating mammography systems have to fulfil the following requirements:

- CE label (no prototypes are allowed).
- Approval according to the German X-ray regulation.
- Acceptance test.
- Radiation Protection Expert test (Sachverständigenprüfung).
- Requirements of the German Quality Assurance Guideline (QS-RL).
- Availability of a 2D mode and 3D mode on the same imaging system.
- All systems have to fulfil the limiting value for the average glandular dose of 2D mammography in the 3D mode according to the German standard DIN 6868-162.
- Compatibility to the Digital Imaging and Communication in Medicine (DICOM) standard, Service-Object-Pair (SOP) class, Breast Tomosynthesis Object (BTO).
- Access to the raw projection images (only for physico-technical quality assurance).

All requirements will be checked by a medical physicist from the Reference Centre Muenster in an initial test before the start of the study. Further additional routine constancy measurements for breast tomosynthesis systems will be implemented in the daily and monthly quality control of all participating mammography systems. These tests will be performed by the radiographers of the screening units and the results will be transferred online to the Reference Centre Münster for permanent quality control monitoring.

Apart from technical requirements, medical doctors and technicians will be trained prior to the study start. Training will include acquirement of two-view tomosynthesis data sets, implementation of a trial specific hanging protocol, reading of study examinations, synchronisation of assessment procedures as well as documentation of study-relevant data. Furthermore, every physician participating as a reader in the trial has to assure his expertise as follows:

- According to national screening standards, having participated in all regular teaching courses for the screening programme and having passed the yearly test of 50 screening case studies.
- According to national screening standards, a volume of at least 5000 screening mammograms the year before participating in the study.

During the clinical trial, all readers are regularly assessed with an emphasis on a comparable amount of sets for DBT+s2D images and 2D-FFDM.

Randomisation and blinding

Eligible patients will be randomised in a 1:1 ratio to one of the two study arms. For the preservation of allocation concealment, randomisation lists will be generated and integrated into the MaSc software so that the allocation of future patients is inaccessible to the user. The randomisation lists will be generated and kept by an independent statistician using the random number generator of the validated software SAS (SAS Institute, Cary, North Carolina, USA). Randomisation will be balanced by blocks and stratified by site. The block length will be fixed and kept confidential by the statistician performing the randomisation.

Given the nature of interventions and images produced, study site personnel conducting the screening and readers cannot be blinded for the randomised intervention. Participants will systematically not be informed by the screening personnel about their randomisation prior to the completion of the screening mammography examination.

Objectives

Primary objectives

The first primary objective of the study is to evaluate whether DBT+s2D leads to a clinically relevant increase in the detection rate of invasive breast cancers at the screening examination (predefined at ≥33%) as compared with standard 2D-FFDM. In accordance with the protocol for evaluation of the German mammography screening programme, a screen-detected breast cancer is classified as invasive carcinoma if the T category (tumour size) of the pTNM classification falls into one of the following categories: 1mic, 1a, 1b, 1c, 1, 2, 3, 4a, 4b, 4c, 4d, 4, X (invasive breast cancer, tumour diameter is missing) or the final categorisation is based on a neoadjuvant therapy (ypTNM), implying an invasive cancer prior to therapy.

The second primary objective is to compare the cumulative incidence of interval cancers between the study arms in order to assess the prognostic importance of the additional cancers that will, predictably, be diagnosed by DBT+s2D and insofar to investigate the potential for overdiagnosis. Interval cancers are DCIS and invasive breast cancers that occur in the 24-month interval after a negative screening examination. The present programme sensitivity in the target study population is in the order of 75%.[22] If the postulated increase in cancer detection with DBT+s2D translates into improved screening efficacy, the

resulting reduction of interval cancers is expected to be up to 30%.

The cumulative incidences of interval cancers will be determined by linking MaSc records with the two state cancer registries. Their complete acquisition requires an expanded study follow-up of further 24 months to accommodate the reporting lags experienced in cancer registries.

Secondary objectives

The secondary objectives of the study are to compare the two screening modalities regarding

- detection rates of DCIS (pTis) at the screening examination,
- detection rates of tumour size pT1 (tumor ≤20 mm in greatest dimension) at the screening examination,
- recall rates for assessment after the screening examination,
- positive predictive values of recall for assessment (PPV1) and
- cumulative incidences of interval cancers in 12 months after the screening examination.

Recruitment and timeline

Assuming a study participation of 75% or more, and a target of recruiting 80 000 study participants, about 107 000 women attending mammography screening examinations will have to be offered trial participation. Based on a state-wide screening response of approximately 50%,[25] 214 000 women will receive routine screening invitations together with the study-specific additional information. The recruitment phase is expected to expand over a 12-month period (figure 1).

Assessment of efficacy

The first primary efficacy endpoint is the detection rate of invasive breast cancers, which will be compared between both study arms. An interim analysis is planned to give an option to stop the trial for futility, if the data indicate that no significant result can be expected with a reasonable sample size in the final statistical analysis or to change the design of the study in the case of important new discoveries. The interim analysis will be conducted when first primary endpoint data (detection rate of invasive breast cancers) of 40 000 patients are available; this is expected 9 months after the study start and will cover data for the first six study months. Efficacy will finally be assessed at the end of the study when primary efficacy endpoint data for the detection rate of invasive breast cancers are available for all study participants.

Analyses of the first primary and secondary endpoints are planned in an interim database that will be completed after

- the target numbers of participants have been recruited and
- data for all participants (except for interval cancers from the cancer registries) have been retrieved.

The second primary endpoint, cumulative incidences of interval cancers at 24 months, will be determined in collaboration with the two state cancer registries. Due to the reporting lag experienced in cancer registries (time span from clinical diagnosis to notification in the cancer registry system), the evaluation of interval cancers requires an expanded study follow-up of further 24 months. The analysis of this second primary endpoint will therefore be performed approximately 60 months after the start of the study.

The end of the entire study is reached when

- the target numbers of participants to be recruited have been reached,
- the data of all participants, including information about interval cancers, have been retrieved and
- the database is locked.

Anticipated study start date is June 2018, and estimated study completion date is June 2023.

Data collection and management

The study specific screening software MaSc will be used for data collection. Study-specific features, for example, documentation form for adverse events (AEs), adverse device effects and serious adverse events (SAEs), will be embedded in the standard software and the study site personnel will be trained on the new features of the system. Only persons authorised to enter data will have access to this system. To prevent the identification of a study participant, study data will be pseudonymised by means of a patient identification number. If patient data are transferred to an institution outside the study site, the patient identification number will be used as the only patient identifier.

All procedures and results from the screening and assessment visits will be entered in MaSc by study site personnel. This set of data will be considered as the source data. The study-specific documentation of screening procedures includes in particular data about the machines used, the radiation doses applied, the duration of reading, mammographic morphology of tomosynthesis-specific cancer detection and, in the case of cancer detection, the diagnostic ascertainment method (eg, kind of biopsy, etc). The screening data documentation comprises demographic data, informed consent form, randomisation allocation, screening results (ie, malignancy yes/no/ open), type of malignancy (invasive breast cancer, DCIS), tumour node metastasis (TNM) staging and recall for assessment (yes/no).

A predefined subset of data will be extracted from MaSc for all study participants and will be transferred quarterly to the Centre for Clinical Trials Münster (ZKS Münster). The transferred data will be stored on servers that are located in a secure data centre and behind a firewall in the network of the University Hospital Münster. A backup of the data will be saved on a daily basis.

For data management, the statistical software package SAS (SAS Institute) will be used. Data checks concerning the quality and completeness will be performed by the

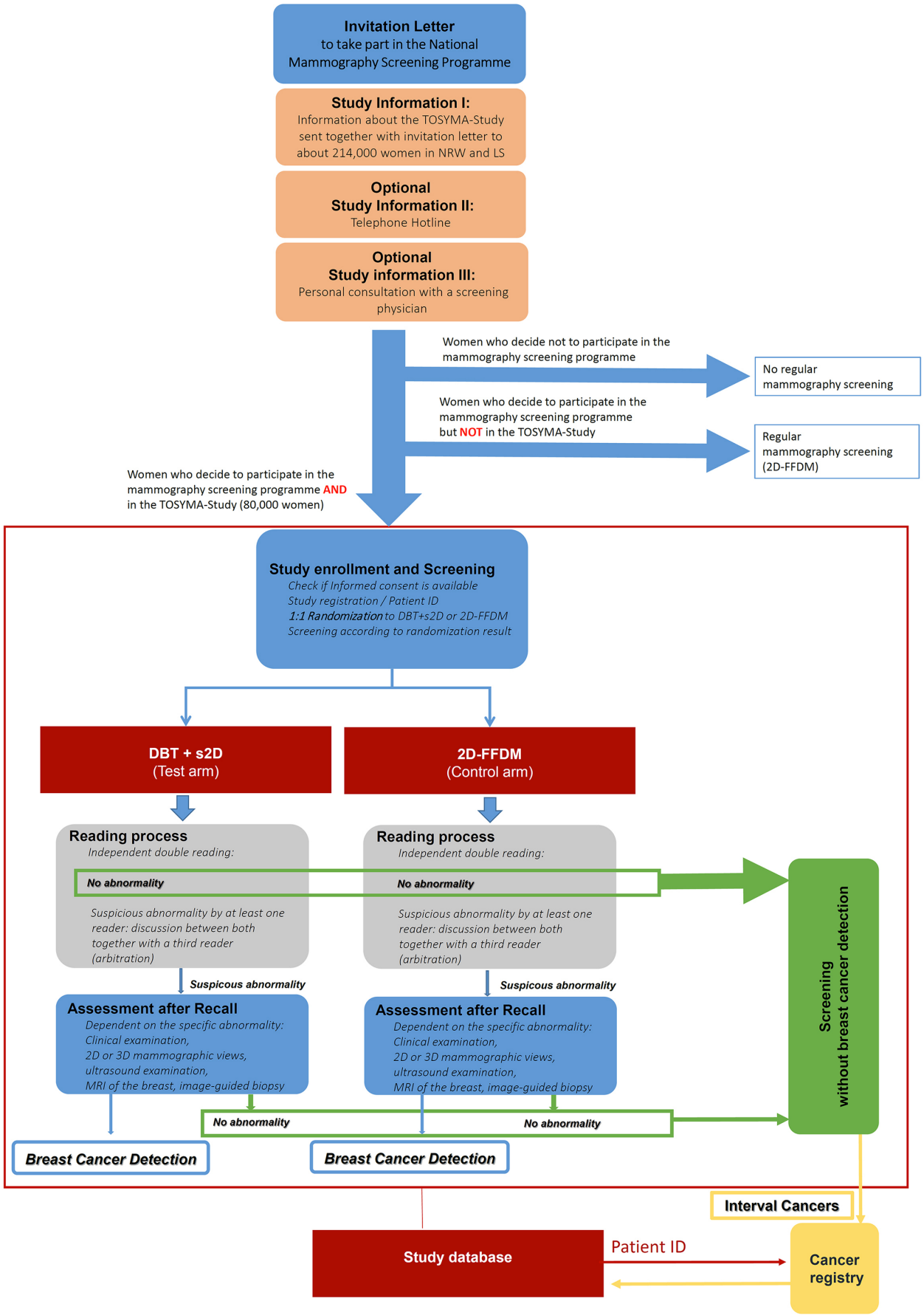

Figure 1 Flow chart of the screening study. DBT, digital breast tomosynthesis; LS, Lower Saxony; NRW, North Rhine-Westphalia; s2D, synthetic two-dimensional mammogram; 2D-FFDM, two-dimensional full-field digital mammography.

ZKS Münster according to the data validation plan. Reports about data quality and completeness will be generated and passed on to the medical project management for review. In case of predefined non-plausible or missing data, queries will be sent to the study centres. The queries must be resolved by authorised members of the study centres in a timely manner.

After completion of data entry and data processing, the database will be locked and the data will be transferred for statistical analysis.

Data on the ascertainment of interval cancers will be provided by the respective cancer registries. A list of pseudonymised identity data of study participants with negative screening results will be sent to the cancer registries after last patient - last visit (LPLV). The cancer registries will be able to link the pseudonyms of each trial participant to the registry database and check if an interval cancer occurred in the 12-month (secondary endpoint) and 24-month (second primary endpoint) period after the negative study screening. A list (.csv file format) with the following variables will be provided by the cancer registries:

- Interval cancer yes or no.
- In case of interval cancer yes: date of diagnosis, type (invasive/in situ), histology, location, TNM, grading.

The data on interval carcinoma will be reconciled with the data already in the study database and constitute the final data set for analysis.

Safety

The safety of the study participants and the safety of the devices used will be assessed by the study site personnel. Safety definitions used in this trial are based on ISO 14155:2011(E) and the MEDDEV Guideline 2.7/3 revision 3, corresponding to relevant definitions in regulatory documents in Germany.

AEs, including SAEs and incidents, and all device deficiencies occurring on the day of mammography screening from the time of registration until the study participant leaves the study site, that is, end of the primary screening process, will be documented as soon as possible in the MaSc database. AEs occurring during later steps, for example, during assessments are outside the scope of this study. Documentation of AEs includes an assessment with regard to seriousness and relatedness to an investigational medical device and, if so, to a device deficiency. A serious AE caused by a device deficiency would be an incident. Documentation of device deficiencies includes an assessment with regard to the potential for resulting in a SAE, which would also be an incident. All AEs and device deficiencies documented in the MaSc database will be re-assessed by the Reference Centre for Mammography in Münster in regular intervals. Listings of AEs and device deficiencies will be presented to the Data Monitoring Committee (DMC). In accordance with the Declaration of Helsinki, SAEs will be submitted to the Ethics Committee in a yearly report by the coordinating investigator.

Statistical analysis

Statistical analyses will be performed using the validated software SAS (SAS Institute). For each group, summary statistics of demographic and other baseline characteristics, including the number of observations, mean, SD, median, minimum and maximum for continuous variables and frequencies and percentages for categorical variables, will be provided.

The specific aim of the study is to show a clinically relevant difference (predefined at ≥33%) between the detection rates of invasive breast cancers diagnosed by DBT+s2D and 2D-FFDM. A second aim is to compare the cumulative incidences of interval cancers between both study arms in order to investigate the potential for overdiagnosis. For this purpose, the null hypotheses of no difference in the detection rates of invasive breast cancers (H_0^{DR}) and the cumulative incidences of interval cancers at 24 months after participation (H_0^{ICR}) will be assessed at an overall two-sided significance level of α=0.05. The analyses will be performed using a Cochran-Mantel-Haenszel test with stratification by site for each hypothesis. To preserve the family-wise error rate, a hierarchical test strategy will be applied in which the hypotheses are tested in a fixed sequence at local level α=0.05 until the first non-rejection occurs. According to clinical importance, H_0^{DR} will be evaluated in the first place of this sequence, whereas H_0^{ICR} will be evaluated in the second place.

An adaptive interim analysis will be performed for each hypothesis based on the conditional error function approach. To avoid directional conflicts caused by two-sided P values, H_0^{DR} and H_0^{ICR} will be tested using two one-sided tests at a level of 0.025. In this study, the following conditional error function (CEF) with weights $w_1 = \sqrt{0.4}$, $w_2 = \sqrt{0.6}$ will be applied to each hypothesis:

$$A(p_1) = \begin{cases} 0.5 \cdot \left[1 - \Phi\left(\frac{1.282 - \sqrt{0.4}\,\Phi^{-1}(1-2p_1)}{\sqrt{0.6}}\right)\right], & 0 \le p_1 \le 0.5 \\ 0, & p_1 > 0.5 \end{cases}$$

In this notation, p_1 denotes one of the two one-sided P values (both corresponding to either H_0^{DR} or H_0^{ICR}) of the first stage and Φ^{-1} denotes the inverse of the standard normal cumulative distribution function. Due to the required time for data completion and cleaning, the results of the interim analysis will not be obtained until the time when the majority of the planned total number of patients has already been recruited. Therefore, no early stopping for benefit is implemented and the full significance level is spent at the second stage. Moreover, the upper limit of the domain and codomain of the CEF is set to 0.5 in order to exclude the possibility of a significant result caused by contradictory one-sided first-stage and second-stage P values (ie, one P value indicating superiority and the other indicating inferiority of DBT+s2D vs 2D-FFDM).

The interim analysis is intended to be performed when first primary endpoint data (detection rate of invasive breast cancer) of approximately half of the planned total number of patients are available. This is expected

to be attained 9 months after the start of recruitment, covering the first 6 months of the recruitment period. Due to a reporting lag in cancer registries, the interim and final analyses of the second hypothesis (H_0^{ICR}) have to be performed simultaneously at the time when primary endpoint data of the 24-month incidence of interval cancers are available.

In rare cases, first primary endpoint data might not yet be available, when the interim analysis of this endpoint is performed, due to delayed documentation of diagnostic assessment results. If a reasonable amount of missing first primary outcome data will be found at the interim analysis, missing values will be imputed, using appropriate imputation techniques. In the final analysis, the result of the interim analysis will be recalculated based on the true primary outcome data of stage 1 patients observed until the final analysis takes place. In the final primary efficacy analyses as well as all further statistical analyses, no imputation of missing values will be performed. An additional sensitivity analysis of the primary endpoints will be conducted in which missing (binary) outcome data will be handled according to Imai.[26]

The primary statistical analysis will be performed on all randomised subjects according to the intention-to-treat (ITT) principle and will provide confirmatory statistical evidence.

All secondary analyses will be carried out on the ITT population and will be interpreted as exploratory. Prespecified secondary endpoints will be compared between both study arms using two-sided Cochran-Mantel-Haenszel tests with stratification by site. Additional multivariable statistical analyses will be performed in order to identify factors that impact the diagnostic performance of both screening methods. Further exploratory analyses will be laid down in the Statistical Analysis Plan before performing the analyses.

Primary and secondary analyses will be repeated on the per-protocol population as exploratory analyses.

Sample size calculation

Since the majority of study participants are expected to be subsequent round participants and breast cancer detection rates are in general lower for these patients, sample size calculation will be based on results from the subsequent rounds of the German mammography screening programme as a conservative approach. The average detection rate of invasive breast cancers in the follow-up rounds of the mammography screening programme in the federal state of North Rhine-Westphalia 2009–2010 was 4.4 per 1000 screened women. Based on evidence from the literature evaluating DBT in addition to 2D-DM, the expected gain in the first primary efficacy endpoint (detection rate of invasive breast cancers) is postulated as ≥33%. Hence, in each site, the expected first primary endpoint rates in the 2D-FFDM and the DBT+s2D arm amount to 4.4 and 5.852 per 1000 women screened, respectively. Based on the specified adaptive design, a total number of 80 000 patients needs to be recruited to provide a power of at least 80% to reject the first primary hypothesis, applying an overall (two-sided) significance level of α=0.05. The interim analysis will be performed at the time when first primary endpoint data of 40 000 patients are available in total across both study arms and offers the possibility of design modifications, in particular sample size adaptations. Power calculations were performed using the ADDPLAN software V.6.0.9 and R (R Core Team) V.3.2.4.

Data monitoring committee

The DMC of the trial will consist of three independent members including a statistician, a radiologist and a radio-epidemiologist who are not involved in the trial. The DMC will meet twice, once after 4 months of recruitment and once at the time of the interim analysis. If necessary, an additional meeting will be scheduled. The DMC will be responsible for supervision of safety data including radiation exposure, data quality and trial conduct on the basis of adverse events listings, central monitoring reports prepared by the ZKS Münster and other information. They will also be involved in the planned interim analysis.

Monitoring and auditing

To ensure a high degree of safety and data quality, the study sites will be monitored regularly. Initiation visits will be conducted as soon as all required approvals, contracts, documentation and procedural information are in place at the study site. During the initiation visit, the monitor will familiarise the site staff with the protocol, the MaSc database, the study documents and the requirements of Good Clinical Practice. Follow-up monitoring visits onsite will take place in appropriate intervals, depending on actual recruitment status.

An audit may be initiated at any study site during the study and after completion. All study-related documentation must be made available to the designated auditor(s).

Ethics, amendments and dissemination

The study will be conducted in accordance with the Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects (current version, October 2013, Fortaleza) and Good Clinical Practice according to DIN EN ISO 14155 and will be consistent with applicable regulatory requirements and laws.

The study protocol has been reviewed and approved by the Medical Ethical Committees (2016-132f-S). In case of protocol amendments, a new application will be submitted to the ethical committees.

Approval of the Federal Office for Radiation Protection (BfS), estimating benefit and radiation-associated harms, has been obtained.[27]

The result of the interim analysis is kept confidential, that is, the coordinating investigator is only informed about (dis)continuation of recruitment and the sample size that is determined for the second stage of the study. Adaptations will be introduced by a protocol

amendment that must be approved by the competent ethical committees.

After completion of the biostatistical evaluation, a final study report will be prepared, including all results, irrespective of whether favourable or not. A summary of the final report will be provided to the ethical committees within 12 months after the end of study. Results will also be submitted to international peer-reviewed journals and presented at scientific conferences.

Patient and public involvement

The study protocol was developed by the Institute of Clinical Radiology at the University of Münster in close collaboration with the installed organisations of the German Mammography Screening Programme (eg, Reference Centres for Mammography Screening, German Mammography Screening Office, National Association of Statutory Health Insurance Physicians, National Association of Statutory Health Insurance Funds, and State Cancer Registries).

Primary and secondary endpoints were chosen according to the current guidelines of the national Mammography Screening Programme. These endpoints are established parameters of screening efficiency. Our approach is driven by scientific experts. Potential screening participants/patients were not involved in design and planning of this study protocol.

The study was registered on the public registry ClinicalTrials.gov. Results of the study will be disseminated to the public and study participants via ClinicalTrials.gov., peer-reviewed publications of study results and press releases.

DISCUSSION

No pragmatic randomised trial has been carried out to date investigating whether DBT plus synthetic image mammography is superior to FFDM for early breast cancer detection in a systematic population-based screening programme conducted in multiple institutions and with devices from multiple vendors.

The present evidence from non-randomised, observational studies indicates that DBT in combination with s2D image reconstruction may increase the rates of screen-detected invasive breast cancers by more than 33% compared with the conventional 2D-FFDM screening (figure 2). At the same time, DCIS rates were only moderately raised with DBT+s2D as compared with standard 2D-FFDM.[9 12 14 15] Furthermore, the recall rates were significantly reduced.[9 12 14 16] To corroborate these results from observational studies, a randomised, controlled, multicentre, multivendor trial is required to confirm on a high level of evidence that screening with DBT+s2D is indeed superior to using the conventional 2D-FFDM technique. TOSYMA sets out to fill this obvious research and evidence gap. The main objective of the TOSYMA trial is to investigate the hypothesis that DBT+s2D leads to a clinically relevant increase in the detection rate of screen-detected invasive cancers compared with standard 2D-FFDM. Moreover, the potential for overdiagnosis by the novel technique will be investigated by means of the 24-month cumulative incidence of interval cancers of screen-negative women. Higher detection rates in screening are expected to result in lower cumulative incidences of interval cancers; otherwise, raised screen detection rates would rather indicate overdiagnosis.

Systematic screening of the female population based on digital mammography techniques offers the perspective of saving lives from cancer death and of reducing the adverse side effects of surgical and systematic treatment by detecting cancer at earlier stages, when it is more responsive to less aggressive treatment.[6] Improved diagnostic technology in population-based screening leading to an impact on screening effectiveness implies a benefit for the participating women and, depending on participation rates, for the screening target population in general. If the trial results confirm the view that the new screening modality DBT+s2D leads to an increased detection of invasive cancers, no rise of false-positive recalls and a concomitant decrease of interval cancers, it carries the potential of challenging the current standard screening modality.

There are ongoing trials by other investigators that aim to evaluate the value of tomosynthesis in population-based mammography screening. In Italy, a randomised trial

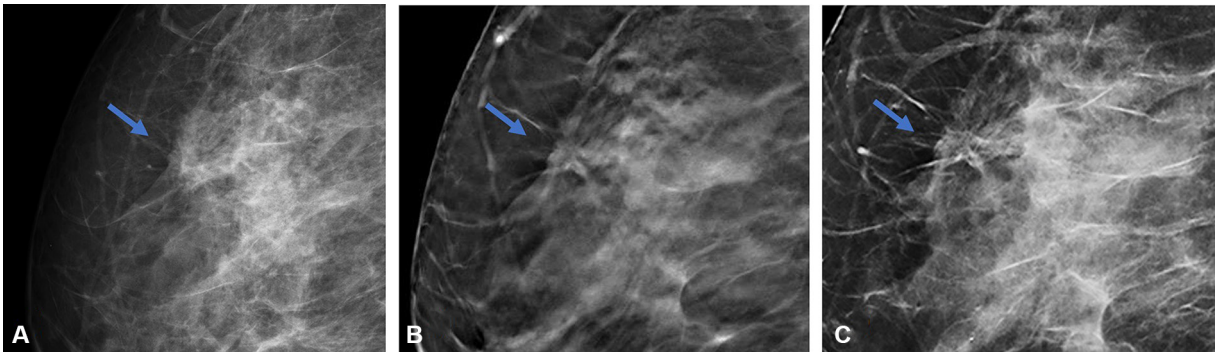

Figure 2 Invasive lobular carcinoma of the right breast depicted by an architectural distortion (A) subtle finding on the two-dimensional full-field digital mammogram (B) pronounced visible on the slice of the digital tomosynthesis as well as on the (C) reconstructed synthetic mammogram.

(about 20 000 tests in the intervention arm) compares the performance of tomosynthesis *plus* 2D-DM versus usual care (2D-DM only) with respect to the incidence of advanced stage breast cancers (interval and following screening examination) and interval cancers. Estimated completion date is December 2018 (ClinicalTrials.gov Identifier: NCT02698202).

In Norway, a screening trial started recruitment in January 2016, planned until 2018, comparing 2D synthetic mammography *plus* DBT with 2D-DM with equipment from one vendor in Bergen. Estimated enrolment comprises 37 000 women who are invited for screening. Estimated primary completion date is January 2020 for the comparison of rates of screening detected breast cancer in tomosynthesis versus DM, and evaluations on interval cancers will be included additionally (ClinicalTrials.gov Identifier: NCT02835625).

In the USA, under responsibility of the Eastern Cooperative Oncology Group (ECOG-ACRIN Cancer Research Group), a randomised study, comparing 2D-DM and digital tomosynthesis mammography, started recruitment in 2017 for women aged 45 to 74 years. The primary outcome is the proportion of women diagnosed with an advanced breast cancer at any time during a period of 4.5 years from randomisation, including the period of active screening (48 months, annual if premenopausal, biennial if postmenopausal) and a period of follow-up after the last screen. Estimated primary completion year is 2030 (ClinicalTrials.gov Identifier: NCT03233191).

There are a number of aspects where TOSYMA may complement and expand the findings from these ongoing trials. Thus, TOSYMA impresses with its mere study size of 80 000 participants, which facilitates precise estimation of effects and a high statistical power for the evaluation of the primary endpoints. The established adaptive design provides the option to recalculate the sample size in the interim analysis and to modify the trial design as the need arises. Due to its multiple screening units, the trial guarantees a fast recruitment process. In addition, the TOSYMA protocol integrates all study procedures within the structure of the regular routine screening programme, which has an acknowledged high level of performance quality regarding image quality, technology and diagnostic procedures. Of note, participants will need no extra checks or recalls related to the trial; the trial is both multicentre and multivendor based. This concept of a truly pragmatic trial carries a high potential for external validity and translational research and, thus, for a swift transfer into screening practice if the trial results prove that DBT+s2D is superior to standard 2DFM. Furthermore, the clinical relevance of changes in screen-detected cancers needs to be backed by the evaluation of the dynamics of subsequently occurring interval cancers: this is methodologically ensured by state-wide population-based cancer registries. Additionally, effects of a higher detection rate by DBT might be found assessing the proceeding interval of the study plus the following screening examination regarding tumour stages between both study arms. A downstaging of screen-detected cancers in the subsequent round could be expected for example for invasive lobular cancers due to its growth pattern, visibility by architectural distortion and less aggressive features compared with other interval cancers. Furthermore, the trial protocol avoids augmentation of radiation doses, which is present in additive DBT protocols.

There will also be limitations. First, TOSYMA investigates the first-time use of DBT+s2D as a screening tool and findings might be different when using DBT in routine screening over multiple subsequent screening rounds. Thus, TOSYMA might include a relevant learning curve of readers although all study participants have to complete training units prior to the start of the clinical trial. Furthermore, differences between screening rounds might occur if the detection of cancers would differ regarding growing progression between 2D and DBT.

In conclusion, TOSYMA will address a clinically relevant topic. It may help to clarify the issue over whether the adoption of new technical developments is able to improve the effective screening for breast cancer in a population-based context.

Author affiliations
[1]Institute of Clinical Radiology and Reference Center for Mammography Münster, University of Münster and University Hospital Münster, Münster, Germany
[2]Institute of Biostatistics and Clinical Research, IBKF, University of Münster, Münster, Germany
[3]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany
[4]Center for Clinical Trials Münster, University Hospital Münster, Münster, Germany
[5]Reference Center for Mammography North, Oldenburg, Germany
[6]State Cancer Registry North Rhine-Westphalia, Münster, Germany
[7]State Cancer Registry Lower Saxony, Oldenburg, Germany

Acknowledgements The authors would like to acknowledge the support of the screening units represented by C. Barduna, J. Bruns, B. Dablé, C. Dietz, M. Ganseforth, A. Gräwingholt, M. Hirning, P. Keulers, U. Krause, T. Lehnen, P. Liersch, U. Meyer-Johann, D. Polenske, J. Porrmann, R. Rathmann, J. Ritter, A. Rudolf, K. Schulze, S. Spital, K. Struckmann, M. Stuhrmann, G. Süttmann, T. Töllner, G. Urban, B. Wein, I. Wendt. We would like to thank T. Decker for his collaboration related to the study.

Contributors Conception and design of the study: SW, JG, HWH, MK, AS, JC, HL, LK, WH. Statistical expertise: JG, LK. Study preparation: SW, JG, HWH, MK, AS, JC, HL, LK, KS, SD, SB, MU, GH, OH, JK, WH. Manuscript preparation: SW, JG, HWH, MK, AS, JC, HL, LK, KS, SD, SB, MU, GH, OH, JK, WH. Principal investigator: WH. All authors contributed to refinement of the study protocol and approved the final manuscript.

Funding This work is supported by the Deutsche Forschungsgemeinschaft (DFG—German Research Council—Project No. HE 1646/5-1). Publication fees are planned to be covered by the Open Access Publication Fund of the University of Münster.

Competing interests HWH and OH receives third-party funding from the German Federal Office for Radiation Protection (BfS, Bundesamt für Strahlenschutz) for a research project on the evaluation of the German mammography screening program. WH runs the National Mammography Reference Center Münster, a third-party-funded project at the University of Münster, Germany.

Patient consent Not required.

Ethics approval The study protocol has been reviewed and approved by the medical ethical committees (2016-132-f-S, Ärztekammer Westfalen-Lippe und der Westfälischen Wilhelms-Universität).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Robert Koch Institute and the Association of Population-based Cancer Registries in Germany. *Cancer in Germany 2011/2012*. 10th edn. Berlin: Robert Koch Institute and the Association of Population-based Cancer Registries in Germany, 2015.
2. Marmot MG, Altman DG, Cameron DA, *et al*. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013;108:2205–40.
3. Paci E. EUROSCREEN Working Group. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* 2012;19(Suppl 1):5–13.
4. Gotzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013;6:CD001877.
5. Lauby-Secretan B, Loomis D, Straif K. Breast-cancer screening—viewpoint of the IARC Working Group. *N Engl J Med* 2015;373:1479.
6. Perry N, Broeders M, de Wolf C, *et al*. *European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis*. Luxembourg: Office for Official Publications of the European Communities, 2006.
7. Rhiem K, Schmutzler RK. [Risk-adapted surveillance: focus on familial breast and ovarian cancer]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2014;57:307–11.
8. Bick U. Intensified surveillance for early detection of breast cancer in high-risk patients. *Breast Care* 2015;10:13–20.
9. Ciatto S, Houssami N, Bernardi D, *et al*. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol* 2013;14:583–9.
10. Skaane P, Bandos AI, Gullien R, *et al*. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013;267:47–56.
11. Friedewald SM, Rafferty EA, Rose SL, *et al*. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311:2499–507.
12. Skaane P, Bandos AI, Eben EB, *et al*. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology* 2014;271:655–63.
13. Gilbert FJ, Tucker L, Gillan MG, *et al*. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme – a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technol Assess* 2015;19:1–136.
14. Gilbert FJ, Tucker L, Gillan MG, *et al*. Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY Trial). *Radiology* 2015;277:697–706.
15. Zuckerman SP, Conant EF, Keller BM, *et al*. Implementation of synthesized two-dimensional mammography in a population-based digital breast tomosynthesis screening program. *Radiology* 2016;281:730–6.
16. Aujero MP, Gavenonis SC, Benjamin R, *et al*. Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. *Radiology* 2017;283:70–6.
17. Bernardi D, Macaskill P, Pellegrini M, *et al*. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. *Lancet Oncol* 2016;17:1105–13.
18. Garayoa J, Chevalier M, Castillo M, *et al*. Diagnostic value of the stand-alone synthetic image in digital breast tomosynthesis examinations. *Eur Radiol* 2018;28:565–72.
19. Caumo F, Zorzi M, Brunelli S, *et al*. Digital breast tomosynthesis with synthesized two-dimensional images versus full-field digital mammography for population screening: outcomes from the Verona screening program. *Radiology* 2018;287:37–46.
20. Houssami N. Evidence on synthesized two-dimensional mammography versus digital mammography when using tomosynthesis (three-dimensional mammography) for population breast cancer screening. *Clin Breast Cancer* 2017. 10.1016/j.clbc.2017.09.012 (Epub ahead of print 28 Sep 2017).
21. Heidinger O, Batzler WU, Krieg V, *et al*. The incidence of interval cancers in the German mammography screening program: results from the population-based cancer registry in North Rhine-Westphalia. *Dtsch Arztebl Int* 2012;109:781–7.
22. Urbschat I, Heidinger O. [Determination of interval cancer rates in the German mammography screening program using population-based cancer registry data]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2014;57:68–76.
23. Houssami N, Bernardi D, Caumo F, *et al*. Interval breast cancers in the 'screening with tomosynthesis or standard mammography' (STORM) population-based trial. *Breast* 2018;38:150–3.
24. Gemeinsamer Bundesausschuss. *Krebsfrüherkennungs-Richtlinie*, 2017.
25. Malek D, Kääb-Sanyal V. Implementation of the German Mammography Screening Program (German MSP) and first results for initial examinations, 2005–2009. *Breast Care* 2016;11:183–7.
26. Imai K. Statistical analysis of randomized experiments with non-ignorable missing binary outcomes: an application to a voting experiment. *Journal of the Royal Statistical Society: Series C* 2009;58:83–104.
27. Nekolla EA, Griebel J, Brix G. [Radiation risk associated with mammography screening examinations for women younger than 50 years of age]. *Z Med Phys* 2008;18:170–9.