Changing case Order to Optimise patterns of Performance in Screening (CO-OPS) Trial

RESEARCH PROTOCOL

Version 2.0
24th July 2013

NIHR Postdoctoral research fellowship for Dr Sian Taylor-Phillips reference NIHR-PDF-2011-01-073 Improving breast cancer screening detection rates through understanding, modelling, and adapting patterns of radiologist performance

Approved by the Coventry and Warwickshire NHS Research Ethics Committee on 27th June 2012 Reference 12/WM/0182

Sponsor: University of Warwick
1. Background

1.1 Epidemiology and burden of condition

Breast cancer is a leading cause of mortality in women, and was the cause of death for 10,280 women in England and Wales in 2010. It was responsible for 12% of all deaths of women in their fifties, and 8% of deaths of women in their sixties. [1]

In England 2.3 million women are projected to attend breast cancer screening each year, after the age extension of eligibility is increased to 47-73. [2,3] The cancer detection rate is currently 8 per thousand women screened. There are also 36 healthy women recalled for further tests (false positive recalls) per thousand screened. [2] 66% of all breast cancers in women aged 50-64 in England are detected by screening, [3] with the remaining 34% including women who choose not to attend screening, fast growing cancers which develop between screening rounds, and cancers missed at screening. The NHS Breast Screening programme (NHSBSP) aims to keep the number of cancers detected between screening rounds (interval cancers) down to 2.3 per thousand women, [4] however these are difficult targets and a more realistic rate may be more than 3.7 per thousand women. [4]

A related issue is that of over-diagnosis as a result of the introduction of breast cancer screening programmes, whereby cancers are detected at screening and treated, but this treatment is unnecessary because the woman would have died of other causes before the cancer ever grew to a stage where it would affect her health. It is very difficult to determine in the screening programme which cancers have been detected early and therefore saved the woman’s life, and which cancers would never have threatened her life.

Therefore missed cancers (false negatives) and recall of healthy women for further tests (false positive recalls) are significant issues in mammography screening, but must be considered in the context of possible over-diagnosis.

1.2 Existing knowledge

In the NHSBSP, each set of mammograms is read by two radiologists who each indicate if there should be a recall for further tests. Disagreements are referred for arbitration, usually by a third radiologist. Radiologists assess two mammograms per minute on average [5] which equates to 120 in a 1-hour session. This is a highly skilled, pressurised but repetitive and frequently intellectually unchallenging activity. A vigilance decrement of performance decrease over time has been observed in similar repetitive visual tasks such as radar operator. [6]

Data from an observational study using the Breast Screening Programme database for eight English breast screening centres for 3 years has shown that the probability of being recalled for further tests and of having cancer detected at screening is dependent on the time since the radiologist had a break from reading. For the first ten cases after a first reader break, the cancer detection rate for the whole process (including second reader and arbitration) is higher. These extra cancers detected are not associated with a higher probability of being DCIS rather than invasive, and have a similar mix of grades 1, 2 and 3 to the rest of the cancers detected at screening so this phenomenon is unlikely to be associated with increased overdiagnosis.
First and second readers read each batch of mammograms in the same order, thus their post-break increases in cancer detection rate occur when reading the same women’s mammograms.

### 1.3 Hypothesis

If the second radiologist examines the batch of mammograms in the reverse order to the first radiologist this will increase the overall cancer detection rates, because each radiologist’s peak performance will be when examining different cases.

### 1.4 Need for a Trial

A previous observational study has shown a reduction in sensitivity with a proxy for time (number of cases) since a break. Currently both radiologists are presented with cases in the same order as one another, so both have reduced sensitivity when reading the same cases. Therefore changing the order in which the second radiologist is presented with the mammograms may increase sensitivity of screening. A trial is required to determine whether such improvements would be realised in practice.

### 1.5 Good Clinical Practice

The trial will be carried out in accordance with Good Clinical Practice (GCP) as detailed in the Medical Research Council (MRC) GCP guidelines [here](http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002416) and in accordance with the following protocol.

### 1.6 CONSORT

The trial will be reported in line with the CONSORT (*Consolidated Standards of Reporting Trials*) statement, and the further guidance available for cluster randomised controlled trials.

### 2 Trial Design

#### 2.1 Trial summary

The trial is a multi-centre cluster randomised parallel group superiority trial. The centres will be English Breast Screening centres with digital mammography equipment. The intervention is reversing the order in which the second radiologist reads each batch of mammograms. A cluster for randomisation is a batch of mammograms (circa 30 women’s mammograms). The unit for analysis is the individual women’s outcome.

The pre-pilot will be conducted in three screening centres for two weeks to ascertain the practical issues with implementation, and to test the Crystal report for data extraction. Concurrently a survey of all English breast screening centres will be conducted to ascertain key characteristics relevant to the trial which may differ between centres, such as the blinding of one reader to another’s decision, and the method of arbitration used.

The trial will be implemented using an adaptation to the NBSS software used at participating centres.
The trial will last 16 months in 44 centres (1.5 million women). The first 12 months data (1.1 million women) will be collected from the NBSS database to determine whether the intervention affected cancer detection rate.

The full data set will be collected three years after the end of the study, to ascertain whether there was a decrease in rate of interval cancers (cancers presenting symptomatically in the three years after screening), and therefore whether the intervention reduced the number of cancers missed at screening.

3. Aims and Objectives of the Trial

3.1 Primary aim
To determine whether reversing the order in which the second reader reads the mammograms affects cancer detection rate.

3.2 Secondary aims

To investigate the mechanism by which the intervention is hypothesised to increase cancer detection rate. This is made up of several parts:

1. What effect does the intervention have on recall rate?
2. How well does the arbitration process perform?
3. Does reversing the order in which the second reader reads the mammograms increase disagreement between readers
4. Does cancer detection rate and recall rate change with time since a break?

To determine whether reversing the order in which the second reader reads the mammograms affects missed cancer rate.
3.3 Outcome Measures

a) Efficacy:
   i) Primary

Cancer detection rate in intervention and control group.

ii) Secondary
   1. Recall rate in intervention and control groups
   2. Referral rate to arbitration, recall rate and cancer detection rate from arbitration, and positive predictive value of arbitration by type of arbitration used. Sensitivity and specificity of arbitration using 3 years follow up data
   3. Disagreement rate in intervention and control group
   4. Change in disagreement rate, recall rate, cancer detection rate, and missed cancer rate with intended reading order (and cases since a cancer, recall, and mid-batch break)
   5. Rate of cancers detected symptomatically in the three years subsequent to screening in intervention and control group. To be collected three years after the trial implementation is complete. This provides a measure of the difference in rates of cancer missed.

iii) Health economics
    Cost per extra cancer detected
    Cost per quality adjusted life year gained from the intervention
2.2 Trial flow diagram
CONSORT 2010 Flow Diagram for cancer detection analysis after 12 months

Enrollment of women screened

Assessed for eligibility (n=1,100,780)
- Excluded (n=0)
  - Not meeting inclusion criteria (n=0)
  - Declined to participate (n=0)

Randomized (n=1,100,780)
- In batches of ~30
  - (44 centres for 1 year)

Allocation

Allocated to case order reversal intervention for second reader (n=550,390)
- Received allocated intervention (n= 547,890)
- Did not receive allocated intervention but included in analysis (radiologists examining cases more than once or changing case order) (n=2250)

Allocated to control (n=550,390)
- Received control (n=550,390)
- Did not receive control (n=0)

Follow-Up

Lost to follow-up (NBSS records not updated with cancer cases) (n=2,500)

Lost to follow-up: Recalled to assessment but did not attend (n=1,000)

Discontinued intervention (n=0)

Analysis

Analysed (n= 546,890)
- Excluded from analysis (n=0)

Analysed (n=546,890)
- Excluded from analysis (n=0)
3.4 Statistical Analysis Plan

i) Primary

To determine whether cancer detection rate is higher in the intervention group in comparison to the control group analysis will be conducted using a multi-level logistic regression model.

Levels to be included in the model are case, batch, and centre. To prevent over-fitting each level will only be included if they explain a sufficient portion of the variability.

It is reasonable to expect characteristics at the level of the women screened to be very similar between intervention and control due to the large number of batches of women randomised. Similarly centre level characteristics such as the arbitration system used will not differ between intervention and control as randomisation is at the batch level within each centre. Therefore the main analysis will simply contain one predictor, intervention or control arm. The intervention will be considered effective if the 5th percentile of the distribution of the coefficient for the intervention in the model is greater than zero.

A second analysis will be conducted including adjustment for characteristics of the woman screened (age and whether they have previously attended screening). Sub-group analysis will be conducted in younger women, those whose cases are read at the beginning or end of the batch, and younger women, as the intervention may be more effective in these groups.

ii) Secondary

Analysis of recall rate and rate of disagreements between intervention and control group will use the same methods as for the primary analysis. The positive predictive value of intervention and control group will also be calculated.

4. Power and Sample Size

The trial is powered to detect at the 5% significance level with 80% power in a two tailed test a difference in cancer detection rate between intervention and control group of one extra cancer detected per 2000 women screened. This would correspond to around 1000 extra cancers detected by screening in England each year. Power calculations used cancer detection rate in the control group (7.8 per thousand women screened) and cancer detection rate in the intervention group (8.3 per thousand women screened). These estimates of cancer detection rate are derived from data from a previous study.

ICC was calculated from previous data, using a logistic binomial-Gaussian model (method B) with 1000 Monte Carlo simulations.\(^7\) Hence, using the derived ICC of 0.002 and a cluster (batch) size of 40 women, this then gives the design effect as 1.09. This gave an overall sample size required of 1,093,780 women, or 44 breast screening centres for 1 year (On the basis that in England there are 82 centres each screening around 25,000 women per year).
There is no adjustment for drop out or crossover because once the intervention is applied to a screening centre each batch will automatically be randomly assigned to intervention or control groups by the NBSS computer system, and the intervention applied automatically by that same system. However, there is a possibility that a small minority of records (circa 5000) will not be updated in time for the data collection. These will be considered missing at random as they will be all the cases from certain dates. A further 2000 women may be recalled for further tests and not attend their appointment. Furthermore, some radiologists may not read cases in the intended order, for example coming back to re-read difficult cases at a later time. This also is expected to be very rare in a busy screening centre, (circa 2250 in intervention arm), these cases will be included in the analysis. Therefore 1,100,780 cases are required to be randomised, equivalent to 44 screening centres for 1 year.

The same method was applied to detect a difference in cancer detection rate between intervention and control group of one fewer cancer missed per 2500 women screened. This gave an overall sample size required of 1.5 million women, or 44 breast screening centres for 1 year and 4 months.

Therefore 44 breast screening centres will be recruited to the trial, which will last for 1 year and 4 months. After the first year of the trial (allowing 2 months for follow up appointments to determine which women have cancer) the data for one year of screening at 44 centres will be downloaded, and analysis of the primary outcome of cancer detection rate, and the secondary outcomes to be completed. Analysis of missed cancer rates will not be until three years after completion of the whole trial duration.

There are 82 Breast Screening centres in England, of which 48 are fully digital, and 25 are partly digital. It is anticipated that by the start of data collection 64 centres will be fully digital. It is anticipated that it will take three months to enrol 44 centres and obtain local R&D approvals. Administering the trial at each centre will take the NHS staff no extra time commitment as it is fully computerised and will be automatically downloaded with their latest software update, which is anticipated will help reach recruitment targets.

5. Eligibility
5.1 Inclusion criteria
Screening centres taking part must use double reading of screening mammograms, and must have at least one piece of digital mammography equipment used for screening
All women who receive mammography screening using digital equipment during the study period at the study centres.

5.2 Exclusion criteria
Centres which use single reading of mammograms
Centres which have no digital mammography equipment used for screening (the software intervention does not work on the old equipment)
Women who attend symptomatic breast clinics
5.3 Post-randomisation withdrawals and exclusions
Centres participating may withdraw from the trial treatment, and/or the whole trial at any time without prejudice. No follow up data will be collected, only the initial survey results describing the centre characteristics will be used to analyse whether there are patterns in withdrawals from the trial, unless the centre requests that their data is not used in this way.

5.4 Compliance/contamination
Compliance is expected to be very high, as after centre level consent is given all implementation is automated. Therefore deviation from the protocol would be very difficult. Similarly contamination is not anticipated to be an issue as there is no method available to move cases between the intervention and control groups. However case order can be changed by selecting the ignore option rather than inputting a screening result and reviewed at a later stage, or a result could be inputted and then revised later. These things would change the reading order and produce some level of contamination. This will be ameliorated by analysing based on intended order of reading. Data will be collected on actual reading order also to measure levels of contamination and their effect on outcomes. These effects are expected to be extremely small (of the order of once per 500 cases) because screening is a fast moving high volume activity and there is no time in practice to come back to cases at a later stage.

6. Consent
Informed consent will be obtained at the centre level, by the director of the service and the lead radiologist. These are usually the same person but if they are not the consent of both will be sought. Consent will be required at the centre level rather than the patient or radiologist as it is at this level that the intervention is applied. The intervention can be considered an alternative and at least equivalently good form of standard practice, as nothing about how the mammograms are reviewed and evaluated changes, just the order in which they are assessed.

Directors of breast screening centres will be contacted in the first instance by email introducing the study with a copy of the participant information sheet and informed consent form attached. They will then receive a follow up phone call one week later. Each centre will be offered a copy of the research protocol and IRAS ethics form, and a visit from one of the investigation team to introduce the study if they wish. The study software which will run the trial will be embedded in the NBSS software in every English breast screening as part of the routine updating of the NBSS system. This software will remain inactive until a signed consent form is received from that centre, at which point it can be activated simply and quickly. The study can commence in each centre by simply changing the software settings to activate the intervention.

In the unlikely event that information becomes available may be relevant to the participant’s willingness to continue in the trial, for example findings from other research studies, then the directors of every participating breast screening centre will be informed immediately by email and follow up phone call.
6.1 Recruitment and Randomisation
The rate of accrual will be monitored at the centre level, if it falls appreciably below the projected level, the reasons will be identified and remedial actions taken in order to protect the power of the trial and alleviate concerns about selective entry and other aspects of quality.

Randomisation will be automatically computer generated at the point at which the batch is ready to be read on the NBSS system. The random number generation will be using the root of the time at which randomisation is required. Only batches of cases to be read as part of the NHS breast screening programme will be randomised. Cases from symptomatic clinics will be excluded prior to randomisation.

6.2 Blinding
There is no expected placebo effect in this trial. No elements of standard practice are to be changed as part of the trial except reading order, to test whether the intervention would be effective in normal practice. The women screened will not be aware whether their mammograms are read as part of an intervention or control batch. The radiologists and radiography advanced practitioners when acting as second reader are not blinded to reading order, but they are blinded to trial arm because both intervention and control groups include reading forwards and backwards. The statistician will be blinded to treatment groups until analysis is complete.

7. Project Timetable and Milestones
Year 1
August 2012: Software developed, ethical approvals complete and trial registered
December 2012: Pilot complete and 44 sites recruited

Year 2
Implementation in all 44 sites throughout 2013

Year 3
May 2014: 1 year data collected in 44 centres.
September 2014: Write up complete
December 2014: Heath economics analysis complete

Year 6
Collection and analysis of interval cancer data. Follow on funding is required to proceed with this part of the project.

8. Protocol Amendments

Version 2.0: agreed on 24th July 2013 – number of centres increased from 36 to 44. (Coventry and Warwickshire REC on 12th July 2013 approved minor amendment to increase the number of centres from 36 to 44).
9. References


