



NIH PUBLIC ACCESS

Author Manuscript

*Acad Radiol.* Author manuscript; available in PMC 2007 September 11.

Published in final edited form as:

*Acad Radiol.* 2007 June ; 14(6): 670–676.

## Two-Modality Mammography May Confer an Advantage over Either Full-Field Digital Mammography or Screen-Film Mammography

Deborah H. Glueck, PhD<sup>1</sup>, Molly M. Lamb, BA<sup>1</sup>, John M. Lewin, MD<sup>2</sup>, and Etta D. Pisano, MD<sup>3</sup>

*1*Department of Preventive Medicine and Biometrics, University of Colorado at Denver and Health Sciences Center, Denver, CO

*2*Diversified Radiology of Colorado, Denver, CO

*3*Departments of Radiology and Biomedical Engineering, the Biomedical Research Imaging Center and the Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC

### Abstract

**Rationale and Objectives**—To compare the cancer detection rate and ROC area under the curve of full-field digital mammography, screen-film mammography, and a combined technique that allowed diagnosis if a finding was suspicious on film, on digital, or both.

**Materials and Methods**—We used the data originally analyzed in Lewin et al. (2002). In that trial, 6,736 paired full-field and digital mammograms were performed in 4,489 women. We used parametric and nonparametric tests to compare the area under the curve for ROC scores of film-screen only, digital mammography only, and the combined test. We used McNemar's test for paired proportions to compare the cancer detection rates.

**Results**—With the parametric test, neither the difference in AUC between the film and combined, nor the difference between the digital and combined ROC curves was significant at the Bonferroni-corrected 0.025 alpha level (film vs. combined difference = 0.0563,  $p = 0.0712$ ; digital vs. combined difference = 0.0894,  $p = 0.0455$ ). The nonparametric test showed that there was a significant difference between both film and combined (difference = 0.073,  $p = 0.008$ ) and digital vs. combined ROC curves (difference = 0.1164,  $p = 0.0008$ ). The continuity corrected McNemar's test showed a significant increase in the proportion of cancers detected by the combined modality over film (chi squared = 7.111,  $df = 1$ ,  $p = 0.0077$ ), and over digital (chi squared = 12.071,  $df = 1$ ,  $p = 0.0005$ ).

**Conclusion**—Using two mammograms, one film and one digital, significantly increases the detection of breast cancer.

### Keywords

Digital mammography; film mammography; combined test; ROC area under the curve analysis

---

Corresponding Author: Dr. Deborah H. Glueck, PhD, University of Colorado at Denver and Health Sciences Center, 4200 East 9<sup>th</sup> Ave, Box B-119, Denver, CO 80262, Phone: (303) 315-7227 Facsimile: (303) 315-3183 Email: [Deborah.glueck@uchsc.edu](mailto:Deborah.glueck@uchsc.edu)

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Introduction

Three recent clinical trials with paired designs have compared full field digital and screen-film mammography (1–3). Each woman was imaged with both digital and film machines. The purpose of these trials was to compare the diagnostic accuracy of the digital and film approaches. Interestingly, in each study, more cancers were detected than would have been detected by either modality alone (Table 1). This suggests that full-field digital mammography and screen film mammography could be used in tandem to improve breast cancer detection.

Although doing two different mammograms, with two different readers is burdensome and costly, there is historical precedent for dramatically increasing the screening protocol. Sickles et al. (4) demonstrated that doing both mediolateral oblique and craniocaudal views increased cancer detection rates. He recommended always doing both views, which became the new standard in the field. Although doing two views was twice as expensive and required much more radiologist time, it allowed the detection of more cancers, which “outweigh(ed) the additional radiation risk and added cost”.

We define a *combined test* as a test that declares a woman to have cancer if the cancer is detected by film mammography alone, by digital mammography alone, or by both modalities. The ROC score assigned is the maximum of the film and digital scores. We assume that a woman would be recalled for a biopsy if either modality suggested follow-up work. We hypothesize that such a combined test will give a larger area under the ROC curve (AUC) than either modality alone. Additionally, we hypothesize that the total number of cancers detected by the combined test will be significantly higher than the total number of cancers detected by either modality alone.

## Materials and Methods

Lewin et al. (1) conducted a trial of 4,489 women, who received both full-field digital and screen-film mammography. They analyzed the results of 6,736 paired examinations, and tested the difference in area under the free response operating characteristic curve (5,6) for the two modalities. We re-analyzed the same data in order to test our new hypotheses.

To conduct the ROC analysis we emulated the methodology of the Digital Mammographic Imaging Screening Trial (DMIST). We used a parametric binormal ROC technique for paired data (7) ([www-radiology.uchicago.edu](http://www-radiology.uchicago.edu)) for the primary analysis. A nonparametric AUC test for paired data (8) was conducted as a planned confirmatory analysis. Both Hajian-Tilaki et al. (9) and Zhou (10) point out that parametric and nonparametric methods produce roughly the same result, without appreciable error or bias.

We used the continuity corrected McNemar’s test for paired nominal data (11) to compare the proportion of cancers detected by the combined modality, versus either modality alone. Because we are testing two hypotheses in both the ROC analysis and in the McNemar’s analysis, (combined modality versus film, and combined modality versus digital) we controlled for multiple comparisons. Thus, we tested each hypothesis at a Bonferroni corrected alpha level of 0.025.

In the trial conducted by Lewin et al. (1), the radiologist produced an ROC score for each finding seen on digital mammography, and a similar ROC score for each finding seen on film mammography. There was no overall ROC score for the digital mammogram or the film mammogram. For each exam, we took the overall maximum digital ROC score, the overall maximum film ROC score, and the overall maximum combined ROC score. The overall maximum combined ROC score provided a measure of how confident the radiologist was that either modality had detected a malignancy. For the breasts which were proven to contain

malignancy, we verified that the highest score actually was assigned to the finding that was diagnosed as cancerous.

Lewin et al. (1) reported a total of 42 cancers diagnosed by mammography, 33 diagnosed on film, 27 diagnosed on digital, 18 diagnosed by both, and 8 interval cancers. After a thorough review of the data, we discovered that one patient was counted as having two cancers, in the same quadrant of the same breast. Biopsy records show that there was one cancer. Therefore, our analysis used a total of 41 cancers detected by mammography: 32 detected on film, 27 detected on digital, and 18 detected by *both* film screen and digital mammography. Additionally, there were 8 interval cancers. The 8 interval cancers plus the 41 cancers detected by mammography yielded a total of 49 cancers diagnosed in this study population.

## Results

The primary analysis used the parametric binormal ROC technique (7). The difference between the digital and film ROC curves was not significant (difference = 0.0418,  $p = 0.51$ ), verifying the results of Lewin et al. (1). The difference between the film and combined ROC curves was also not significant at the Bonferroni-corrected 0.025 alpha level (difference = 0.0563,  $p = 0.0712$ ), as was the difference between the digital and the combined curves (difference = 0.0894,  $p = 0.0455$ ). (Figure 1 and Table 2)

Interestingly, the nonparametric tests (8) that were run to confirm the parametric tests showed a significant difference between the film vs. combined ROC curves (difference = 0.073,  $p = 0.008$ ) and digital vs. combined ROC curves (difference = 0.1164,  $p = 0.0008$ ). As expected, the nonparametric test of the film vs. the digital ROC curves was not significant (difference = 0.0434, 0.3863) (Figure 2 and Table 3).

Although the direction of the difference in AUC is the same in both the parametric and nonparametric tests, the parametric  $p$ -values were not significant, while the nonparametric ones were. The two types of tests will usually produce similar  $p$ -values. However, they use estimation methods that are different enough that occasionally, one will be significant while the other is not. The major theoretical reason why the areas under the curve are different is that the parametric method used a binormal fit, while the non-parametric method used a trapezoidal rule to estimate the area. Thus, the estimated operating points of the two methods have different locations. In the presence of these conflicting results, a definite conclusion regarding the benefit of the combined modality based on ROC curves cannot be drawn. Future studies would be needed to confirm the ROC result.

Another approach to understanding the contribution of a combined testing approach is to focus on the number of cancers detected. Screen film mammography detected 65.3% of the 49 cancers, full field digital mammography detected 55.1%, and the combined modalities detected 83.7% (Table 1). The continuity corrected McNemar's test showed a significant increase in the proportion of cancers detected by the combined modality over film (chi squared = 7.111,  $df = 1$ ,  $p = 0.0077$ ), and over digital (chi squared = 12.071,  $df = 1$ ,  $p = 0.0005$ ). Again, the film versus digital comparison showed no significant difference (chi squared = 0.696,  $df = 1$ ,  $p = 0.4042$ ).

## Discussion

The increase in area under the curve for the combined modality reflects a larger increase in sensitivity than the decrease in specificity. The ROC analysis is important, because it allows researchers to look at the performance of diagnostic methods over the entire range of sensitivity. However, for many women with breast cancer, whether their breast cancer is detected is the most important factor in measuring the benefit of screening mammography. The McNemar's

tests directly address this issue, by comparing the proportion of cancers that were diagnosed by each method.

In the Lewin trial (1), the detection rate of both modalities combined was higher than either film mammography or digital mammography alone (Table 1). The number of cancers diagnosed by combined modalities may have increased for any of three reasons: 1. Film-screen mammography and full-field digital mammography detected different populations of cancers, 2. Four different views of each breast (cranial-caudal and mediolateral on both digital and film) totaling four different compressions of each breast increased detection. 3. Independent reading by two different radiologists increased detection.

Double reading of mammography has been shown to increase detection. The existing data on double reading mostly comes from Europe. Anttinen et al. (12) looked at double reading in a screening population of 17,000 women in Finland, and suggested that the double reading increased cancer detection rates by 9%. Thurfjell et al. (13) conducted a study of double reading in a Swedish population of more than 11,000 women and increased the number of cancers detected by almost 15%. Ciatto et al. (14) evaluated the effect of a simulated double reading technique in a highly enriched training set and found that that double reading improved sensitivity from 50.2% to 64.8%.

If double reading accounts for only some of the observed increase in cancer detection, it seems likely that increasing the number of views or using two different modalities was an additional cause of the better performance. We believe that two clinical trials would be needed in order to determine whether the number of compressions or the different modalities increased the number of cancers detected. It cannot be determined which factor is responsible without additional data collection.

Lewin et al. (1, Table 1) summarized the reasons why findings were only detected by one modality. The most common reason cited was “fortuitous positioning”. For demonstration, Lewin et al. (1, Figure 1) showed a mammogram in which a density is visible on a screen-film mammogram due to overlapping tissue, but invisible on a digital mammogram due to different positioning and compression. We suspect that in the Lewin trial (1), positioning and compression differences affected the detection rate, but we cannot tell the proportion of increase in detection that was due to this reason. The DMIST group is currently conducting an analysis of their images to determine the extent that positioning and compression differences or other factors affected the detection rate.

In fact, with the current clinical trial data, we cannot determine whether the number of readers, the number of compressions, or the use of two different modalities is responsible for the increased cancer detection rate. Lewin et al. (1) and Pisano et al. (3) did not intend to answer these questions, or even to consider using two-modality mammography as a screening method. In the section that follows, we describe the study designs that would be needed to answer these questions. These studies have not been carried out, and are purely theoretical exercises designed to demonstrate how one could differentiate between the proposed different causes of the increased detection rate.

A three arm randomized controlled trial is the best way to test whether the two screening modalities find different populations of cancer. The three arms would employ the following screening modalities: arm one: two film mammograms; arm two: two digital mammograms; arm three: one digital and one film mammogram. Each mammogram would include both mediolateral oblique and craniocaudal views. Each mammogram would be read by one of two readers, randomly chosen from a pool of qualified mammographers. A woman would be recalled for a biopsy if either reader suggested follow-up work. This trial fixes the number of readers (two), and the number of views (four). The only factor remaining is the modality.

The following four arm clinical trial could test whether the number of compressions increases the detection rate. The four arms would employ the following four different screening modalities: arm one: two film mammograms; arm two: one film mammogram; arm three: two digital mammograms; arm four: one digital mammogram. Each mammogram would include two views (mediolateral oblique and craniocaudal), and two compressions. There would be one reader for each exam, assigned at random from a pool of trained mammographers. We would compare the single film mammogram arm to the double film mammogram, and the single digital mammogram arm to the double digital mammogram arm. This trial fixes the number of readers (one), and the modality (film or digital). The only factor left that could affect the number of cancers detected is the number of views.

We speculate that using two different sorts of machines will improve cancer detection more than increasing the number of views. In the DMIST trial, digital mammography had an advantage in young women, pre-menopausal women, and women with dense breasts. One explanation of this finding is that there are three different populations of cancer/background dyads. One population of cancer/background dyads is film-detectable. The second population of cancer/background dyads is digital-detectable. The third population cannot be detected on either modality, either because the cancer has not developed yet or because it is invisible against its background.

How big might the increase in detection due to using the combined modality be? We have no data to answer this question. However, published sources of information can help us to guess the size of the increase. We made several assumptions so that we could conduct a thought experiment. First, we assumed that the Lewin trial (1) missed approximately 10% of the total number of cancers, i.e. that 10% of the cancers were neither detected by film, nor by digital, nor seen as interval cancers. Second, we extrapolated from the increased detection rate seen in three double *reader* studies (Ciatto, Thurfjell and Antinnen, 12–14) to a double modality/double reader study. This extrapolation led to the assumption that performing two digital mammograms or two film mammograms would increase the overall number of cancers detected by some 10%. Finally, we assumed that the proportion of interval cancers, and the proportions of cancer/background dyads detectable by film mammography alone, digital mammography alone, or both film and digital mammography occurred in the population at roughly the rates seen in the Lewin trial (1).

These four populations are shown in Figure 3A. 3B shows that double screening and double reading with film mammography will capture roughly the same cancers each time, resulting in the detection of approximately 65% of all cancers. 3C shows similarly that double screening and double reading with digital mammography will also capture roughly the same cancers each time, resulting in the detection of approximately 55% all cancers. Notice that these cancers are mostly from a different population of cancer/background dyads from those detected in 3B. Finally, 3D shows that double screening, once with film and once with digital mammography, should detect more cancers (roughly 75% of all cancers) since the combined modality has greater efficacy in two different populations of cancer/background dyads. These numbers are speculative (see assumptions described above), and only give an approximation of what this trial might yield.

What we cannot know is what fraction of the cancers not detected by a given modality in the Lewin trial are truly not detectable by that modality. In other words, what fraction of the assumed 10% increase in cancer detection from a double film or double digital trial would come from cancers seen only by the other modality in the Lewin trial, what fraction would come from the interval cancers and what fraction would come from a presumed fourth population of cancers present in the population but not detected by any means in the Lewin trial.

The trials described above would raise many ethical and monetary issues. Screening twice increases the radiation dose to the breast. Double screening certainly increases the recall rate and the biopsy rate, which leads to increased patient anxiety and morbidity. Double screening will also increase cost. It is more expensive to acquire mammograms using two different modalities than to acquire two different mammographic views. The breast screening clinics would need to buy and maintain two machines, and there would be a large increase in the need for radiologist time. Additionally, it will increase time, effort, and discomfort for the woman.

From the perspective of the insurer, double screening is too costly. From the perspective of a clinician, double screening may seem too time-consuming. From the perspective of a patient, however, the increased cancer detection rate may outweigh any other consideration. The increase in the number of cancers detected by the combined modality in the Lewin (1) should draw immediate patient and physician interest.

### Acknowledgements

Many thanks to Dr. Lorenzo Pesce for his assistance with the parametric ROC analysis. Glueck and Lamb were supported by NCI grant K07CA88811. Drs. Lewin and Pisano are mentors to Glueck on this grant.

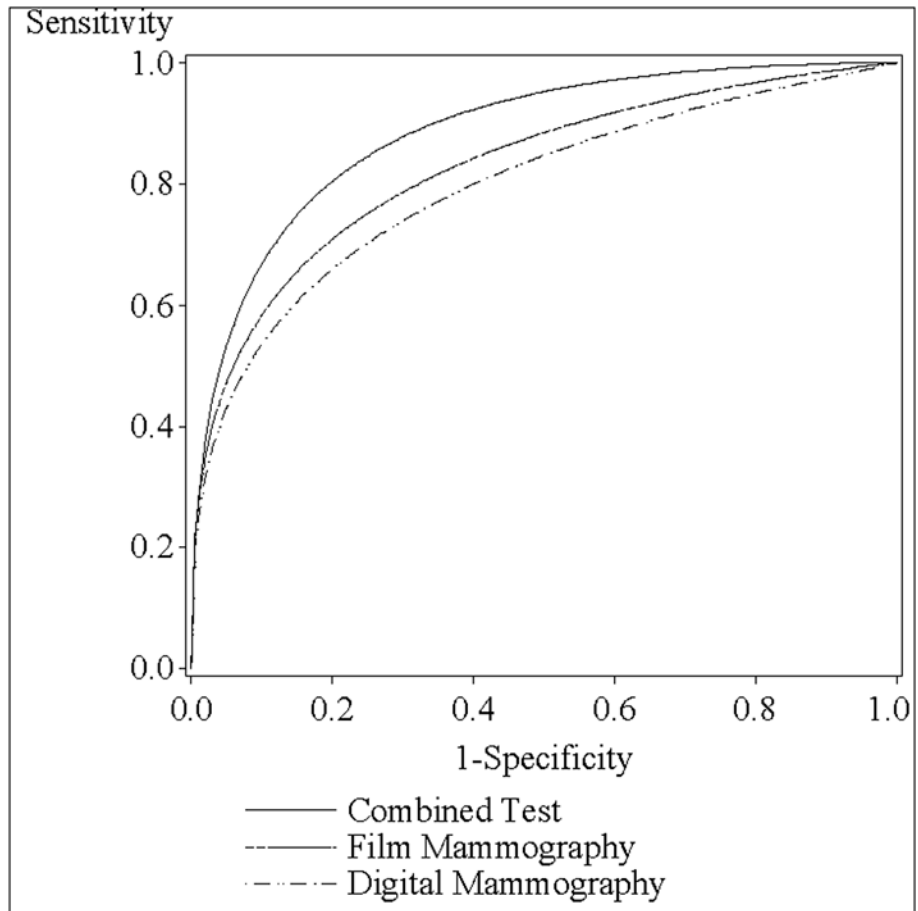
Research supported by NCI K07CA88811.

### References

1. Lewin JM, D'Orsi CJ, Hendrick RE, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *AJR Am J Roentgenol* 2002;179:671–677. [PubMed: 12185042]
2. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading--Oslo I study. *Radiology* 2003;229:877–884. [PubMed: 14576447]
3. Pisano ED, Gatsonis C, Hendrick E, et al. Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005;353:1773–1783. [PubMed: 16169887]
4. Sickles EA, Weber WN, Galvin HB, et al. Baseline screening mammography: one vs two views per breast. *AJR Am J Roentgenol* 1986;147:1149–1153. [PubMed: 3490749]
5. Chakraborty DP, Winter LH. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology* 1990;174:873–881. [PubMed: 2305073]
6. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839–843. [PubMed: 6878708]
7. Metz, C.; Wang, P.; Kronman, HA. New approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck, F., editor. *Information processing in medical imaging*. The Hague, the Netherlands: Nijhoff; 1984.
8. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845. [PubMed: 3203132]
9. Hajian-Tilaki KO, Hanley JA, Joseph L, et al. A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic tests. *Med Decis Making* 1997;17:94–102. [PubMed: 8994156]
10. Zhou, X-H.; McClish, DK.; Obuchowski, NA. *Statistical Methods in Diagnostic Medicine*. New York, NY: John Wiley and Sons Inc; 2002.
11. Daniel, WW. *Applied Nonparametric Statistics*. Second Edition. Pacific Grove, California: Duxbury Thomson Learning; 1990.
12. Anttinen I, Pamilo M, Soiva M, et al. Double reading of mammography screening films--one radiologist or two? *Clin Radiol* 1993;48:414–421. [PubMed: 8293648]
13. Thurfjell EL, Lernevall KA, Taube AA. Benefit of independent double reading in a population-based mammography screening program. *Radiology* 1994;191:241–244. [PubMed: 8134580]

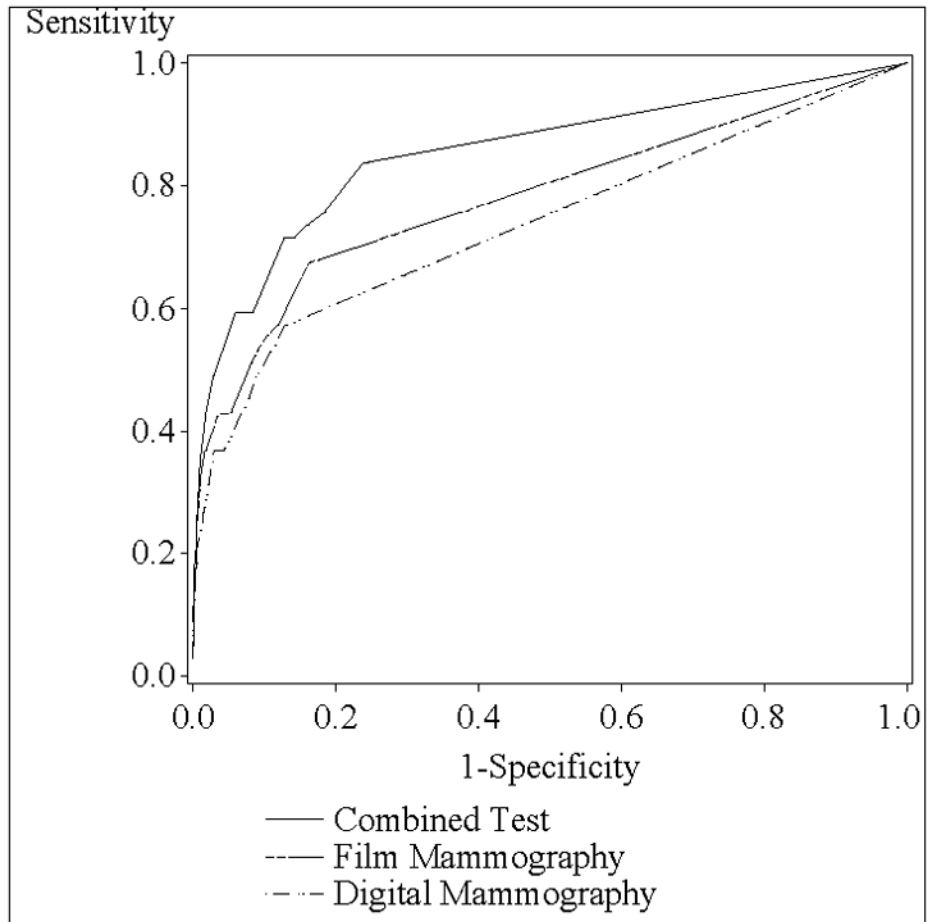


14. Ciatto S, Rosselli Del Turco M, Burke P, et al. Comparison of standard and double reading and computer-aided detection (CAD) of interval cancers at prior negative screening mammograms: blind review. *Br J Cancer* 2003;89:1645–1649. [PubMed: 14583763]

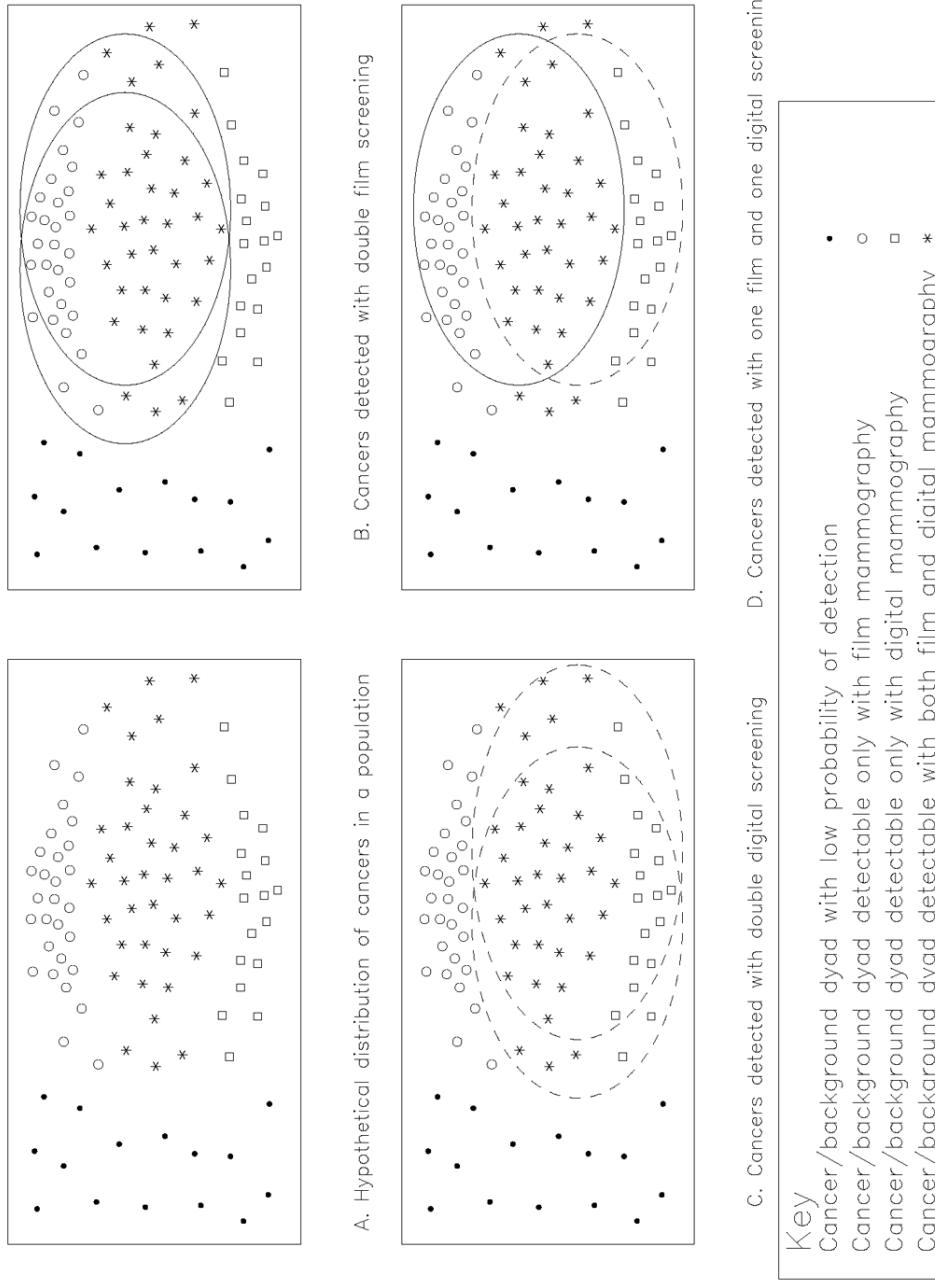


**Figure 1.** ROC curves for Film Mammography Results, Digital Mammography Results and the Combined Test Results in the Parametric Analysis.





**Figure 2.**  
Graph of AUC lines for ROC Nonparametric Analysis.



**Figure 3.**  
Hypothetical Cancer Detection Results of Proposed Trial #2.

**Table 1**  
Proportions of Cancer Detected by Each Modality in Three Clinical Trials

<b>Trial</b>	<b>Screen film mammography</b>	<b>Full field digital mammography</b>	<b>Combined modalities</b>
Lewin et al.(1)	32 / 49 = 65.3%	27 / 49 = 55.1%	41 / 49 = 83.7%
Skaane et al. * (2)	28 / 31 = 90.3%	23 / 31 = 74.2%	31 / 31 = 100%
Pisano et al. (3)	174 / 335 = 51.9%	185 / 335 = 55.2%	237 / 335 = 70.7%

\* No interval cancers were reported in this study

**Table 2**

## Parametric ROC Analysis Results

Statistics	Film vs. Combined	Digital Vs. Combined	Film vs. Digital
First modality AUC	0.831	0.7956	0.8283
Standard Error	0.0406	0.0267	0.0423
Second modality AUC	0.8873	0.885	0.7865
Standard Error	0.0259	0.0509	0.0538
AUC difference	0.0563	0.0894	0.0418
Standard Error difference	0.03122	0.04467	0.06357
Z-score	1.8045	1.9996	0.6587
p-value	0.0712	0.0455	0.5101

**Table 3**

## Nonparametric ROC Analysis Results

Statistics	Film vs. Combined	Digital Vs. Combined	Film vs. Digital
First modality AUC	0.7811	0.7377	0.7811
Standard Error	0.037	0.0381	0.037
Second modality AUC	0.8541	0.8541	0.7377
Standard Error	0.0315	0.0315	0.0381
AUC difference	0.073	0.1164	0.0434
Standard Error difference	0.0275	0.0347	0.0501
Chi squared (df = 1)	7.0397	11.281	0.7507
p-value	0.008	0.0008	0.3863