



139
247
THS

This is to certify that the
thesis entitled

AN IMPROVEMENT IN THE METHOD OF ASSESSING
SPECIFICITY OF LUMP DISCRIMINATION IN MAMMACARE
SILICONE MODELS

presented by

MICHAEL ROGER BRENNAN

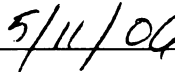
has been accepted towards fulfillment
of the requirements for the

Master of
science

degree in

Epidemiology


Major Professor's Signature



Date

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY
Michigan State
University

**AN IMPROVEMENT IN THE METHOD OF ASSESSING SPECIFICITY OF
LUMP DISCRIMINATION IN MAMMACARE® SILICONE MODELS.**

By

Michael Roger Brennan

A THESIS

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

MASTER OF SCIENCE

DEPARTMENT OF EPIDEMIOLOGY

2006

ABSTRACT

AN IMPROVEMENT IN THE METHOD OF ASSESSING SPECIFICITY OF LUMP DISCRIMINATION IN MAMMACARE® SILICONE MODELS.

By

Michael Roger Brennan

When teaching the clinical breast examination, the sensitivity and specificity of lump detection is often assessed using six silicone breast models with a known number and location of lumps. The current method of calculating the specificity of lump detection is inaccurate due to the use of the entire silicone model as the unit of analysis.

Using a computer simulation of the clinical breast exam a new cellular method of evaluating learners' sensitivity and specificity was created. This new method using cells was then applied to actual data.

The computer simulations of 448 virtual examinations of 2,688 virtual breasts were performed. The results of the simulation and application to actual data demonstrated the current method of evaluating specificity was invalid.

The results of the simulation and the data indicate that when educators assess learners' clinical breasts exam skills using silicone models a cellular approach to calculating specificity should be applied.

Copyright by
MICHAEL ROGER BRENNAN
2006

Dedication

To anyone's hands that become better at CBE as a result of this thesis, and the lives they help.

Acknowledgement

I must thank Dr. Janet Rose Osuch for all of her help and in being a great friend. She has been my friend and mentor for the last six years, without her I would be lost and less of the person I am today. I thank Dr. Paneth his discussions over my career have been invaluable. Dr. Barry I thank for having interest in the computer aspects of my thesis, and keeping me motivated to create the simulation. Dr. Dorothy Pathak who first introduced me to the MammaCare system, and information I learned from her grant, DAMD 17-98-1-8118, entitled Improved Follow-up of Breast Abnormalities Through Comprehensive Breast Care in Women 40 to 70 years of age. The data from a study sponsored by Michigan State University Institute for Health Care Studies with support by the Michigan Cancer Consortium Division of the Michigan Department of Community Health, headed by Barbara Given R.N., Ph.D. F.A.A.N. To my friends along the way particularly my intellectual friend Josh Woods, his ability to understand has help me through many critiques in my thesis and in life. I have to thank my wonderful Brennan family, particularly my father, mother and sister who have always been supportive and loving in my life. Last but not least, I want to thank my best friend and fiancé Kate Myers who I love above all, her support over the past year and half made it possible for this thesis to be complete.

TABLE OF CONTENTS

INTRODUCTION.....	1
MEHTODS.....	5
RESULTS.....	13
APPLICATION.....	25
DISCUSSION.....	30
CONCLUSION.....	35
APPENDIX.....	36
BIBLIOGRAPHY.....	44

INTRODUCTION

The clinical breast examination (CBE) is a recommended approach for the early detection of breast cancer, and is considered to be a valuable complement to mammography^{1,2}. There is general agreement that proper teaching of CBE techniques is of fundamental importance. When teaching the CBE, the sensitivity and specificity of lump detection is often assessed using six silicone breast models with a known number and location of lumps. The current method³ of calculating the specificity of lump detection is inaccurate due to the use of the entire silicone model as the unit of analysis. Using this method, a learner who identifies multiple false positive lumps in one of the models has the same calculated specificity as another learner who identifies one false positive lump in the same silicone model. Similarly, this method is insensitive to learner progression of skill acquisition. These anomalies make it difficult to provide effective feedback to learners with varying skill levels.

The teaching of CBE skills conventionally use a combination of didactic lessons, standardized patient instructors, and the silicone breast models. Although no official standard of teaching or performance of CBE exists, the MammaCare[®] method³ is currently the most widely implemented, published, and studied tool. This tool is the only teaching method capable of independently and directly evaluating a learner's lump discrimination, using the silicone breast model examination³⁻⁸. Effective use of this teaching technique requires proper evaluation and feedback of a learner's performance.

The MammaCare® method uses standardized patients to teach visual inspection and some CBE palpation skills. A series of six silicone breast models are used to reinforce the detection of breast abnormalities during CBE skills training. These six silicone breast models contain a known number, size, location, and hardness of lumps and were developed by the MammaCare® Corporation.^{3-5,9} One silicone model in each set of six contains no lumps; each of the other five silicone models contains between one and five lumps. Among the six standardized silicone models a total number of eighteen lumps are present. The MammaCare® series is the only validated tool to assess learners' lump detection and discrimination skills^{1,3}.

Traditionally, the two components of learner feedback for CBE lump discrimination are sensitivity and specificity^{3,5,6}. In terms of lump discrimination, the sensitivity is the probability of detecting a lump given that a lump exists, and the specificity is the probability of not detecting a lump given that no lump exists. This allows the evaluator to describe a learner's skill in correctly identifying lumps (true positives), correctly reporting the absence of lumps (true negatives), detecting a lump when no lump exists (false positives), and not detecting a lump when a lump exists (false negatives).

Currently, a learner's performance is evaluated based upon the following definitions for sensitivity and specificity⁵:

$$\text{Sensitivity} = \frac{\text{The number of correctly identified lumps}}{\text{The total number of lumps in all breasts}}$$

$$\text{Specificity} = 1 - \frac{\text{The number of breast models with at least one false positive}}{\text{The total number of breast models}}$$

This definition of sensitivity is directly derived from the epidemiological definition, but the method of calculating specificity, although practical, is not.

Calculating specificity using this formula creates a dilemma, which results in difficulty discerning varying levels of learners' skill. The dilemma can be portrayed in the following examples. Two students, A and B, each examine the standard set of six silicone breast models for the presence of lumps. Student B identifies three false positives in each of the six silicone breast models (total of eighteen false positive lumps). Student A identifies only one false positive in each of six silicone breast models (total of six false positive lumps). These two students have different skills, yet using the above definition the measure of their specificity of lump detection would be identical. A similar problem occurs when measuring the lump detection skills of an individual tested pre and post CBE training. For example, before training an individual detects seven false positives per each of the six silicone models. After being trained in the proper examination techniques, the same individual improves his or her lump discrimination to find only one false positive per each of the six silicone models. Under the current method of analysis, the individual's specificity of lump detection will be zero, and unchanged, for both pre and post training evaluations, even though there was a marked improvement in discrimination skills.

In the past studies evaluating specificity of learners' CBE skills have reported some odd results. Studies assessing the specificity of the CBE in

nurses, medical students, residents, and house officers have all demonstrated declines in specificity soon after going through CBE training using the MammaCare® method^{10, 11, 12, 13}. Other studies assessing sensitivity using the MammaCare® method have chosen either not to report specificity^{14, 15} or reported other proxy variables such as number of false positives^{16, 17}. These study results could be attributed to the method of assessing CBE specificity.

The dilemmas of this method of calculating specificity are due to the use of a large unit of analysis, i.e. the entire breast model. In this paper a computer simulation is presented that demonstrates how the frequency of lump assignment and units of analysis affect the calculation of specificity.

METHODS

A computer simulation of the examination of six silicone breast models was developed to mimic the MammaCare[®] models. Computed simulations of four hundred and forty eight virtual examiners each examining six silicone models were performed. The simulations implemented varying frequencies of lump assignment, and different units of analysis; i.e. whole unit, quadrants, or thirty-six cells. Sensitivity and specificity were calculated for each of the frequencies of lump assignment and each unit of analysis. The simulation demonstrated the dependence of the calculated specificity upon the unit of analysis applied.

Computer simulation of six silicone breast models

Computer replicas of the six silicone models (MammaCare, Inc.®) used in training settings were created using a two dimensional grid in Microsoft Excel. Each of the computer simulated breasts consisted of thirty-six uniform cells allowing for three units of analysis: the whole breast, quadrant, and individual cells. A cell that is shaded designates the location of a lump in the computer-simulated breast, see appendix figure one.

The MammaCare[®] breast models used in research and evaluation consist of six silicone models containing a total of eighteen lumps of various size, depth, degrees of hardness, and location. One model in each set of six contains no lumps; each of the other five silicone models contains between one and five

lumps. Thus there are eighteen shaded areas in the computer simulation as illustrated in figure one.

Computer simulation of lump discrimination

To simulate a learner attempting to discern the location of lumps in a breast, a logic statement using a Microsoft Excel formula was inserted into each of the thirty-six cells that comprised the computer simulated silicone breast. The final outcome of the statement was the assignment of either a "1" or a "0" to each cell. A "1" meant that the virtual examiner detected a lump, while a "0" indicated that the virtual examiner detected no lump.

The unique property of the implemented formula was that while it allowed for the control of the frequency of lump assignment, the location of lump assignment was completely random. To illustrate, in a breast consisting of thirty-six cells, a researcher could set the frequency of lump assignment to be ten percent. This would mean that the average expected number of lumps to be detected by the virtual examiner in this breast would be thirty six multiplied by zero point one (0.1), which equals three point six (3.6). Similar to an actual clinician the virtual examiner either detects a lump or does not detect a lump. This means that in a thirty-six cell breast at a frequency of lump assignment of ten percent, the number of lumps assigned by the virtual examiner would typically range between three and four. Now, even though the fact that frequency of lump assignment dictates that three to four lumps on average will be

assigned in each breast comprised of thirty-six cells, the location in which of thirty-six cells the lump assignment occurs is completely random.

The following is the logic statement used in the Microsoft Excel spreadsheet.

Statement: If (RAND() >= RATE, 1,0)

Where:

If(): the "if/then" statement

RAND(): Random number, a value randomly assigned to a cell by the computer ranging from 0 to 1.

RATE: a number entered by the user of the simulation, ranging from 0 to 1, that when implemented into the Excel formula determines the probability with which a "1" or a "0" is assigned to a cell. Therefore, this value dictates the frequency of lump assignment by the virtual examiner.

1: one of two possible output values given by the statement and reported in a cell, infers the virtual examiner detected a lump.

0: one of two possible output values given by the statement and reported in a cell, infers the virtual examiner detected no lump.

With this statement, the computer assigned a random number and the researcher assigned a value for the probability of a "1" or "0" occurring in a single cell. If the random number generated by the computer for a cell equaled or exceeded the value assigned as the rate, then the cell was assigned a "1". The "1" was then interpreted as the virtual examiner detecting a lump. Alternatively, if

the random number was less than the value assigned as the rate, then the cell was assigned a “0”. The “0” was then interpreted as the virtual examiner not detecting a lump.

It is important to clarify that this statement was applied uniquely in each cell of each breast that comprised the entire simulation. Given that a random number of a known range is generated in each cell and that a predetermined number (rate) was assigned to each cell to compare to the random number; then when the sample is large enough, as in our simulation, a predictable frequency of lump assignment was created. To simulate varying levels of examiner skills, multiple simulations were performed. The simulations were performed with values of the rate so that the frequency of lump assignment was 0, 10, 20, 50, and 100 percent.

Procedures and definitions for calculating specificity and sensitivity

The results of the virtual CBE using the current method of calculating the specificity of lump detection⁵ (using the whole simulated breast as the unit of analysis), was contrasted and compared to two alternative methods of calculating specificity. Alternative one calculated specificity using the quadrants of the computer simulated breast as the unit of analysis. For alternative two, the cell was applied as the unit of analysis.

The calculations used to derive specificity based on the results of the computer simulated CBE are as follows:

Current or Whole Breast Method:

$$\text{Specificity} = \frac{\text{The number of breasts with areas void of lumps identified as having no lumps}}{\text{The total number of breasts}}$$

Alternative 1, quadrant method:

$$\text{Specificity} = \frac{\text{The number of quadrants with areas void of lumps identified as having no lumps}}{\text{The total number of quadrants}}$$

Alternative 2, cell method:

$$\text{Specificity} = \frac{\text{Each cell void of a lump correctly identified as having no lump}}{\text{The total number of cells not containing lumps}}$$

Regardless of the method used to calculate specificity, the calculation of sensitivity stayed constant. This was due to the fact that the current calculation of sensitivity used in the MammaCare[®] method is the correct epidemiologic formula. The following formula was used to calculate sensitivity based on the computer simulated virtual examiners' CBE results.

$$\text{Sensitivity} = \frac{\text{The number of correctly identified lumps}}{\text{The total number of lumps in all breasts}}$$

The operational definitions of true positive, true negative, false positive, and false negative for each method, or unit of analysis, in the computer simulation are summarized in table one (in the appendix). As the definitions in the table demonstrate, the value of the true positives and false negatives remain constant across all three units of analysis. Hence, the calculation of sensitivity is unchanged throughout the computer simulation regardless of which method of deriving specificity is implemented.

While the location of lump detection was random in the computer simulation, the same simulation results were the basis of calculating the specificity across all three methods, or units of analysis. That is, regardless of the method of analysis used to calculate specificity, the computer simulated CBE virtual examiner's frequency of lump assignment results stayed constant as the method of deriving specificity was changed and contrasted. This was to assure that the only variable that was altered when comparing the calculated specificity of the simulated CBE at a given frequency of lump assignment was the unit of analysis.

Sensitivity and specificity have a dependent relationship with the frequency of lump assignment in the computer simulation because the rate of lump discrimination is fixed, while the placement to areas containing and not containing lumps is completely random. The location of lump assignment is completely random, therefore making the simulation of a virtual examiner completely impartial with regard to lump discrimination skills. Since the virtual examiner only discovers an actual lump by random chance, over an extremely large sample the actual sensitivity will become the frequency of lump assignment by the virtual examiner. Similarly, if the location of lump assignment by the virtual examiner is random, then the assignment of no lump must also be random. Given that the frequency of not detecting lumps is one minus the frequency of lump assignment, and the location of lumps is arbitrary, then over a large sample size the specificity must equal one minus the frequency of the lump assignment by the virtual examiner.

The complete haphazard lump placement by virtual examiner is reflected in cellular method's analysis of the computer simulation's relationship between sensitivity, specificity, and frequency of lump assignment. Usually sensitivity and specificity are mutually exclusive characteristics of a test. The computer simulation was designed so when the cellular method is implemented the results could be predicted and demonstrate the sensitivity's and specificity's dependence on the frequency of lump assignment. This is because the virtual examiner's location of lump assignment is made completely by random chance. This forces the virtual examiners ability over numerous trails to only be as good as the frequency of lump assignment. The randomness of the placement of lumps in the simulation drives the virtual examiner's sensitivity to equal the frequency of lump assignment, and the specificity to equal one minus the frequency of lump assignment. So when the sample size was great and the entire breast is quantifiable to the smallest detectable unit as in the cellular method of the computer simulation, the sensitivity and specificity can be predicted based on the frequency of lump assignment.

Once the computer simulation was developed it became quite simple to adjust the frequency of lump assignment for each simulation and calculate the specificity for each unit of analysis. After the simulated CBEs were performed the results were summarized into two by two tables. The summarized comprehensive results in the two by two tables for each method were then used to calculate sensitivity and specificity. To calculate specificity and sensitivity for each unit of analysis at varying frequencies of lump assignment the program

simulated four hundred and forty eight virtual examiners each of whom examined the six two dimensional computer models designed to emulate the MammaCare[®] breasts. With such a large number of virtual examiners each inspecting six breasts apiece, the simulated CBE for each frequency of lump assignment resulted a total virtual examination of two thousand six hundred and eighty eight breasts, ten thousand seven hundred and fifty two quadrants, and eighty eight thousand seven hundred and four cells. This was to assure that the results had extremely narrow confidence intervals for each method used to calculate specificity.

RESULTS

The simulation of four hundred and forty eight physicians each examining six silicone models demonstrated the inherent failures of the current and quadrant methods to accurately estimate specificity. The current method of calculating specificity, using the whole breast as the unit of analysis, grossly underestimates specificity compared to the cellular method. The results demonstrate the progressive diminishing of specificity that occurs as the unit of analysis increases in area. Also, when the frequencies of false positives were high, as with unskilled examiners, the ability of the current and quadrant methods to discern varying levels of inter and intra examiner skills were diminished or even absent. Simulations that implemented the thirty six cell method resulted in pin point estimates and confidence intervals of specificity providing practically perfect accuracy and precision.

Figure Two

Figure two displays the results of an example of the computer simulated CBE of one breast. The frequency of lump assignment in this example was ten percent. Since there are thirty six cells that comprise the breast, the expected number of lumps to be detected by the virtual examiner should be thirty six multiplied by zero point one (0.1). So for this thirty six celled breast the expected number of lumps to be detected by the virtual examiner would be three point six (3.6). Just as in practice, the virtual examiner cannot find zero point six (0.6) of a lump. So typically at a frequency of lump detection of ten percent the computer

simulation detects three to four lumps per breast. In the example displayed in figure two during the simulated CBE the virtual examiner detected four lumps, signified by the cells that contain a "1".

Though the frequency of lump assignment in figure two can be predicted, the location of the lump detection by the virtual examiner is not predictable. This is due to the fact that a random number of a fixed range is generated for each individual cell, and that number is then compared with the programmed frequency of lump detection to determine if a lump is detected. So in this example, as with the entire simulation, though the virtual examiner's frequency of lump assignment in this breast is predetermined, where the lumps are recorded as being detected is unknown until after the simulation is performed. In the example displayed in figure two, the virtual breast is interpreted as actually containing five lumps. This is because a shaded cell in the virtual breast is interpreted as containing a lump.

After the simulated CBE was performed, at a frequency of lump assignment of ten percent, the virtual examiner indicates having discriminated four lumps. This is seen by the four cells which contain a "1" within them. In this example, the virtual examiner only detected one lump where a lump actually existed. This is indicated in the example as the shaded cell that contains a "1" within it, located in the lower right quadrant of the simulated breast. The other three lumps discriminated by the virtual examiner occurred in locations where no lump existed. This is indicated in the example by a "1" occurring in three of the

cells that are not shaded. In the example this can be noted in the upper left, upper right, and lower right quadrants of the virtual breast in table two.

In the example displayed in figure two the calculated sensitivity of lump detection is the same for the current, quadrant, and cellular methods. As explained and defined earlier, the sensitivity is the same regardless of the method used to determine specificity. In this example of a simulated CBE of one breast, the virtual examiner only found one of five lumps that were actually present. This means that for this example, the virtual examiner had one true positive and four false negatives. Therefore in this one breast the virtual examiner had a sensitivity of one out of five, or twenty percent.

The results of specificity in this example are more complex because the specificity is defined differently for each method of analysis that is implemented. The definitions of specificity for each method of analysis are explained in the methods section of this paper. In the example shown in figure two, the specificity calculated using the current, quadrant, and cellular methods are zero, twenty-five, and ninety percent, respectively. As displayed by the results, the method of analysis plays a large role in the calculated specificity.

Implementing the current method in this example calculates a specificity of zero percent. Since the example demonstrates the analysis of only one breast, the denominator of the specificity is one. In this method the breast would have to be completely devoid of false positives in order for it to be deemed as a true negative. As seen in figure two, the virtual examiner detects three lumps where no lump exists, indicated by the "1"s located in the unshaded cells. So the true

negatives in this example are zero. Therefore in this example using the whole breast method, the calculated specificity of the virtual examiner is zero divided by one, or zero percent specific.

Implementing the quadrant method in this example calculates a specificity of twenty five percent. Clinically and in this simulation, a single breast is divided into four quadrants. So in this example using only one breast the resulting denominator for calculating specificity is four. In the quadrant method a quadrant has to be completely devoid of any false positives for it to be deemed a true negative. In the example displayed in figure two, a true negative defined by the quadrant method occurs once in the lower left quadrant. The other three quadrants in this example each contain a false positive. Therefore in this example using the quadrant method, the calculated specificity of the virtual examiner is one divided by four, or twenty five percent specific.

Implementing the cellular method in this example calculates a specificity of ninety percent. All unshaded cells in this breast interpreted as not containing a lump contribute to the denominator. In the cellular method an unshaded cell containing a "0" indicates that the virtual examiner identified a true negative. An unshaded cell with a "1" is interpreted as the virtual examiner having detected a false positive. In the example shown in figure two the virtual examiner identified twenty-eight true negatives and three false positives. Therefore in this example using the cellular method, the calculated specificity of the virtual examiner is twenty-eight divided by thirty one, or ninety percent specific.

Given that the frequency of lump assignment of the virtual examiner was set at ten percent it can be predicted that the sensitivity of lump detection in the series of six models should be ten percent and the specificity of the absence of lumps should be ninety percent. The resulting sensitivity in figure two is twenty percent. A little higher than expected, but this is a sample of only one breast. When the simulation of the virtual CBE is run across thousands of breasts the sensitivity approaches the predicted value for this frequency of lump assignment. Similarly, the specificity using the cellular method attains the predicted value.

In figure two the whole breast and quadrant methods grossly underestimate the predicted specificity. By definition the whole breast and quadrant methods increase their probability of a false positive occurring, and as the example displays, these methods inherently diminish and incorrectly report the specificity.

Figure Three

Figure three again demonstrates a simulated CBE performed by a virtual examiner. The location and number of lumps in the virtual breast, and the methods of calculating sensitivity and specificity in figure three are the same as in figure two. The difference is that the example displayed in figure three has a frequency of lump assignment set at fifty percent. In figure two the frequency of lump assignment was set at ten percent. This increase in the frequency of lump assignment in figure three produces a noticeable enhancement of the number of "1"s recorded in the cells of the figure.

The values of the calculated sensitivity and specificity have now changed since the virtual examiner has become more hastened in detecting lumps. In figure three as in figure two, the calculated sensitivity is the same regardless of the unit of analysis, unlike specificity where the calculation is dependant upon the unit of analysis that is implemented.

In figure three during the CBE simulation the virtual examiner correctly detected three out of a possible five lumps. This is indicated in figure three as the three out of five shaded squares in which a "1" was marked. In this example the virtual examiner has detected three true positives and two false negatives. This means that in figure three for this one breast the virtual examiner had a sensitivity of three divided by five, or sixty percent.

In figure three as in figure two, the value derived for specificity is dependant on the unit of analysis. In figure three the simulated CBE produced a virtual examiner with a frequency of lump assignment of fifty percent. In this example the specificity for the current, quadrant, and cellular methods was zero, zero, and fifty one percent, respectively.

In figure three the denominator for calculating specificity using the whole breast method is one, since only one breast was examined in this example. Again, in this method the breast would have to be completely devoid of false positives in order for it to be deemed a true negative. With such a high frequency of lump assignment the probability of that occurring is extremely low. As seen in figure three, the virtual examiner discriminated fifteen lumps where no lump existed, indicated by the "1"s located in the unshaded cells. So the true

negatives for whole breast method in this example are zero. Therefore in this example using the whole breast method, the calculated specificity of the virtual examiner is zero divided by one, or zero percent specific.

The quadrant method in this figure, similar to the whole breast method, also derives a specificity of zero percent. The denominator when calculating the specificity for a single breast using the quadrant method is four. In the quadrant method, a quadrant has to be completely devoid of any false positives for it to be deemed a true negative. Similar to the whole breast method, when the virtual examiner possesses such a high frequency of lump discrimination the probability of that occurring is extremely low. By the operational definition for the quadrant method, no true negatives are detected by the virtual examiner in this CBE simulation. Therefore in this example using the quadrant method, the calculated specificity of the virtual examiner is zero divided by four, or zero percent specific.

In contrast, figure three demonstrates for this example that the cellular method calculates a specificity of fifty one percent. All unshaded cells in this breast contribute to the denominator. In the cellular method an unshaded cell containing a "0" indicates that the virtual examiner identified a true negative. An unshaded cell with a "1" is interpreted as the virtual examiner detecting a false positive. Figure three indicates that the virtual examiner identified sixteen true negatives and fifteen false positives. Therefore in this example using the cellular method, the calculated specificity of the virtual examiner is sixteen divided by thirty-one, or fifty-one percent specific.

Given that in this example the frequency of lump assignment of the virtual examiner was set at fifty percent, it can be predicted that sensitivity of lump detection should be fifty percent and the specificity of the absence of lumps should be fifty percent in this simulation. The resulting sensitivity in figure three is sixty percent. A little higher than expected, but this is a sample of only one breast. When the simulation of the virtual CBE is run across thousands of breasts the sensitivity approaches the predicted value for this frequency of lump assignment. Similarly, the specificity using the cellular method almost attains the predicted value.

The whole breast and quadrant methods in figure three both calculate specificities of zero percent. In these methods the operational definition of true negative are difficult to attain since the unit of analysis is large and frequency of lump assignment is often. The results of the high frequency of lump discernment and the large unit of analysis create a scenario where no true negatives are demonstrated by this virtual examiner. With no true negatives attained in either method the specificity is zero. The example displays that these methods inherently diminish and incorrectly report the specificity.

Comparing Figure Two and Three

When comparing figure two to figure three the difficulty of discriminating varying levels of examiner skill becomes more understandable. In these two figures the frequency of lump assignment by the virtual examiner changes forty

percent. The current, i.e. whole breast method, and quadrant methods are unable to report this change in lump detection frequency. In contrast, the cellular method calculates specificity so that this change in frequency of lump assignment by the virtual CBE examiner is accurately reported.

When deriving specificity using the current method (whole breast as the unit of analysis), the values displayed in figure two and figure three demonstrate how this method is clearly erroneous. In these two examples the frequency of lump discernment by the virtual examiner has changed dramatically. The frequency of lump detection has changed forty percent, which noticeably affects the rate of false positives indicated by the virtual examiner. Both of the simulation results using the current method in these two examples calculate a specificity of zero percent. As explained earlier this is tremendously incorrect compared to the known change in performance. Interestingly, when comparing the two virtual examinations, the whole breast method reports a specificity that is not helpful in distinguishing different levels of examiner skill. The examples demonstrate how implementing the whole breast method of analyzing specificity fails to produce results where intra and inter examiner skill progression can be observed. These results were not just witnessed in these examples but were ubiquitous in the entire simulation.

Similar to the current method, the results of specificity using the quadrant method when figure two and figure three are compared are still poor. The quadrant method is more accurate at low levels of lump detection than the current method. In figure two where the frequency of lump detection is low, the

quadrant method reports a more accurate specificity relative to the current method. However, this derived specificity is still extremely invalid. When comparing varying levels of the virtual examiner's frequency of lump assignment, the quadrant method can report a difference. Although the difference of examiner specificity reported in these examples does not properly reflect the actual change in the virtual examiner's frequency of lump assignment. The actual difference in lump discernment between figure two and three is forty percent, but in this example the quadrant method only calculates a difference of twenty five percent. Like the current method, the quadrant method is invalid and fails to properly report differences in examiner's lump discrimination skills.

The cellular method of deriving specificity actually reflects the change in the virtual examiner's frequency of lump detection. Regardless if the frequency of lump detection was often or rare, the cellular method attains an accurate specificity. When comparing varying frequencies of lump detection among virtual examiners, the cellular method is far superior to either of the other two methods. It provides accurate and precise information that reports valid learner feedback. Also being so accurate and precise, the inter and intra examiner skills' can be clearly reported and correctly assessed. The results comparing these two examples obviously demonstrates the superiority of the cellular method, and makes obvious its necessity in reporting specificity.

Overall results of the simulation

The definitions to calculate the specificity of lump detection for each unit of analysis were applied to varying frequencies of lump detection. The simulations of virtual CBEs were run so that the virtual examination of two thousand six hundred and eighty eight breasts, ten thousand seven hundred and fifty two quadrants, and eighty eight thousand seven hundred and four cells were performed. For each unit of analysis, these numbers of examinations yielded extremely narrow 95% confidence intervals. As displayed in figure two, the confidence interval of the calculated specificity using this magnitude of simulations for the whole breast, quadrant, and cellular method were $\pm 5 \times 10^{-3}$, $\pm 1 \times 10^{-5}$, and $\pm 1.25 \times 10^{-6}$, respectively.

Table two also summarizes the relationship between the frequency of lump assignment, unit of analysis, and specificity. The table demonstrates that the whole breast unit of analysis is capable of discriminating varying levels of specificity only at low frequencies of lump assignment, while the quadrant method is moderate, and the cellular method is the best at differentiating skill levels.

Not only is the cellular method of calculating specificity accurate, it is excellent at differentiating varying levels of frequency of lump discrimination. The predicted specificity can be determined as one minus the frequency of lump assignment. As table two indicates the cellular method always reports the predicted specificity. The results of the entire simulation further strengthen the notion that the current method of calculating specificity is incorrect and that the

cellular method could be an accurate and precise alternative to evaluating learner specificity using the MammaCare® models.

APPLICATION

An application of calculating specificity using the current (whole breast method) and cellular methods was applied to a small sample of data. The data came from a study sponsored by Michigan State University Institute for Health Care Studies with support by the Michigan Cancer Consortium Division of the Michigan Department of Community Health, headed by Barbara Given R.N., Ph.D. F.A.A.N. The data collected from a study assessing nurse practitioner's (NP) CBE skills using the standardized silicone models produced by MammaCare[®]. A small sample of data collected five years ago and without any personal identifiers of ten NPs was extracted from this database. These ten individual's results of the silicone model CBE were reassessed to see how the calculated specificity would be affected when the current and cellular methods were applied.

The data obtained was from ten NPs who each had their CBE skills evaluated using the standard set of six MammaCare[®] silicone breast models. The models were of the same design as those used in previously validated studies and described earlier in the paper.

The learners were asked to evaluate each silicone model and report lump location, size, hardness, and depth. After evaluating each model, the learner would translate their assessment to a form and indicate the location of the lump and describe its characteristics of size, depth, and hardness. The sensitivity and specificity of lump location in these NPs were assessed using the current (whole breast) and cellular methods. The calculations for sensitivity and specificity using

current and cellular methods of analysis are described previously in the methods section.

The application of the whole breast and cellular methods to derive sensitivity and specificity among these ten learners are displayed in table three. Similar to the results of the computer simulation, the application of the cellular method compared to the whole breast method displays that learner evaluations with regard to specificity change drastically. The application further demonstrates the need to implement the cellular method into learner assessment, particularly in the case of reporting specificity.

As table three demonstrates, the sensitivity for both methods have similar results. This is expected since the definitions to assess sensitivity are the same for both methods. Differences in the calculated sensitivity occurred because when using the cellular analysis the actual lump was not always located in the center of the cell. This tended to skew the assessment for some learners. This problem could be easily rectified by having the MammaCare[®] Corporation place the actual lumps in the center of cells decided as having lumps. However, the evaluation of sensitivity using the two different methods was very similar.

Table three demonstrates that the reported learner's specificity changed drastically when the unit of analysis switched from the whole breast to the cellular method. The reasons for this large change in the reported specificity are due to the increased precision and accuracy of the cellular method. The cellular method attains better precision and accuracy by more completely quantifying the areas that do not contain lumps. By better quantifying this region of the silicone

models, the denominator becomes larger which allows more precision. Also, the actual differences in learner performance could be more clearly discerned which improved accuracy.

The specificity in table three using the whole breast method ranges from zero to sixty six percent, while using the cellular method ranges from ninety two point four to ninety eight point nine percent. This inflation of the value and decrease in range of the specificity in the cellular method compared to the whole breast method is largely due to the cellular method's larger denominator. In the whole breast method the denominator is six, since each student examined six silicone models. In the cellular method the denominator is one hundred and ninety eight. The cellular denominator is equal to six breasts times the thirty six cells that comprise them, minus the eighteen cells that contain lumps. This increases the precision of the specificity since the results in the cellular method are reported in a scale of one one-hundred and ninety eighth, versus the whole breast method which reports information in terms of one sixth.

When looking at table three the improved accuracy of the specificity using the cellular method is noticed when the proficiency of learners are compared by both methods. In table three, comparing the specificity of learner seven to learner nine using the whole breast method, reports a marked difference in favor of learner seven. Paradoxically, when comparing the specificity of the same learners using the cellular method, learner nine is evaluated much better than learner seven. The truth is that with regards to reporting false positive in these models, learner nine reported far fewer than learner seven. It happens that

learner nine marked a few false positives over six models, whereas learner seven marked several false positives within two models.

The whole breast method biases results to favor learners who only mark errors in one breast, regardless of how many errors occur in that breast.

Similarly, the whole breast method biases results to penalize learners severely for making only one error in multiple breasts. So in actuality, learner nine has a better specificity than learner seven. Unfortunately, when the current method assesses learner performance, both the learner and evaluator are misled to believe that learner nine had very poor specificity. The truth is that learner nine's specificity of lump discrimination was good.

The inadequacy of the whole breast method can be further scrutinized using learner seven. Using the whole breast method, learner seven's specificity is reported as being in a three way tie for the best in this group of ten. In actuality using the cellular method, learner seven's ranking drops to become sixth out of ten. This is a substantial change in rank for learner seven when the cellular method is implemented. In this group of learners, the whole breast method misleads the learner and evaluator with regard to lump discrimination proficiency.

Overall the results of the application demonstrate that when the MammaCare[®] method is being used, the cellular method should be implemented in evaluating learners' CBE lump discrimination skills. The cellular method in the application demonstrates greater precision and accuracy in comparison to the current method. Most importantly the cellular method reports specificity in a way

that is exact and truthful so as to not mislead the learner, evaluator, or studies involving the MammaCare[®] method.

DISCUSSION

The simulation and application demonstrate that using the cellular unit of analysis best discriminates examiners of different skill levels and best documents learner progression or setbacks. The computer simulation and application of in silicone models demonstrated that specificity is more accurate when the units of analysis are smaller. Furthermore, when larger units of analysis are used and the false positive rate is high, the reported value of specificity is unreliable, underestimated, and unable to differentiate multiple learners' skill and improvement. In the cellular method, where the unit of analysis can contain only one false positive, the specificity is accurately calculated.

Deficiencies of the computer simulation and application of the cellular method

The computer model and application both have limitations. First, these studies apply only to teaching settings, where the location and number of lumps are known. In particular, these studies were done to display the shortcomings of the current method of assessing specificity using the MammaCare[®] silicone models. Therefore there is little direct function of these results except in learning settings using the MammaCare[®] models. Secondly, the simulation and practical application cannot duplicate the subtle vagaries of interaction among the physician, patient, and environment that influence specificity of lump detection in a real clinical breast examination.

Deficiencies of the computer simulation

Certain limitations only pertain to the computer simulation. First the computer model is very simple. The computer model neglects variables that influence CBE including visual inspection, area of tissue examined, types of pressure, types of motion, part of finger or hand used, number of fingers used, search pattern, and duration of search¹. Another limitation is that the computer simulation assumes that all lumps have an equal probability of being discovered. This neglects lump characteristics such as varying depth, size, and hardness of lumps, which have been shown to influence an examiner's ability to perceive lumps¹. The translation of the simulation to include these variables requires additional work.

Deficiencies of the application of the cellular method.

The application of the cellular method to current MammaCare[®] models in this paper has two flaws that need to be rectified for future applications of the cellular method. One problem with the application was that the location of actual lumps was not centered within a single cell. The other problem was that the area of silicone model is less than the area of the cellular grid used to evaluate it.

The MammaCare[®] silicone models' lump locations were not designed with consideration for a cellular method of analysis. The MammaCare[®] Corporation can place lumps according to requested locations. The requested location of future lumps in MammaCare[®] silicone models should occur so that they are situated in the center of a cell used for analysis. This would benefit the learner

and evaluator in grading the sensitivity and specificity of a learner's lump detection. If the actual lumps are in the center of the cell this increases the probability of proper evaluation.

The second problem with application of the cellular method is that the actual silicone models evaluated by the ten learners are smaller in area than the cellular grid used to grade them. When the area of the cells overlay the area of the silicone model, as in figure four, each corner of the grid extends the area of the silicone model. Mathematically, the summation of the overextending areas, in figure five, needs to be withdrawn from the denominator of specificity used to analyze the silicone models. Preferably, the overextending area should be described in terms of cellular area so that it can be directly subtracted from the denominator used to calculate specificity. In the cellular form of evaluation used in the application portion of this paper, the denominator of the specificity was one hundred and ninety eight cells. A more precise dominator would be one hundred and ninety eight minus six times the overextending area of one grid on one model. This would enable the cellular method of analysis to become even more precise in quantification of areas that do not contain lumps.

Strengths of the computer simulation and application of the cellular method

The computer simulation contains several strengths. The computer model evaluated a large number of virtual examiners providing tight confidence intervals for the sensitivity and specificity of each unit of analysis. Furthermore, the computer simulation tested the interaction of skill level (measured as frequency

of lump application) and the units of analysis effect on sensitivity and specificity. This simulation obviously demonstrates the failure of the current method to report specificity in a manner that is reliable and reflects a learner's skill level. Finally, this simulation provides a conceptual framework that can be applied to an existing and widely used teaching tool.

Applying the cellular method to actual data further demonstrated the strengths of this method of analysis. As reported earlier, the cellular method is more accurate and precise than the current (whole breast) method. When applied, the cellular method reports learner feedback in a manner that is helpful and reliable. This precise and accurate feedback will hopefully create evaluator and learner feedback that enhances CBE skills.

Validity of the cellular method and biases of the current method

The validity of the cellular method is clearly demonstrated in the computer simulation by the relationship between sensitivity, specificity, and their dependence upon the frequency of lump assignment. Since the assigned value for the frequency of lump detection was previously determined, the true sensitivity and specificity could be predicted. For example in table two, when the frequency of lump assignment is set at ten percent, the predicted sensitivity and specificity of lump detection are ten percent and ninety percent respectively. Each setting of the frequency of lump assignment has a unique and calculable sensitivity and specificity, independent of the unit of analysis. However as is displayed in table two, in the computer simulation only the cellular unit of analysis

consistently approached the predicted value for specificity. The other two methods, using the whole breast and quadrant, underestimate the specificity because the unit of analysis is larger than the smallest detectable element. Therefore this large unit of analysis increased the probability of collecting false positives in a given area, and even allowed the containment of multiple unmeasured false positives.

The classic epidemiologic definition of specificity is the foundation for the cellular model's accuracy. Models where multiple false positives can occur within a unit of analysis will increase the possibility of a false positive occurring and underestimate specificity. This underestimating bias of specificity that occurs with larger units of analysis should be further examined. This form of bias should be properly defined, and searched for in other medical tests, screening modalities and procedures, such as colonoscopy¹⁸ or prostate screening.

Ideally there is a point, somewhere between the whole breast as the unit of analysis and the smallest quantifiable unit, that balances pragmatism and rigor to more accurately evaluate learners' CBE skills. It has been recommended that when teaching the CBE with the MammaCare[®] models that a learner use eight or nine vertical strips to examine the breast model¹⁹. This would lend it self aptly to creating an eight by eight cellular method of analysis to evaluate sensitivity and specificity. The optimal unit of analysis must be one that minimizes the capacity to contain multiple false positives, is practical, and can be easily applied to currently available teaching tools.

CONCLUSION

The computer simulation and application of the cellular method to actual data demonstrate that the current method of analyzing learners' specificity of CBE skills, implementing the whole breast as the unit of analysis, is insufficient. Furthermore, the simulation demonstrates that implementing quadrants as the unit of analysis is also insufficient to accurately estimate specificity. The methods of calculating specificity using whole breast and quadrants as the unit of analysis are severely limited in their capacity to compare inter and intra examiner improvement regarding specificity.

The cellular method used in this computer simulation and application provides near perfect accuracy. Based on the computer simulation and the application of the cellular method, when using the standardized MammaCare[®] silicone breast models to teach CBE, educators should use a cellular approach to calculate specificity. The cellular approach more accurately implements the epidemiologic definition of specificity, and is more beneficial to society, the patient, and the clinician by providing better CBE evaluation. Hopefully better CBE evaluation will translate into more standardized and effective CBE skills in doctors, physician assistants, nurses, and nurse practitioners, and may even lead to better breast cancer diagnosis and treatment.

APPENDIX

The Computer Simulated Breast, Lumps, and Units of

Method Analysis

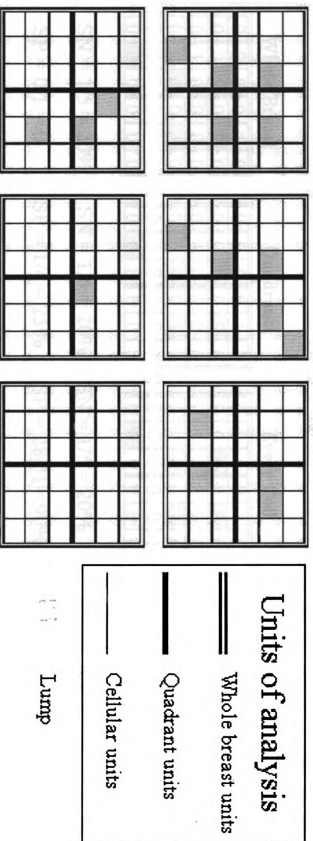


Figure 1: Two-dimensional template simulating the six standardized breast models (Mammacare Inc ©). Shaded areas represent location of lumps.

The Calculated Specificity Using Different Methods of Analysis

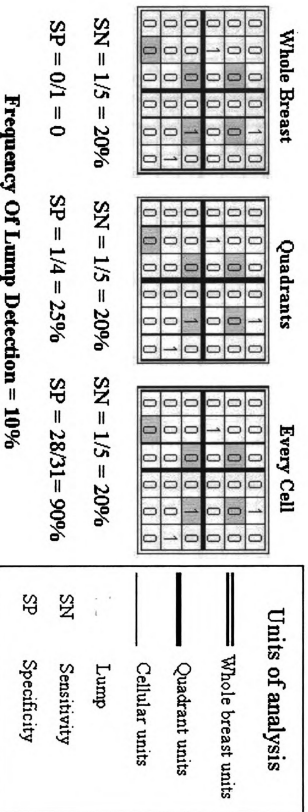


Figure 2: Computer output of a trainee's virtual examination of a single breast, with a 10% Frequency of lump detection. The effect that the three different units of analysis have on the calculation of specificity is demonstrated.

The Calculated Specificity Using Different Methods of Analysis

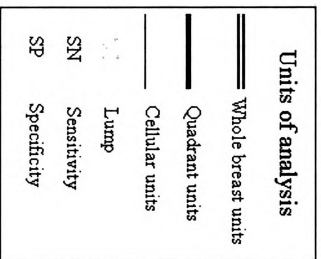
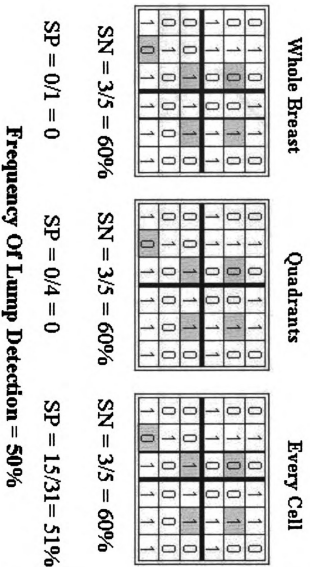


Figure 3: Computer output of a trainee's virtual examination of a single breast, with a different frequency of lump detection. The effect that the three different units of analysis have on the calculation of specificity is demonstrated.

The Area of the Silicone Model with an Overlaying Analysis Grid

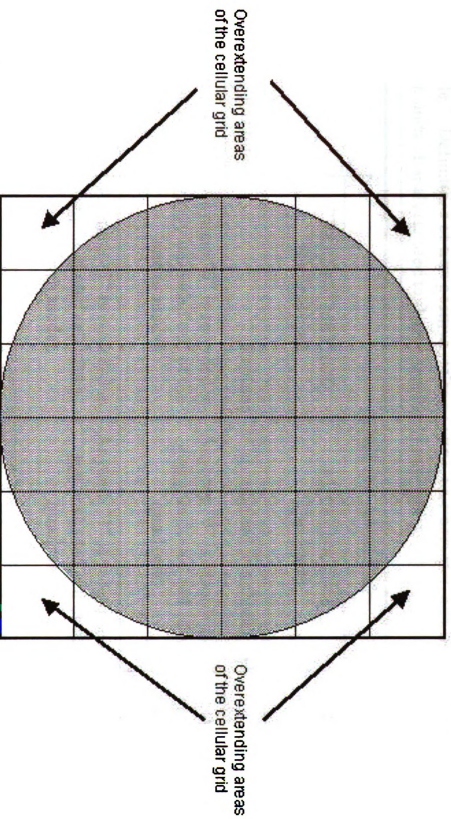


Figure 5 displays the area of the cellular grid used to analyze the silicone model overlaying the area of an actual silicon model. The figure demonstrates in each corner the overextending regions of the cellular grid.

Table 1 : Definitions of true positive, false positive, true negative, and false negative for the three units of analysis used to derive sensitivity and specificity in the computer simulated CBE.

		UNIT OF ANALYSIS		
		CURRENT METHOD Whole Breast	ALTERNATIVE 1 Quadrants	ALTERNATIVE 2 Cells
True Positive		Anytime "1" is recorded in a shaded cell		
True Negative		Anytime "0" is recorded in <u>every</u> unshaded cell of the entire model.	Anytime "0" is recorded in <u>every</u> unshaded cell of a quadrant.	Anytime "0" is recorded in an unshaded cell.
False Positive		Anytime "1" is recorded in <u>any</u> unshaded cell of the entire model.	Anytime "1" is recorded in <u>any</u> unshaded cell of a quadrant.	Anytime "1" is recorded in an unshaded cell.
False Negative		Anytime "0" is recorded in a shaded cell		

Table 2: The relationship between the frequency of lump detection and specificity depending on the unit of analysis.

Frequency of lump detection	Specificity*		
	Whole breast†	Quadrants*	Cells*
	N=2,688	N=10,752	N=88,701
0%	1	1	1
10%	.03	.42	.90
20%	0	.16	.80
50%	0	0	.50
100%	0	0	0

* 95% Confidence interval † = $\pm 1.5 \times 10^{-3}$, * = $\pm 1.1 \times 10^{-5}$, * = $\pm 1.125 \times 10^{-6}$.

Table 3: The calculated sensitivity (Sens) and specificity(Spec) of ten learners of Clinical Breast Exam (CBE) evaluated using the whole breast (WB) and cellular(Cell) methods.

CBE learner	WB Sens (%)	Cell Sens (%)	WB Spec (%)	Cell Spec (%)
1	33.3	38.8	0.0	92.4
2	38.0	50.0	0.0	93.9
3	55.5	55.5	50.0	98.4
4	55.5	55.5	66.6	98.9
5	44.4	38.8	33.3	97.4
6	33.3	38.8	50.0	98.4
7	44.4	44.4	66.6	94.9
8	50.0	50.0	66.6	98.9
9	27.7	27.7	16.6	97.9
10	61.1	66.6	0.0	93.9

BIBLIOGRAPHY

1. The American Cancer Society. "Clinical Breast Examination: A Review of the Literature." September 19, 2002.
2. Miller AB, To T, Baines CJ, Wall C. "Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years." *J Natl Cancer Inst.* 2000 Sep 20;92(18):1490-9.
3. Fletcher SW, O'Malley MS, Bunce LA. "Physicians' abilities to detect lumps in silicone breast models." *JAMA.* 1985 Apr 19;253(15):2224-8
4. Hall DC, Adams CK, Stein GH, Stephenson HS, Goldstein MK, Pennypacker HS. "Improved detection of human breast lesions following experimental training." *Cancer.* 1980 Jul 15;46(2):408-14.
5. Fletcher SW, O'Malley MS, Pilgrim CA, Gonzalez JJ. "How do women compare with internal medicine residents in breast lump detection? A study with silicone models." *J Gen Intern Med.* 1989 Jul-Aug;4(4):277-83.
6. Trapp MA, Kottke TE, Vierkant RA, Kaur JS, Sellers TA. "The ability of trained nurses to detect lumps in a test set of silicone breast models." *Cancer.* 1999 Nov 1;86(9):1750-6.
7. Lannotti RJ, Finney LJ, Sander AA, De Leon JM. "Effect of clinical breast examination training on practitioner's perceived competence." *Cancer Detect Prev.* 2002;26(2):146-8.
8. Aliabadi-Wahle S, Ebersole M, Choe EU, Beech DJ. "Training in clinical breast examination as part of a general surgery core curriculum." *J Cancer Educ.* 2000 Spring;15(1):10-3.
9. Pennypacker HS, Bloom HS, Criswell EL, Neelakantan P, Goldstein MK, Stein GH. "Toward an effective Technology of instruction in breast self-examination." *Int J Ment Health* 1982; 11: 98-116.
10. McDermott MM, Dolan NC, Huang J, et al. "Lump detection is enhanced in silicone breast models simulating postmenopausal breast tissue." *J Gen Int Med* 1996; 11: 112-114
11. Campbell HS, McBean M, Mandin H, Byrant H. Teaching medical students how to perform a clinical breast examination. *Acad Med* 1994; 69: 993-995.

12. Pilgrim C, Lannon C, Harris RP, et al. Improving clinical breast examination training in a medical school. *J Gen Int Med* 1993; 8: 685–688.
13. Campbell SH, Fletcher SW, Lin S. Improving physicians' and nurses' clinical breast examination: a randomized controlled trial. *Am J Prev Med* 1991; 7: 1–8.
14. Smith RL, Hanchak NA, Bloom H, et al. The effectiveness of postgraduate education on the clinical breast examination skills of primary care physicians. *Am J Man Care* 1996; 2: 989–995.
15. Costanza ME, Luckmann R, Quirk ME, et al. The effectiveness of using standardized patients to improve community physician skills in mammography counseling and clinical breast exam. *Prev Med* 1999; 29: 241–248.
16. Benincasa TA, King ES, Rimer BK, et al. Results of an office-based training program in clinical breast examination for primary care physicians. *J Cancer Ed* 1996; 11: 25–31.
17. Trapp MA, Kottke TE, Vierkant RA, et al. The ability of trained nurses to detect lumps in a test set of silicone breast models. *Cancer* 1999; 86: 1750–1756.
18. Cotton PB, et al “Computed tomographic colonography (virtual colonoscopy): a multicenter comparison with standard colonoscopy for detection of colorectal neoplasia.” *JAMA*. 2004 Apr 14;291(14):1713-9.
19. Pennypacker HS, Pilgrim CA. Achieving competence in clinical breast examination. *Nurse Practitioner Forum* 1993;4:85-90.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02736 7568