



Published in final edited form as:

AJR Am J Roentgenol. 2012 June ; 198(6): W562–W567. doi:10.2214/AJR.11.7971.

Web-based Mammography Audit Feedback

BM Geller¹, L Ichikawa², DL Miglioretti^{2,3}, and D Eastman^{2,4}

¹University of Vermont, Health Promotion Research, 1 South Prospect Street, Burlington, VT 05401-3444

²Group Health Research Institute, 1730 Minor Avenue, Ste. 1600, Seattle, WA 98101

³University of Washington Department of Biostatistics, Seattle, WA

⁴Kaiser Permanente Center for Health Research Southeast, Eleven Piedmont Center, Suite 110, 3495 Piedmont Road NE, Atlanta, GA 30305

Abstract

Introduction—Interpreting screening mammography accurately is challenging and requires ongoing education to maintain and improve interpretative skills. Recognizing this, many countries with organized breast screening programs have developed audit and feedback systems based on their national performance data to help radiologists assess and improve their skills. We developed and pilot tested an interactive website to provide screening and diagnostic mammography audit feedback with comparisons to national and regional benchmarks.

Methods and Materials—Radiologists who participate in three Breast Cancer Surveillance Consortium registries in the United States were invited during 2009 and 2010 to use a website that provides tabular and graphical displays of mammography audit reports with comparisons to national and regional performance measures. We collected data on the use of and perceptions of the website.

Results—Thirty-five of 111 invited radiologists used the web site from 1–5 times in a year. The most popular measure was sensitivity for both screening and diagnostic mammography while a table with all measures was the most visited page. Of the 13 radiologist who completed the post use survey, all found it easy to use and navigate, 11 found the benchmarks useful, and 9 reported that they intend to improve a specific outcome measure this year.

Conclusions—An interactive website to provide customized mammography audit feedback reports to radiologists has the potential to be a powerful tool in improving interpretive performance. The conceptual framework of customized audit feedback reports can also be generalized to other imaging tests.

INTRODUCTION

Mammography remains the best available breast cancer screening technology, but United States (US) radiologists vary widely in their accuracy of mammographic interpretation.(1–4) Although some radiologists specialize in breast imaging, radiology generalists interpret over two thirds of mammograms in the US.(5) The effectiveness of mammography depends on the ability to perceive mammographic abnormalities and interpret these findings accurately; both tasks are quite challenging and require ongoing education to improve interpretive skills - or even maintain them. That is why many countries with organized screening programs

Corresponding author: Berta M. Geller, Ed.D., Telephone: 802-656-4187, Fax: 802-656-8826, berta.geller@uvm.edu.

This work is solely the responsibility of the authors and not the funding agency.

have developed audit and feedback systems based on their national data to help radiologists assess and improve their skills.(6, 7) The Mammography Quality Standards Act of 1992 (MQSA)(8) requires a minimal audit of US radiologists who interpret mammography; but these measures require no comparisons to benchmarks and are probably insufficient to identify deficits in the performance of mammography interpretation.(9)

The NCI-funded Breast Cancer Surveillance Consortium (BCSC) was originally started to meet the MQSA mandate to establish a breast cancer screening surveillance system.(10) Radiologists who participate in the BCSC receive paper outcome audit reports from their local regional registries. However each participating BCSC registry includes varying outcome measures and uses different formats for their reports and sends these reports either as facility or radiologist level data.(11) None of the BCSC registry outcome audits offer national comparisons.

In response to the need for audit feedback with national comparisons, we developed and pilot tested an interactive audit website with comparisons to regional and national performance benchmark data. Performance measures were displayed in graphs and tables developed based on information from a recent qualitative study with radiologists from three BCSC registries about mammography outcome audits (Bowles, geller 2009).

METHODS

Data protection

Data from the Breast Cancer Surveillance Consortium (BCSC) were used to identify eligible radiologists and calculate performance measures and benchmarks. Because of the sensitive and confidential nature of the data the website was created as a secure ASP.NET 2.0 application. This allows the website to use forms authentication with the SQL Server membership provider to create individual login credentials and track user activity. All data from the website for reporting was aggregated by year and radiologist using an anonymous BCSC study ID. Radiologists access the website via an assigned user ID (different from their BCSC study ID) and password. The user ID is not stored in the aggregated data and is known only to the BCSC Statistical Coordinating Center, the BCSC registry that sent the user ID and password, and the radiologist.

Data sources and definitions

All mammography and cancer data on the website came from the Breast Cancer Surveillance Consortium from 2000 through 2007 with a full year of cancer follow-up for the 2006 mammograms. Regional data was comprised for one group as all data from one registry and for the other two registries we combined two state adjacent registries where some radiologists read in both states. *Screening performance measures* are based on the initial assessment of mammograms indicated by the radiologists to be for screening purposes. We classified a screening mammogram as positive if it was given an initial BI-RADS assessment of 0 (need additional imaging evaluation), 4 (suspicious abnormality), or 5 (highly suggestive of cancer).(12) We classified as negative those mammograms given an assessment of 1 (negative) or 2 (benign). A BI-RADS assessment of 3 (probably benign) was considered positive if there was a recommendation for immediate work-up and negative otherwise. Women were considered to have had breast cancer if a diagnosis of invasive carcinoma or ductal carcinoma *in situ* occurred within 1 year of the examination and prior to the next screening mammogram.

For *diagnostic mammography examinations*, performance is based on the final BI-RADS assessment at the end of imaging workup. Diagnostic mammography include examinations performed for the additional workup of an abnormal screening mammogram; short-interval

follow-up; or evaluation of a breast problem. We classified diagnostic mammograms as positive if given a BI-RADS 4 or 5. We classified mammograms as negative if given a final BI-RADS assessment of 1 or 2. A BI-RADS assessment of 0 or 3 with a recommendation for FNA, biopsy, or surgical consult is considered positive and negative if no such recommendation is made.

On the website radiologists can choose to view several performance measures based on the number of true-positive (TP), false-positive (FP), true-negative (TN), false-negative (FN), and total mammograms (N) summed over the years of interest. The performance measures are calculated as follows: Sensitivity = $TP/(TP + FN)$, Specificity = $TN/(TN + FP)$, FP rate = $FP/(TN + FP)$, abnormal interpretation rate (AIR, sometimes called recall rate) = $(FP + TP)/N$, CDR (cancer detection rate) = TP/N , PPV1 using initial assessment = $TP/(TP + FP)$, and PPV2 using final assessment = $TP/(TP + FP)$. Regional results included data from radiologists in the local area. National results included data from all seven registries in the BCSC. Benchmarks were based on the 25th and 75th percentile for each performance measure.

Website design

Website design was based on *Research-Based Web Design and Usability Guidelines*, which were updated in 2006.(13) We followed basic tenets such as involving users in the design, types of navigation tools, and page layout. We worked with a graphic designer and editor to create pages that were visually pleasing and easy to read. Radiologists could choose to view data as tables, bar graphs, or line graphs. A scatter plot of cancer detection rate vs. recall rate was also available (see figure 1 examples of page). In addition to the various reports available to the radiologists, the website had a Frequently Asked Questions (FAQ) page and a Definitions page so that radiologists could learn more about how to use the website and how to understand the data in the tables and graphs.

The website has a log-in page as described above. Once a BCSC radiologist is logged on, he or she is able to select the years of data to be viewed, screening or diagnostic mammography, and type of audit statistics (with a default to sensitivity). Then they choose among five display formats to customize the report. Each page displays the radiologist's user ID. The reports are able to be printed locally and the bottom of each page states, "Confidential - please file in a locked file cabinet or shred when finished with report."

Study population

One hundred and eleven radiologists who participate in three Breast Cancer Surveillance Consortium (BCSC) registries in the United States (New Hampshire Mammography Network, North Carolina Registry and Vermont Breast Cancer Surveillance System) were invited to use the website. BCSC radiologists were eligible to participate if they contributed data any time during 2000–2006, interpreted at least 480 mammograms in any one year (the minimum volume required by MQSA)(8), and were still practicing at a BCSC facility in 2007. The BCSC principal investigators sent radiologists in their registries an invitation letter explaining the website and provided a unique user ID and password in the summer of 2009 and in the spring of 2010. The participating radiologists entered this log-in information, which let the website identify the radiologist and provide them with a report of their BCSC data in comparison to regional and national data. At any time during the year, they could log on to the website and review their data.

We captured data electronically each time radiologists logged on to the website including what pages they visited, whether they looked at performance measures for screening and/or

diagnostic mammography and the visual format they viewed. We also asked the radiologists to complete an online survey at the end of their session about the website.

Each registry and the Statistical Coordinating Center (SCC) have received institutional review board approval for either active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analytic studies. All procedures are Health Insurance Portability and Accountability Act (HIPAA) compliant and all registries and the SCC have received a Federal Certificate of Confidentiality and other protection for the identities of women, physicians, and facilities in this research. In addition for this study our invitation letters stated the benefits and harms of participating, and we assumed passive consent if the radiologist logged on to the website.

Results

From July 2009 through July of 2010, 35 of the 111 (32%) radiologists invited logged on and used the website. Of the 35 radiologists who used the website, 13 (37%) completed a post use survey. All 13 radiologists who completed the survey had previously reviewed their outcome audit data sent by their registry, nine of whom reviewed their outcome data annually, two every six months and two reported reviewing outcome data occasionally. Eight radiologists reported that they have been reading mammograms for more than 20 years while the remaining five radiologists reported reading mammograms between 10 and 20 years. All 13 of the radiologists who completed the survey agreed or strongly agreed that the website was easy to use and easy to navigate. Nine radiologists reported that they intend to improve a specific outcome measure this year (data not shown).

Table 1 reports other results from the website evaluation survey. Most radiologists found the various visual formats useful. Examples of the visual formats can be viewed in Figure 1. Four comments at the end of the survey provided some additional feedback. Comment 1 said “Breaking it down per year would also help us see if we are improving our call back and sensitivity.” Comment 2 said “Explaining more about Confidence Intervals (CI) and percentages would be helpful. Telling us the range we should strive for would be helpful. What are the national averages for all the percentages and numbers?” The third comment was “How do we get specific patient names to review false negatives?” And finally the last comment was “To be more in the norm I need to increase sensitivity which means more recalls and more biopsies. This will lower my specificity as my negative biopsy rate will increase.”

Table 2 reports the number of radiologists who used the website by month, the number of sessions and the number of pages reviewed. The three registries sent their initial invitations out during three different months and there was highest use one month after the invitation was mailed (August – October, 2009; April and May 2010). Twenty-three radiologists visited the website once, while 10 visited it twice, 1 visited 3 times and 1 radiologist visited the website 5 times for a total of 51 visits.

The web page viewed most frequently was the Outcomes Table for Screening Mammography (Table 3). It was viewed 221 times by all 35 radiologists. When a radiologist logs on and chooses “Go To Reports” it automatically goes to the Outcomes Table for Screening Mammography for all seven years. The second most popular web page was the Outcomes Table for Diagnostic Mammography which was viewed 39 times by 15 radiologists, followed by the 2 × 2 Table for Screening Mammography which was viewed 37 times by 17 radiologists. The line and vertical bar graphs for sensitivity of screening mammography were visited 35 times each. The FAQ page was accessed 22 times by 10 radiologists while the Definitions page was accessed 24 times by 9 radiologists.

Discussion

Many countries with organized breast screening programs have developed audit and feedback systems based on their national data to help radiologists assess and improve their skills; however, the US does not offer a similar program to their radiologists. Audit feedback can improve medical practice (14–16) but currently is not routinely used for mammography in the US other than the minimal MQSA requirement. Based on the results of focus groups with radiologists in three states (WA, NH, VT) (17) we developed and piloted a web-based outcome audit and feedback system for radiologists participating in three BCSC registries.

Radiologists who participate in the BCSC have long benefited from receiving paper outcome audits from their local registries, albeit without national comparisons. Benchmarks let radiologists compare their performance with that of others and to accepted practice guidelines. Many of the early performance benchmarks were developed based on the evaluation of outcomes from small groups of breast imaging specialists or using the opinions of experienced radiologists.(12, 18)

The BCSC published benchmarks for both screening and diagnostic mammography based on the performance of community radiologists in 2005 and 2006.(1, 2) These benchmarks are updated annually on the BCSC website (<http://breastscreening.cancer.gov/>). Subtle differences in the way data are calculated (adjusted or unadjusted) and variations in definitions used to determine a positive and negative exam (including BI-RADS[®] assessment category 3 as negative regardless of the management recommendation), make it complicated for radiologists to make an exact comparison to the published benchmarks. An advantage of our audit website is that the same definitions were used for the radiologist's individual performance measures and the regional and national benchmarks.

Although radiologists in our previous focus groups wanted the flexibility of seeing the data in different visual formats, the outcomes table was the most common format used (17). Perhaps busy radiologists did not want to take the time to look at the data in more than one format and because the outcome table provides all of the measures on one page it was the most convenient to review.

The four comments on the survey were informative. Comment 1, “Breaking it down per year” and Comment 2 “Explaining more about Confidence Intervals (CI) and percentages.... Telling us the range we should strive for would be helpful. What are the national averages for all the percentages and numbers?” informed us that these radiologists were not aware of the existing functions on the website and that we need clearer descriptions of what is available and more detailed definitions. We were surprised by the next comment “How do we get specific patient names to review false negatives?” because we thought that the local registries who provide papers audits also provided lists of patients who had false negatives. The BCSC does not have access to names of patients and cannot provide this information yet it is vitally important for radiologists to learn from the review false negative cases.(19) An important function of radiology information systems would be to produce these types of lists for radiologist review.

Only a small proportion of invited radiologists used the website and only 37% of those who used the website completed the survey so our results may not be generalizable to all U.S. radiologists who read mammography. The 22 radiologists who did not complete our survey may have not liked using the website. Currently this website is only available to radiologists who participate in the BCSC and not to all radiologists in the US. We do not know whether other breast imaging facilities are able to export TP, FP, TN and FN data separated into screening and diagnostic mammography from their computer systems. With the advent of the American College of Radiology's National Mammography Database, more radiologists

will be able to export these data elements in the future. The BCSC matches mammograms with pathology and cancer registries to identify TP and FN exams, so can calculate performance measures such as sensitivity and specificity. Most breast imaging practices, even those participating in the National Mammography Database, are not able to completely capture cancers matched to mammograms. Although we provided most of the information that was mentioned in the comments of the survey, the radiologists did not know it was available and did not access this information. Cancer registry data are only available to the BCSC two to three years after the cancer diagnosis date so all measures requiring cancer status will always be several years behind current mammography assessments. Also technology is changing rapidly in the field of breast imaging. The use of digital and computer assisted detection are rapidly disseminating throughout the US and this will influence the outcome audit results.(20)

Most radiologists who accessed the website did so in the month or two following the invitation letter. Because we are planning to update the data annually it may not be necessary to visit the website more frequently than once a year. However, to enhance the use of the audit feedback one would need to develop a reminder system for the radiologists to check the website at least annually or develop incentives such as CME credit or documenting regular participation to become and maintain an ACR “Center of Excellence.” Or our legislators could consider making it mandatory to review complete audit reports when MQSA is reissued.

The website cost about \$55,000 to develop. This covered the cost of a programmer, graphic and word editors, a data manager and two investigators. We estimate that there will be a modest cost to annually update the data and to maintain the website.

Along with these limitations are considerable strengths. The radiologists who used the website found it useful to help guide changes in their interpretative goals. Many radiologists are not familiar with published interpretive goals (Jackson under review) and are not always accurate in knowing what their outcome statistics are compared to their peers (Cook under review). This website continually provides accurate individual radiologist data with comparisons to national and regional data that are calculated all the same. Radiologists appeared to use different types of visual formats and reviewed different outcome measures. This can only occur with an interactive website. The American College of Radiology’s National Mammography Database has recently started to provide some audit feedback to their participating facilities using the BCSC data as benchmark comparisons but it is currently not interactive (personal communication, Mythreyi Chatfield, 7/13/11).

We are expanding the website to also provide information at the facility level which should be ready for BCSC facilities in the summer of 2012, and we plan to make the website public in early 2013, so that all radiologists and facilities will be able to get audit reports, with benchmark comparisons, after entering their own information. The shell of our website is available to be used by other countries or screening programs.

An interactive website to provide customized mammography audit feedback reports to radiologists has the potential to be a powerful tool in improving interpretive performance. The conceptual framework of customized audit feedback reports can also be generalized to other imaging tests.

Acknowledgments

We thank Dr. Bonnie Yankaskas and her staff at the Carolina Mammography Registry, Dr. Tracy Onega and her staff at the New Hampshire Mammography Network, and Kathleen Howe at the Vermont Breast Cancer Surveillance System for their help in inviting participating radiologists into the study. Data collection for this work

was supported by a NCI-funded Breast Cancer Surveillance Consortium co-operative agreement (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040). A special competitive supplement to U01CA70013 funded this specific study. The collection of cancer incidence data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: <http://breastscreening.cancer.gov/work/acknowledgement.html>

References

1. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology*. 2006; 241(1):55–66. [PubMed: 16990671]
2. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. *Radiology*. 2005; 235(3):775–90. [PubMed: 15914475]
3. Barlow WE, Lehman CD, Zheng Y, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. *J Natl Cancer Inst*. 2002; 94(15):1151–9. [PubMed: 12165640]
4. Elmore JG, Jackson SL, Abraham L, et al. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology*. 2009; 253(3): 641–51. [PubMed: 19864507]
5. Lewis RS, Sunshine JH, Bhargavan M. A portrait of breast imaging specialists and of the interpretation of mammography in the United States. *AJR Am J Roentgenol*. 2006; 187(5):W456–68. [PubMed: 17056875]
6. van der Horst F, Hendriks J, Rijken H, et al. Breast cancer screening in the Netherlands: Audit and training of radiologists. *Seminars in Breast Disease*. 2003; 6(3):114–22.
7. Perry NM. Interpretive Skills in the National Health Service Breast Screening Programme: performance indicators and remedial measures. *Seminars in Breast Disease*. 2003; 6(3):108–13.
8. Regist, F., editor. *Quality mammography standards--FDA. Final rule*. 1997. p. 55852-994.
9. Nass, SB.; John. *Improving Breast Imaging Quality Standards*. Institute of Medicine and National Research Council of the National Academies; Washington, DC: 2005.
10. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR Am J Roentgenol*. 1997; 169(4): 1001–8. [PubMed: 9308451]
11. Elmore JG, Aiello Bowles EJ, Geller B, et al. Radiologists' attitudes and use of mammography audit reports. *Acad Radiol*. 2010; 17(6):752–60. [PubMed: 20457418]
12. D'Orsi, C.; Bassett, L.; Berg, W., et al. *ACR BI-RADS-Breast Imaging Atlas*. 4. Reston VA: American College of Radiology; 2003.
13. USHHS. *Research-Based Web Design and Usability Guidelines*. 2006.
14. Jamtvedt G, Young JM, Kristoffersen DT, et al. Does telling people what they have been doing change what they do? A systematic review of the effects of audit and feedback. *Qual Saf Health Care*. 2006; 15(6):433–6. [PubMed: 17142594]
15. Jamtvedt G, Young JM, Kristoffersen DT, et al. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2006; (2):CD000259. [PubMed: 16625533]
16. Veloski J, Boex J, Grasberger M, et al. Systematic review of the literature on assessment, feedback and physicians' clinical performance BEME Guide No. 7. *Med Teach*. 2006; 28:117–28. [PubMed: 16707292]
17. Bowles E, Geller B. Best ways to provide feedback to radiologists on mammography performance. *AJR Am J Roentgenol*. 2009; 193(1):157–64. [PubMed: 19542408]
18. Bassett, L.; Hendrick, R.; Bassford, T., et al. *Clinical practice guidelines No 13*. Agency for Health Care Policy and Research; Rockville, MD: 1994. Quality determinants of mammography.
19. Siegal E, Angelakis E, Hartman A. Can peer review contribute to earlier detection of breast cancer? A quality initiative to learn from false-negative mammograms. *Breast J*. 2008; 14(4):330–4. [PubMed: 18540960]
20. Glynn C, Farria D, Monsees B, et al. Effect of Transition to Digital Mammography on Clinical Outcomes. *Radiology*. 2011

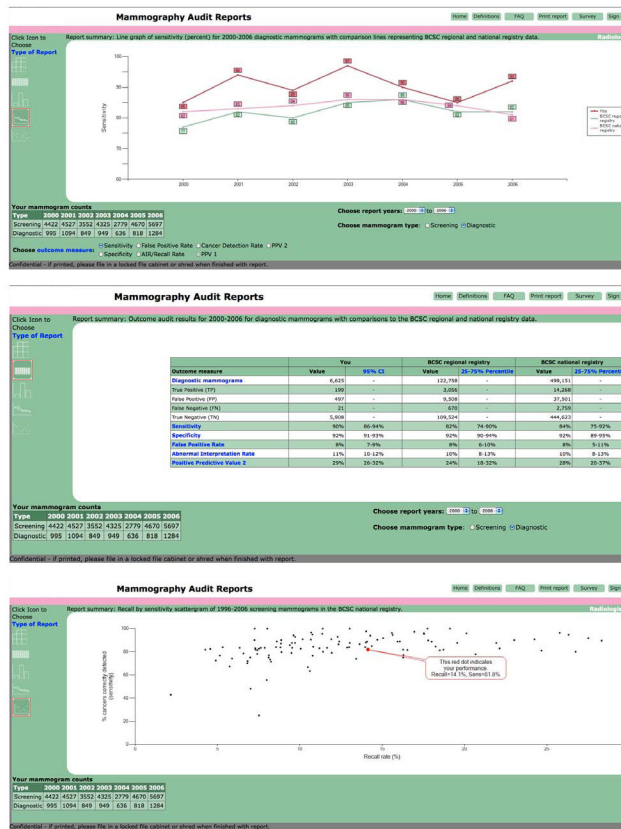


Figure 1. Sample outcome audit formats. Top of figure is a line graph for sensitivity. Middle of figure is an outcomes table with all measures for diagnostic mammography. Bottom of figure is a scattergram of recall by sensitivity for screening mammography.

Table 1

Results of post use website survey

Survey Questions	Disagree	Neutral	Agree	Strongly agree	NA (did not use)
The website was easy to use			1	12	
The website was easy to navigate			1	12	
I found all the explanations I wanted		1	5	6	1
I liked have the regional comparisons		1	5	7	
I liked having the national comparisons			6	7	
The 2 × 2 table was useful	1	1	3	3	3
The one page audit table was useful			4	9	
The bar graphs were useful		4	5	2	2
The line graphs were useful	1	3	5	1	3

Table 2

Radiology feedback web site use by month

Year/Month of session	Number of pages viewed	Number of sessions	Number of radiologists
2009/07	97	2	2
2009/08	138	7	6
2009/09	99	4	4
2009/10	187	10	9
2009/11	29	2	2
2010/02	10	1	1
2010/03	35	2	1
2010/04	177	6	6
2010/05	289	14	13
2010/06	18	2	2
2010/07	25	1	1
Total	1104	51	

* 12/2009 and 01/2010 had no visits to the website

Table 3

Diagnostic and screening: Outcome measures by visual format

Type of mammogram	Visual format	Outcome measure	Number of times viewed	Number of radiologists who viewed at least once
Screening	Outcomes table	N/A	221	35
	2 × 2 table	N/A	37	17
	Line graph	CDR*	9	7
	Line graph	False positive rate	7	6
	Line graph	PPV1	6	5
	Line graph	Recall rate	3	3
	Line graph	Sensitivity	35	19
	Line graph	Specificity	10	7
	Vertical bar graph	CDR	8	5
	Vertical bar graph	False positive rate	4	4
	Vertical bar graph	PPV1	9	6
	Vertical bar graph	Recall rate	6	6
	Vertical bar graph	Sensitivity	35	18
	Vertical bar graph	Specificity	7	6
	Scattergram	N/A	30	15
Diagnostic	Outcome table	N/A	39	15
	2 × 2 table	N/A	16	7
	Line graph	Recall rate	6	1
	Line graph	Sensitivity	6	5
	Vertical bar graph	CDR	3	2
	Vertical bar graph	False positive rate	4	2
	Vertical bar graph	PPV2	2	2
	Vertical bar graph	Recall rate	5	2
	Vertical bar graph	Sensitivity	18	6
Vertical bar graph	Specificity	5	2	

• CDR – Cancer Detection Rate