



Published in final edited form as:

*J Am Coll Radiol.* 2020 November ; 17(11): 1420–1428. doi:10.1016/j.jacr.2020.07.019.

## Trends in attaining mammography quality benchmarks with repeated participation in a quality measurement program: Going Beyond the Mammography Quality Standards Act to address breast cancer disparities

Garth H Rauscher, PhD<sup>1</sup>, Katherine Tossas-Milligan<sup>1,2</sup>, Teresita Macarol<sup>3</sup>, Paula M Grabler<sup>4</sup>, Anne Marie Murphy, PhD<sup>5</sup>

<sup>1</sup>Division of Epidemiology and Biostatistics, University of Illinois at Chicago, Chicago, IL, USA

<sup>2</sup>University of Illinois Cancer Center, Chicago, IL, USA

<sup>3</sup>Advocate Health Care, Downers Grove, IL, USA

<sup>4</sup>Rush University Medical Center, Chicago, IL, USA

<sup>5</sup>Metropolitan Chicago Breast Cancer Task Force, Chicago, IL, USA.

### Abstract

**Background:** The Mammography Quality Standards Act (MQSA) requires that mammography facilities conduct an audit but there are no specifications on the metrics to be measured. In a previous mammography quality improvement project, we examined whether breast cancer screening facilities could collect the data necessary to show that they met certain quality benchmarks.

**Purpose:** Here we present trends from the first five years of data collection, to examine whether continued participation in this quality improvement program was associated with an increase in the number of benchmarks met for breast cancer screening.

**Materials and Methods:** Participating facilities across the state of Illinois (N=114) with at least two time points of data collected in (years 2006, 2009, 2010, 2011 and/or 2013) were included. Facilities provided aggregate data on screening mammograms and corresponding diagnostic follow-up information, which was used to estimate 13 measures and corresponding benchmarks for patient tracking, callback, cancer detection, loss to follow-up and timeliness of care.

---

Address correspondence to Dr. Garth Rauscher, Division of Epidemiology and Biostatistics, School of Public Health, MC 923, University of Illinois at Chicago, Chicago, Illinois 60612. garthr@uic.edu, phone (312) 413-4317.

#### AUTHOR CONTRIBUTIONS

Conception and design: AMM, GHR

Collection and assembly of data: KTM, GHR

Data analysis and interpretation: GHR, AMM

Manuscript writing: all authors

Final approval of manuscript: all authors

Accountable for all aspects of the work: all authors

No potential conflicts of interest exist.

The authors declare that they had full access to all of the data in this study and the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis.

**Results:** The number of facilities able to show that they met specific benchmarks increased with length of participation for many but not all measures. Trends towards meeting more benchmarks were apparent for cancer detection, timely imaging, not lost at biopsy, known minimal status ( $p<0.01$  for all), and proportion of screen-detected cancers that were minimal and early stage ( $p<0.001$  for both).

**Conclusion:** Participation in the quality improvement program appeared to lead to improvements in patient tracking, callback/detection, and timeliness benchmarks.

### Keywords

Breast Cancer; Mammography; Quality Improvement; Mammography Quality Standards Act

---

## INTRODUCTION

The purpose of a high quality mammography program is to detect breast cancer at its earliest and most minimal stages when there is the best chance of effective treatment and to enhance follow-up with identified abnormalities in a timely fashion, which also maximizes the best chance at survival. The federal Mammography Quality Standards Act (MQSA) requires mammography facilities to carry out a medical audit so as to measure quality of mammography but provides no guidelines as to what measures should be included in such an audit. We previously used data collected by the Chicago Breast Cancer Quality Consortium (a project of Equal Hope (formerly the Metropolitan Chicago Breast Cancer Task Force) to examine whether a wide array of different types of mammography facilities could collect the data necessary to examine their mammography quality and demonstrate whether they met certain quality benchmarks (1). In our prior research we found that collection of a wide range of benchmarks by many different types of facility are feasible but most facilities did not meet at least some of these benchmarks (1). In the present analyses, we use data collected across five separate calendar years of mammography screening to examine whether facilities that are in the program for longer periods of time would demonstrate greater attainment of these benchmarks. Participation provided each facility with its own confidential site-specific report that was sent to individuals in leadership positions at each institution (Chief Executive Officer, Chief Medical Officer, Vice President for Quality). The project did not involve any public reporting of results so as to promote widespread participation.

## MATERIALS AND METHODS

As described previously, retrospective, aggregate data for these analyses were collected by the Chicago Breast Cancer Quality Consortium (Consortium), which was a project of Equal Hope (2,3). Participation in the program involved voluntary collection of data on mammography quality metrics and care processes; feedback to all participating institutions regarding benchmark attainment or lack thereof and best practices, and additional data collection at select institutions demonstrating potential quality/care process deficits for the purpose of assisting sites in the development of quality improvement projects.

Facilities were located throughout the state of Illinois, with more participation by facilities located within metropolitan Chicago as these were invited to participate first and these facilities had more representatives involved with the establishment of Equal Hope. After legislation was passed into law, the statewide program was initiated in 2012 and collected data for the years 2011 and 2013. Facilities of all types were represented in the dataset including large academic facilities, multi-institutional health system facilities, small rural facilities, public facilities, safety net facilities, urban, suburban and rural facilities. Of the 25 facilities with data at all five time points, nine were BICOE facilities, and five were designated as Disproportionate Share facilities by the state of Illinois which receive additional payments for serving underserved populations (1).

Data submitted by facilities across the state of Illinois for screening mammograms performed in 2006, 2009, 2010, 2011 and 2013 were included in these analyses. We established a standardized data entry form with controls built in to check for invalid data values and consistency across data elements, and data were reviewed centrally upon receipt to check for errors and request resubmissions when needed. Standard operating procedures for data collection and submission were provided. However, each facility had their own individualized, site-specific process for collecting data. Many utilized data from their commercially purchased mammography tracking systems, or their electronic medical records system but smaller less resourced facilities sometimes used a paper log system. Facilities submitted raw counts and were not involved in calculating the metrics and determining whether they met the benchmarks or not. For calendar years 2006, 2009 and 2010, each participating institution obtained a signed data sharing agreement and Institutional Review Board (IRB) approval for the study from their institution; this was done with the federal protections of the Patient Safety and Quality Improvement Act. Institutions that lacked an IRB used the Rush University IRB.

Electronic data collection forms were designed for collecting data on the screening mammography process pertaining to screening mammograms conducted during each calendar year (1). Each year, prior to data submission, a series of webinars were conducted in order to familiarize staff at each institution with the data collection form and submission process, and emphasize specific points pertaining to submitting accurate data. Each facility submitted aggregate counts pertaining to the number of screening mammograms overall and by finding, using the Breast Imaging and Reporting Data System (BIRADS). BIRADS 0,4,5 indicate an abnormal finding that requires diagnostic follow-up, BIRADS 1,2 indicate a normal or benign finding that typically results in a recommendation for routine follow-up such as another screening mammogram in 12 months, and BIRADS 3 indicates a probably benign finding that typically results in a recommendation for short-interval (e.g. 6 month) follow-up. Facilities reported on the number of screening mammograms performed in a 12-month period and the number that were determined to be abnormal (BIRADS 0, 4, 5) so as to calculate a recall rate. For abnormal screens, facilities submitted the number of screens that received diagnostic follow-up within 12 months (not lost to follow-up) and of those with follow-up whether it was done within 30 days (timely); number that resulted in a recommendation for biopsy; number that received a biopsy within 60 days of abnormal screen (timely) and within 12 months of the abnormal screen (not lost to follow-up); number of cancers ultimately detected through screening (cancer detection rate); number of cancers

for which minimal status and stage at diagnosis were known; and number of cancers that were identified as early stage and minimal. From these data we estimated 13 measures of the screening mammography process, grouped into two categories: Radiologist related (recall rate, cancer detection rate, and yield of cancers from abnormal screens and from biopsies) and care process related (timeliness, patient tracking, and loss to follow-up).

Benchmarks and the decisions regarding thresholds for meeting benchmarks were informed from a variety of sources, including American College of Radiology benchmarks, population-based estimates and ranges for these measures from the Breast Cancer Surveillance Consortium (4-6) and clinical experts on the Chicago Breast Cancer Quality Consortium's Mammography Quality Advisory Board. We calculated each of the above 13 measures separately for each institution when both numerator and denominator data were available. When either numerator or denominator was missing, then a facility was defined as not being able to show that they met that particular benchmark.

### Statistical analysis

We tabulated the proportion of facilities meeting each specific benchmark by time point and examined trends in these proportions, overall and after stratifying results by the number of time-points available for analysis (2, 3, 4 or 5). Due to changes in data collection over time, data on timely follow-up imaging was not collected for calendar year (CY) 2010 and data on known minimal status and a known stage at diagnosis were not collected for CY 2006. Therefore, for the corresponding benchmarks the maximum number of time points was four. Tests for trend were based on P-values from a Wald test for trend in logistic regression with generalized estimating equations (GEE) to account for the clustering of multiple time-points by facility. Sample size was based entirely on the number of facilities participating at each time point. In secondary analyses of the 25 facilities with data at all 5 time points, we stratified facilities into those that were designated as an American College of Radiology (ACR) Breast Imaging Center of Excellence (BICOE) (<http://www.acr.org/quality-safety/accreditation/bicoe>) (N=9) and those that were not (N=16). BICOE facilities are fully accredited in mammography, stereotactic breast biopsy and breast ultrasound by the ACR. We repeated the above models, including an interaction between calendar year and BICOE status. We reported the continuous P-value for that product term to indicate statistical evidence (or lack thereof) that trends in benchmarks being met over time differed for BICOE and non-BICOE facilities.

The following seven measures of recall and cancer detection were calculated for all participating facilities: These are standard measures that relate to the ability to detect breast cancer, to detect it when it is small, and to call back an appropriate number of patients for diagnostic follow-up care.

- **Recall rate:** The proportion of screening mammograms interpreted as abnormal (BI-RADs 0, 4 or 5). Too low a recall rate would increase the chances of a missed breast cancer detection, whereas too high a recall rate would lead to unnecessary diagnostic follow-up imaging. The benchmark for recall rate was met if no less than 5% and no greater than 14% of screening mammograms were interpreted as abnormal.

- Biopsy recommendation rate: The proportion of abnormal screening mammograms resulting in a recommendation for biopsy (benchmark of 8-20%). Too low a biopsy recommendation rate would increase the chances of a missed breast cancer detection, whereas too high a rate would lead to unnecessary biopsies.
- Cancers from abnormal screens (PPV1): The proportion of abnormal screens that received a breast cancer diagnosis within 12 months of the screen, also known as positive predictive value 1 (PPV1, benchmark of 3-8%).
- Cancer from biopsied (PPV3): The proportion of patients biopsied following an abnormal screen that received a breast cancer diagnosis within 12 months of the screen, also known as PPV3 (benchmark of 15-40%).
- Cancer detection rate: The number of breast cancers detected following an abnormal screen for every 1000 screening mammograms performed (benchmark of 3-20 per 1000).
- Proportion minimal: Breast cancer screening is intended to detect tumors when they are small. The benchmark for the proportion of screen-detected breast cancers that were either in situ or no greater than 1 cm in largest diameter was established at >30%. Breast cancers with unknown minimal status were excluded from both numerator and denominator of this measure. While we attempted to collect information on lymph node status for minimal cancers, many institutions were unable to provide this data reliably; therefore, lymph node status was not included in our definition of minimal cancer.
- Proportion early stage: Breast cancer screening is intended to detect tumors when they are early stage. The benchmark for the proportion of screen-detected breast cancers that were either in situ or stage 1 was established at >50%. Breast cancers with unknown stage were excluded from both numerator and denominator of this measure.

The following two measures of timeliness were calculated for all participating facilities: These are standard measures that relate to the ability to provide diagnostic follow-up care within a reasonable time frame following an abnormal screening mammogram result.

- Timely follow-up imaging: The receipt of diagnostic imaging within 30 days of an abnormal screen, among those receiving diagnostic imaging within 12 months of the screen (benchmark of 90% and above).
- Timely biopsy: The receipt of a biopsy within 60 days of the abnormal screen, among those receiving a biopsy within 12 months of the screen (benchmark of 90% and above).

The following four measures of loss to follow-up were calculated for all participating facilities: If patients are lost during diagnostic follow-up, this affects cancer detection rates and other benchmarks, because it is not possible to include their cancer diagnoses in any of the detection benchmarks or to know whether their cancer was diagnosed early.

- Not lost at imaging: The proportion of abnormal screening mammograms receiving follow-up diagnostic imaging within 12 months of the screening mammogram (benchmark of 90% and above).
- Not lost at biopsy: The proportion of women with a biopsy recommendation that received a biopsy within 12 months of the abnormal screen (benchmark of 70% and above).
- Known minimal status: The proportion of screen-detected cancers known to be either minimal or not minimal (benchmark of at least 80%).
- Known stage at diagnosis: The proportion of screen-detected cancers with known stage (benchmark of at least 80%).

## RESULTS

There were 114 mammography facilities that contributed at least two time points of data: 53 sites contributed exactly 2 data points, 20 contributed exactly 3 time points, 16 contributed exactly 4 time points, and 25 sites contributed exactly 5 time points. Sites were grouped by number of time points provided and analyses were stratified in this manner; analyses were also conducted on all facilities combined. Duration of participation in the program was associated with greater attainment of benchmarks for some but not for all measures.

### Imaging quality measures:

The proportion of sites meeting the benchmark for recall rate did not change while the proportion meeting the benchmark for biopsy recommendation improved only among the 20 sites contributing 3 time points (from 40% to 70%,  $p=0.06$ ). Cancer detection measures (cancer yield from abnormal screens or PPV1, cancer yield from biopsies or PPV3 and screen cancer detection rate) increased or marginally increased for 5 of 12 measures ( $p=0.02 - 0.15$ ). When all time points were combined for analyses, screen-detection benchmarks appeared to strongly improve across time points with  $p$ -values for trend ranging from 0.004 to 0.007. Attainment of benchmarks for proportion minimal and early stage also strongly improved with duration of participation (Table 2 and Figure 1).

### Timeliness:

The proportion of facilities meeting the specific timeliness benchmarks increased or marginally increased for 4 of 8 analyses ( $p=0.01 - 0.12$ ). When all time points were combined for analyses, corresponding benchmarks appeared to improve across time points for timely imaging but timely biopsy did not show improvement (Table 3 and Figure 1).

### Loss to follow-up and results tracking:

The proportion of facilities meeting specific benchmarks increased or marginally increased for 3 of 14 analyses ( $p = 0.02$  to 0.16). When all time points were combined for analyses, corresponding benchmarks appeared to improve across time points for not lost to follow-up at biopsy and known minimal status, but not for not lost to follow-up at imaging or known stage (Table 4 and Figure 1).

### Trends for meeting benchmarks by BICOE status:

While trends for meeting most benchmarks did not vary by BICOE status ( $p>0.20$ ), there was some evidence that improvements in recall rate ( $p=0.16$ ), not lost to follow-up at imaging ( $p=0.13$ ), biopsy recommendation rate ( $p=0.05$ ), and proportion early stage ( $p=0.01$ ) were greater for the 17 non-BICOE than for 8 BICOE facilities (results not shown).

## DISCUSSION

The American College of Radiology recommends that facilities meet certain additional quality benchmarks above and beyond the requirement of the Mammography Quality Standards Act to perform an audit. These include measuring the proportion of screening mammograms determined to be abnormal (recall rate), timeliness of follow-up, extent of screen-detection (i.e., cancer detection rate for screening mammograms), and ability to detect small and early stage tumors. In this quality improvement program, we estimated proportions for 13 quality indicators of each facility's breast screening process, and the proportion of sites that met each of the 13 corresponding benchmarks for patient tracking, callback, cancer detection, and timeliness. Our results suggest that increased duration of participation in this program resulted in an overall improvement in the attainment of benchmarks. Nonetheless, lack of timeliness of diagnostic imaging and biopsy services remain a clear shortcoming for many facilities. In addition, the ability to receive results for follow-up not performed at the same facility, and the ability to determine and review staging information for cancers detected from an abnormal screen remains challenging for many facilities. This is especially true for facilities which do not perform biopsies or which are likely to have their patients go elsewhere for biopsies. We found some evidence that non-BICOE facilities showed larger improvements for certain benchmarks, consistent with the notion that these facilities benefitted more from participation in the intervention; however, our ability to detect differences in trends was limited for this interaction analysis due to reduced sample size after stratifying facilities by BICOE status.

We believe that much of the observed improvement in meeting benchmarks stems from an increased ability to collect and report the data needed to construct these measures. Prior to each round of data collection, a series of webinars were conducted that included an emphasis on specific points, which pertain to submitting accurate data. It is likely that with each successive training webinar, sites continued to learn about the importance of separating screening from diagnostic mammograms, having complete follow-up of diagnostic results, and other issues. More accurate and complete collection of screen detected cancer over time would have also produced more accurate cancer detection metrics. To the extent that sites were learning how to do a better job of tracking results and therefore were capturing more of their screen-detected cancers, this could help explain the increase in benchmarks met for cancer detection over time.

Only 46% of timely imaging and 38% of timely biopsy measures across facilities and time points met their corresponding benchmarks. These metrics are associated with specialized clinical services that might require additional organizational investment in order to expand services that are over taxed or absent at the facility, requiring patients to seek diagnostic follow-up elsewhere. Many of the more limited service locations are safety net facilities with

few resources to support navigation of a patient to another institution without any anticipation of revenue by the referring facility.

This project demonstrates the feasibility of implementing the American College of Radiology recommendations broadly and suggests that improvement in the quality of mammography would come from such implementation. One potential limitation is that if practices employed a higher number of fellowship trained breast imagers over time that could account for the increase in performance separate from participation in this program. However, there is no reason to suspect that this happened. Most observed improvement was related to better data collection which could be the result of better processes resulting from the intervention. Another limitation would arise if a large health system updated or improved their mammography tracking systems across all or most of their sites, this might also lead to improvements that were unrelated to this project; however, during our interactions with these facilities we did not observe any system-wide updates in mammography tracking systems.

To date, none of the metrics discussed and recommended by the American College of Radiology have been incorporated into MQSA guidelines. An initiative called the Enhancing Quality Using the Inspection Program (EQUIP), launched in 2017, augments the existing MQSA inspection system to more closely monitor the quality of mammographic imaging and ensure that corrective actions are taken and documented when poor quality imaging is identified (7). In the same way that image quality has not been rigorously overseen by the MQSA inspection system in the past, there is very little oversight of the requirement that sites have a system for tracking mammogram results (1) and what constitutes a system is left up to each facility to determine on their own. In this quality improvement program, we found that when sites are trained on how to collect and submit data on mammography outcomes, their benchmark measures improve over time, providing evidence that could inform an initiative towards improving tracking of mammography outcomes to augment the MQSA.

It should be also be possible for quality assessments to drive improvement in image quality benchmarks, though there is likely insufficient volume at smaller facilities to adequately monitor these metrics individually. From a policy perspective, payors may want to consider whether they require providers to implement some of these quality measures as a condition for participation in network. Additionally, this project provided confidentiality to participating providers. A public project that required transparency in results might drive more rapid improvement. However, given that this was a voluntary program, this program encouraged participation by providing both confidentiality as well as technical assistance and free provider training to participants. Ultimately, a more rigorous system than is currently present through MQSA is needed to reduce variability in the quality of mammography and to more adequately guarantee all women high quality mammography.

### **Acknowledgment:**

We thank the facilities who participated in the Chicago Breast Cancer Quality Consortium and who provided data for these analyses. This research was funded through a grant (R01HS018366) from the Agency for Health Research and Quality, and two grants from the National Cancer Institute (P01CA154292 and 2P01CA15292).



Funding: This work was supported by grants from the Illinois Department of Healthcare and Family Services from March 1, 2012 to February 28, 2016, the Avon Breast Cancer Crusade and Susan G. Komen.

## Abbreviations:

<b>BICOE</b>	Breast Imaging Centers of Excellence
<b>BIRADS</b>	Breast Imaging and Reporting Data System
<b>CY</b>	calendar year
<b>EQUIP</b>	Enhancing Quality Using the Inspection Program
<b>IRB</b>	Institutional Review Board
<b>MQSA</b>	Mammography Quality Standards Act
<b>nHB</b>	non-Hispanic Black
<b>nHW</b>	non-Hispanic White
<b>PPV1</b>	positive predictive value 1
<b>PPV3</b>	positive predictive value 3.

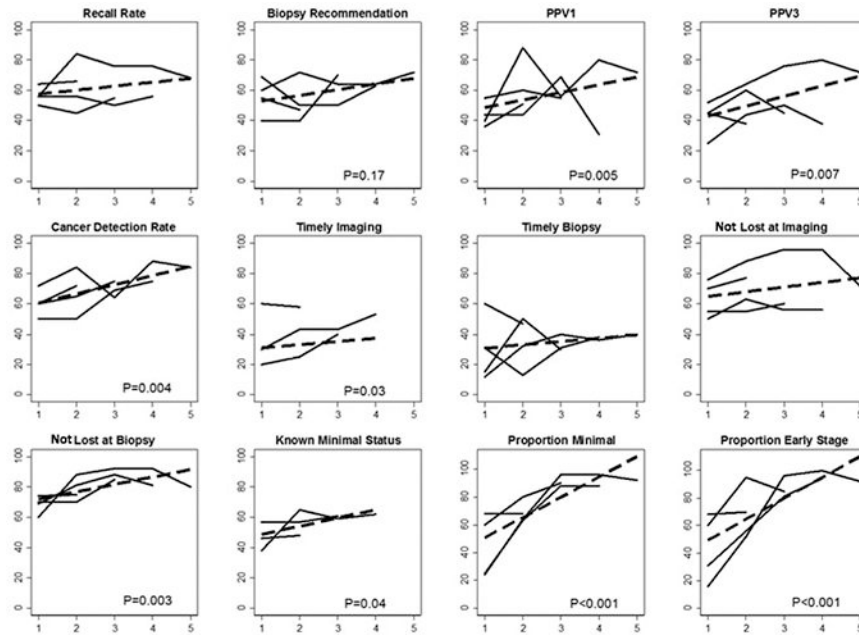
## REFERENCES

1. Rauscher GH, Murphy AM, Orsi JM, Dupuy DM, Grabler PM, Weldon CB. Beyond the mammography quality standards act: Measuring the quality of breast cancer screening programs. *Am J Roentgenol* 2014;202(1):145–51. [PubMed: 24261339]
2. Hirschman J, Whitman S, Ansell D. The black:white disparity in breast cancer mortality: the example of Chicago. *Cancer Causes Control* 2007;18(3):323–33. [PubMed: 17285262]
3. Ansell D, Grabler P, Whitman S, Ferrans C, Burgess-Bishop J, Murray LR, et al. A community effort to reduce the black/white breast cancer mortality disparity in Chicago. *Cancer Causes Control* 2009;20(9):1681–8. [PubMed: 19688184]
4. Sickles EA, Miglioretti DL, Ballard-Barbash R, Geller BM, Leung JWT, Rosenberg RD, et al. Performance Benchmarks for Diagnostic Mammography. *Radiology* 2005 6 1;235(3):775–90. [PubMed: 15914475]
5. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, et al. Performance Benchmarks for Screening Mammography. *Radiology* 2006 10 1;241(1):55–66. [PubMed: 16990671]
6. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DSM, Kerlikowske K, et al. National performance benchmarks for modern screening digital mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology* 2017;283(1).
7. Faguy K Improving Mammography Quality Through EQUIP. *Radiol Technol.* 2019;90(4):369M–385M. [PubMed: 30886033] Hirschman J, Whitman S, Ansell D. The black:white disparity in breast cancer mortality: the example of Chicago. *Cancer Causes Control* 2007;18(3):323–33. [PubMed: 17285262]

Summary statement: Using data collected from the first five years of a robust mammography quality improvement project, we found that participation in the quality improvement program appeared to lead to improvements in patient tracking, callback/detection, and timeliness benchmarks.

**Key Results**

- The number of facilities able to show that they met specific benchmarks increased with length of participation for many but not all measures.
- Trends towards meeting more benchmarks were apparent for cancer detection, timely imaging, not lost at biopsy, known minimal status ( $p < 0.01$  for all), and proportion of screen-detected cancers that were minimal and early stage ( $p < 0.001$  for both).



**Figure 1.** Change in the number of facilities meeting 12 specific benchmarks. Benchmark for known stage at diagnosis not shown (data are in Table 4). Solid lines represent changes in the proportion of of facilities meeting a specific benchmark depending on the number of available time-points for analysis (2,3,4, or 5). The dashed line represents the linear trend in the proportion of of facilities meeting the benchmark over time using all available information. P-values>0.15 are suppressed. Chicago Breast Cancer Quality Consortium (2006-2013).

**Table 1.**

Distribution of facilities by calendar years of data submission.

# Time Points	Calendar Years	Facilities (N=114)	%
2	2009, 2010	1	1
2	2009, 2011	2	2
2	2010, 2011	2	2
2	2011, 2013	48	42
3	2006, 2009, 2011	4	4
3	2009, 2010, 2011	1	1
3	2009, 2011, 2013	5	4
3	2010, 2011, 2013	10	9
4	2006, 2009, 2010, 2013	1	1
4	2006, 2009, 2011, 2013	5	4
4	2006, 2010, 2011, 2013	1	1
4	2009, 2010, 2011, 2013	9	8
5	2006, 2009, 2010, 2011, 2013	25	22
		114	100

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2.**

Change in the proportion of facilities meeting specific benchmarks for call back and cancer detection in the Chicago Breast Cancer Quality Consortium (2006-2013).

Number of time points available for analysis	Recall Rate			Biopsy recommendation Rate			Cancer/Abnormal (PPV1)			Cancer/Biopsied (PPV3)			Screen Detection Rate			Proportion Minimal			Proportion early stage				
	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>d</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>		
2 time points						0.07																	
1	53	64		53	55		53	36		53	45		53	60		53	68		53	68		53	68
2	53	66		53	47		53	51		53	38		53	72		53	68		53	70		53	70
3 time points						0.06																	
1	20	50		20	40		20	55		20	45		20	60		20	60		20	60		20	60
2	20	45		20	40		20	60		20	60		20	65		20	80		20	95		20	95
3	20	55		20	70		20	55		20	45		20	75		20	90		20	85		20	85
4 time points																							
1	16	56		16	69		16	44		16	25		16	50		16	25		16	31		16	31
2	16	56		16	50		16	44		16	44		16	50		16	63		16	56		16	56
3	16	50		16	50		16	69		16	50		16	69		16	88		16	81		16	81
4	16	56		16	63		16	31		16	38		16	75		16	88		16	94		16	94
5 time points						0.03																	
1	25	56		25	60		25	40		25	52		25	72		25	24		25	16		25	16
2	25	84		25	72		25	88		25	64		25	84		25	64		25	52		25	52
3	25	76		25	64		25	56		25	76		25	64		25	96		25	96		25	96
4	25	76		25	64		25	80		25	80		25	88		25	96		25	100		25	100
5	25	68		25	72		25	72		25	72		25	84		25	92		25	92		25	92
Combined						0.17																	
1	114	59		114	55		114	41		114	44		114	61		114	51		114	50		114	50
2	114	65		114	52		114	60		114	48		114	70		114	68		114	68		114	68
3	61	62		61	62		61	59		61	59		61	69		61	92		61	89		61	89
4	41	68		41	63		41	61		41	63		41	83		41	93		41	98		41	98
5	25	68		25	72		25	72		25	72		25	84		25	92		25	92		25	92

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

<sup>b</sup>Number of facilities with data at specified time points.

<sup>c</sup>Percentage of facilities meeting the benchmark at each time point.

<sup>c</sup> P-values from a Wald test for trend via logistic regression with generalized estimating equations to account for clustering by facility. All P-values >0.20 are suppressed. Abbreviations: PPV1, positive predictive value 1; PPV3 positive predictive value 3.

**Table 3.**

Change in the proportion of facilities meeting specific benchmarks for timeliness in the Chicago Breast Cancer Quality Consortium (2006-2013).

Number of time points available for analysis	Timely follow-up imaging			Timely biopsy		
	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>
2 time points						0.12
1	62	60		53	60	
2	62	58		53	47	
3 time points			0.05			
1	20	20		20	15	
2	20	25		20	50	
3	20	40		20	30	
4 time points						
1	30	30	0.01	16	31	
2	30	43		16	13	
3	30	43		16	31	
4	30	53		16	38	
5 time points <sup>d</sup>						0.01
1				25	12	
2				25	32	
3				25	40	
4				25	36	
5				25	40	
Combined			0.03			
1	112	45		114	38	
2	112	48		114	39	
3	50	42		61	34	
4	30	53		41	37	
5				25	40	

<sup>a</sup>Number of facilities with data at specified time points.

<sup>b</sup>Percentage of facilities meeting the benchmark at each time point.

<sup>c</sup>P-values from a Wald test for trend in logistic regression with generalized estimating equations to account for clustering by facility. All P-values >0.20 are suppressed. Data on timely follow-up imaging specific for this benchmark was not collected for CY 2010, therefore no facilities have 5 time points available on this benchmark.

<sup>d</sup>Only 4 time points available for analysis of timely follow-up imaging.



**Table 4.**

Change in the proportion of facilities meeting specific benchmarks pertaining to patient tracking in the Chicago Breast Cancer Quality Consortium (2006-2013).

Number of time points available for analysis	Not Lost at imaging			Not Lost at biopsy			Known minimal status			Known stage at diagnosis		
	N <sup>a</sup>	% <sup>b</sup>	P <sup>d</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>d</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>	N <sup>a</sup>	% <sup>b</sup>	P <sup>c</sup>
2 time points												
1	53	70		53	74		56	46		56	46	
2	53	77		53	75		56	48		56	46	
3 time points												
						0.16						
1	20	55		20	70		23	57		23	48	
2	20	55		20	70		23	57		23	52	
3	20	60		20	85		23	61		23	61	
4 time points												
									0.06			
1	16	50		16	69		34	38		34	38	
2	16	63		16	81		34	65		34	82	
3	16	56		16	88		34	59		34	50	
4	16	56		16	81		34	62		34	41	
5 time points												
						0.02						
1	25	76		25	60							
2	25	88		25	88							
3	25	96		25	92							
4	25	96		25	92							
5	25	72		25	80							
Combined												
						0.003			0.04			
1	114	66		114	69		113	46		113	44	
2	114	74		114	78		113	55		113	58	
3	61	74		61	89		57	60		57	54	
4	41	80		41	88		34	62		34	41	
5	25	72		25	80							

<sup>a</sup>Number of facilities with data at specified time points.

<sup>b</sup>Percentage of facilities meeting the benchmark at each time point.

<sup>c</sup>P-values from a Wald test for trend in logistic regression with generalized estimating equations to account for clustering by facility. All P-values >0.20 are suppressed. Data on known minimal status and known stage were not collected for CY 2006, therefore no facilities have 5 time points available on these benchmarks.