Published in final edited form as:

Cancer. 2014 January 1; 120(1): 103–111. doi:10.1002/cncr.28395.

Breast Cancer Treatment across Healthcare Systems: Linking Electronic Medical Records and State Registry Data to Enable Outcomes Research

Allison W. Kurian, M.D., M.Sc.[1,2], **Aya Mitani, M.S.**[1], **Manisha Desai, Ph.D.**[1], **Peter P. Yu, M.D.**[3], **Tina Seto, M.S.**[1], **Susan C. Weber, Ph.D.**[1], **Cliff Olson, M.S.**[3], **Pragati Kenkare, M.S.**[3], **Scarlett L. Gomez, Ph.D.**[2,4], **Monique A. de Bruin, M.D.**[1], **Kathleen Horst, M.D.**[5], **Jeffrey Belkora, Ph.D.**[3,6], **Suepattra G. May, Ph.D.**[3], **Dominick L. Frosch, Ph.D.**[3,7,8], **Douglas W. Blayney, M.D.**[1], **Harold S. Luft, Ph.D.**[3,6], and **Amar K. Das, M.D., Ph.D.**[1,9]

[1]Department of Medicine, Stanford University

[2]Department of Health Research & Policy, Stanford University

[3]Palo Alto Medical Foundation Research Institute

[4]Cancer Prevention Institute of California

[5]Department of Radiation Oncology, Stanford University

[6]University of California, San Francisco

[7]Department of Medicine, University of California at Los Angeles

[8]Gordon and Betty Moore Foundation

[9]Department of Psychiatry and The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine

Abstract

Background—Understanding of cancer outcomes is limited by data fragmentation. We analyzed the information yielded by integrating breast cancer data from three sources: electronic medical records (EMRs) of two healthcare systems and the state registry.

Methods—We extracted diagnostic test and treatment data from EMRs of all breast cancer patients treated from 2000–2010 in two independent California institutions: a community-based practice (Palo Alto Medical Foundation) and an academic medical center (Stanford University). We incorporated records from the population-based California Cancer Registry (CCR), and then linked EMR-CCR datasets of Community and University patients.

Results—We initially identified 8210 University patients and 5770 Community patients; linked datasets revealed a 16% patient overlap, yielding 12,109 unique patients. The proportion of all Community patients, but not University patients, treated at both institutions increased with

Address correspondence to: Allison W. Kurian, M.D., M.Sc., Assistant Professor of Medicine and of Health Research and Policy, HRP Redwood Building, Room T254A, 150 Governor's Lane, Stanford, CA 94305-5405, Telephone: 650-724-7375, Fax: 650-725-6951, akurian@stanford.edu.

The authors declare no financial disclosures.

worsening cancer prognostic factors. Before linking datasets, Community patients appeared to receive less intervention than University patients (mastectomy: 37.6% versus 43.2%; chemotherapy: 35% versus 41.7%; magnetic resonance imaging (MRI): 10% versus 29.3%; genetic testing: 2.5% versus 9.2%). Linked Community and University datasets revealed that patients treated at both institutions received substantially more intervention (mastectomy: 55.8%; chemotherapy: 47.2%; MRI: 38.9%; genetic testing: 10.9%; $p<0.001$ for each three-way institutional comparison).

Conclusion—Data linkage identified 16% of patients who were treated in two healthcare systems and who, despite comparable prognostic factors, received far more intensive treatment than others. By integrating complementary data from EMRs and population-based registries, we obtained a more comprehensive understanding of breast cancer care and factors that drive treatment utilization.

Keywords

Breast cancer; electronic medical records; bioinformatics; SEER registry; data linkage; outcomes research; comparative effectiveness

INTRODUCTION

Advances in breast cancer diagnosis and treatment[1–4] offer many effective options, and raise questions about the comparative effectiveness of different care pathways.[5–7] National initiatives prioritize comparing the effectiveness of treatments in diverse practice settings,[8–10] requiring demographic and long-term follow-up data from their populations.[11–13] Studies of real-world cancer outcomes, outside of clinical trials, have been limited by the fragmentation and lack of detail in available data. Population-based registries such as the Surveillance, Epidemiology and End Results (SEER) program excel at tracking demographics and incidence, but lack essential details about treatments and diagnostic tests.[14, 15] Institutional electronic medical records (EMR) contain extensive treatment information; however, they are subject to a measurement bias of unknown magnitude, namely the under-reporting of care delivered outside the institution and its outcomes.

Linking EMR-derived data across healthcare systems offers the promise of more complete information, but the challenge of disagreement between institutions, which may require laborious review of patients' charts for resolution. We linked data from the EMRs of an academic medical center and a multi-site community practice in the same catchment region. To provide a gold-standard for patient identification and treatment summaries, we also linked to the statewide population-based California Cancer Registry (CCR, a SEER component).[16] Our hypothesis was that this three-way data linkage would offer a practical and scalable approach to identifying patients treated in more than one healthcare system, and would provide information about variability in cancer care which could not be obtained otherwise.

METHODS

Data Resource Environment

Our project (Oncoshare) began in 2009 to integrate data from EMRs of Stanford University Hospital (SU) and Palo Alto Medical Foundation (PAMF). SU is an academic medical center; PAMF is a multi-site community practice in Alameda, San Mateo, Santa Clara and Santa Cruz counties, California. SU (University) is within one mile of the nearest PAMF (Community) site. Community patients have health maintenance organization (HMO) and fee-for-service insurance; University patients have various insurance plans, including Medicaid. Although inpatient care provided by Community physicians sometimes occurs in University facilities, the institutions are legally and financially separate, with non-overlapping staff. All research was approved by University and Community Institutional Review Boards (IRB) and the State of California IRB (for use of CCR data).

Clinical Data Extraction

We extracted data from University and Community EMRs (Epic, Verona, WI) and from a University warehouse for clinical data collected before Epic implementation in 2007. All University clinical systems data since the mid-1990s reside in the Stanford Translational Research Integrated Database Environment (STRIDE), a warehouse and integration platform for research data extraction and analysis.[17] Real-time electronic data feeds supply clinical information to STRIDE via HL7 technology; extract, transform and load processes out of Epic and into STRIDE occur daily. STRIDE contains one terabyte of data in the form of transcribed dictations and physicians' text notes, billing codes, laboratory and pharmacy orders, medication and radiotherapy administration records, laboratory results, radiology and pathology reports. University chemotherapy data are available from the Epic Beacon provider order entry system since 2008. Community clinical data are housed in three EMR systems: Epic for everything except chemotherapy orders, IDX for billing information, and IntelliDose, an ancillary computer system dedicated to chemotherapy and used since 2000. To ensure uniform coding, chemotherapy data elements in each EMR were mapped to RxNorm,[18] a standardized drug lexicon, and diagnostic test data elements were mapped to National Cancer Institute codes.[19] We identified clinically important interventions, including surgery, chemotherapy, radiation, and emerging diagnostic tests: breast magnetic resonance imaging (MRI), positron emission tomography (PET), and genetic testing for *BRCA1* and *BRCA2* (*BRCA1/2*) mutations. We excluded interventions occurring more than 90 days before cancer diagnosis.

CCR Data Addition

We requested CCR records, with all data fields including age, race/ethnicity, tumor stage, grade, histology, receptors [estrogen receptor (ER), progesterone receptor (PR) and HER2]; and treatment summaries (comprising reports from any California institution of receiving surgery, chemotherapy, and/or radiation) for all breast cancer patients treated at University and/or Community facilities from 2000–2010. Census block groups were geocoded based on patients' residential addresses at the time of diagnoses. The 3% of cases whose address could not be precisely geocoded were assigned to a census block group within their county of residence. We assigned neighborhood socioeconomic status (SES) using a previously

developed and widely used index that incorporates 2000 United States Census data on education, income, occupation and housing costs, based on selection via principal components analysis.[20] We categorized this measure by quintiles based on the distribution of the composite SES index across California. CCR and EMR records were linked using names, social security numbers, medical record numbers and birthdates. All personal identifying information was removed, and clinical encounter dates randomly offset by 30 days, before research use of the data.[21]

Patient Cohort Identification

We defined cohorts representing all patients treated for breast cancer at Community and/or University facilities from January 1, 2000 through January 1, 2010. Eligible patients were female, ≥18 years old, and met at least one of the following criteria within the period: 1) the CCR reported a breast cancer diagnosis and/or treatment at Community and/or University facilities; 2) University and/or Community billing records included a diagnostic code for breast cancer or ductal carcinoma in situ [International Classification of Diseases-9 (ICD-9) codes 174.9 or 233.0], billed by a breast cancer specialist (defined as a surgeon, medical oncologist or radiation oncologist). Treating institution was based on clinician affiliation, not location; a Community surgeon operating at the University was coded as Community. Institution was determined first by EMR-based billing records: patients who had University records of breast cancer-specific interventions (surgery, chemotherapy, radiation) were coded as University, and likewise for Community, as confirmed by the CCR. For patients lacking treatment records, institution was defined by billing records for cancer-related diagnostic tests including PET and genetic testing, and if there were no such records, by presence in University or Community internal tumor registries, which report to the CCR. MRI was not used to determine treating institution because before 2006 some Community patients visited the University for MRI only. After generating separate University and Community cohorts (defined hereafter as “EMR-CCR cohorts”), we linked these two EMR-CCR cohorts to identify patients treated at both institutions.

Quality Assurance and Analytical Cohort Development

We validated and applied an algorithm to link records across data sources.[21, 22] To ensure subjects’ eligibility, we developed analytical cohorts, from which we excluded patients lacking data on all of the following (considered essential for analyzing breast cancer care): stage, tumor receptors (ER, PR, HER2), and any diagnostic or treatment intervention. We applied more stringent inclusion criteria for patients identified in EMRs only but not in the CCR, because review of physicians’ notes and pathology reports in EMRs revealed that many such patients had received breast cancer ICD-9 codes erroneously, often coincident with prophylactic mastectomy or tamoxifen used for breast cancer risk reduction. These stringent inclusion criteria were cancer-specific pathology data (stage and/or tumor receptors) and treatments (chemotherapy and/or radiation). This algorithm was applied within each institution before linking EMR-CCR cohorts, and to the overall cohort after linkage.

Statistical Analysis

Patient characteristics, receipt of treatments and diagnostic tests were tabulated before and after linkage of University and Community EMR-CCR cohorts. After linkage, measures for patients treated at University, Community, and both institutions were compared using the Chi-squared statistic. All p values were two-sided.

RESULTS

Analytical Cohorts

We identified a maximally inclusive University cohort of 8892 patients. Applying our eligibility criteria left 8210 patients (92.3%) in the University analytical cohort. Repeating these steps, we identified a maximally inclusive Community cohort of 6304 patients, and retained 5770 (91.5%) in the Community analytical cohort; adding these cohorts produced an apparent total of 13,980 patients. Linked records from the University and Community EMR-CCR cohorts yielded a maximally inclusive cohort of 13,238 unique patients, of whom we retained 12,109 (91.5%) in the Combined analytical cohort (Figure 1a–c).

Patient Characteristics, Before and After EMR-CCR Cohort Linkage

Before linking University and Community EMR-CCR cohorts, University patients appeared younger, with lower SES and worse cancer prognostic factors than Community patients (Table 1). Linked EMR-CCR cohorts identified a third group of patients who were treated at both institutions (defined hereafter as “Both”). “Both” patients were significantly more likely to be Asian (University-only 14%, Community-only 13.9%, “Both” 17.2%) and of highest-quintile SES (University-only 49.2%, Community-only 64.6%, “Both” 75.2%). “Both” patients had intermediate prognostic factors, including age (<40 years: University-only 10.9%, Community-only 3.7%, “Both” 10%), stage (III or IV: University-only 13.6%, Community-only 6.8%, “Both” 10.2%), tumor receptor subtype (for the poor prognosis subtypes,[23] HER2-positive or ER-, PR- and HER2-negative: University-only 29.1%, Community-only 14.5%, “Both” 25.9%), and grade (3: University-only 32.3%, Community-only 19.8%, “Both” 29.5%; $p<0.001$ for each reported three-way comparison). As prognostic factors worsened, including decreasing age, increasing stage, increasing grade, and less favorable receptor subtype,[24–26] an increasing proportion of Community patients (but not University patients) fell into the “Both” category.

Treatments and Diagnostic Tests, Before and After EMR-CCR Cohort Linkage

Treatment information was most often available from the CCR, but diagnostic test information was available only from EMRs, through providers’ notes and billing (Table 2). For example, CCR data identified about 95% of all women with evidence from any source of having received mastectomy, but institution-specific data identified only 25–50% of these cases. For women in the “Both” category, the “institution-specific” data performed better, reflecting a greater yield from combining EMR-derived data from two institutions. For chemotherapy, Community billing data offered somewhat more complete case finding than that from the University. Linked University and Community EMR-CCR cohorts revealed that the usage of all interventions was highest among the “Both” patients. For example,

mastectomy utilization was as follows: University-only 39.7%, Community-only 30.5%, "Both" 55.8%, and similarly for bilateral mastectomy: University-only 8%, Community-only 5.2%, "Both" 13.2%. Figure 2 illustrates another example: the differential use of MRI among University-only (32.9%), Community-only (32.8%), and "Both" (66%) patients by 2009 ($p<0.001$ for each three-way comparison).

DISCUSSION

To study breast cancer care beyond the walls of a single institution, we linked state registry records to data extracted from the EMRs of two healthcare systems, one community-based and one university-affiliated. This three-way data linkage generated unique insights. We found a 16% patient overlap between nearby healthcare systems, which enables an estimate of the magnitude of missing treatment information in single-institution studies. We discovered a striking care pattern, with Community patients increasingly likely to be treated at both institutions as cancer prognosis worsened, and with "Both" patients receiving the most intensive intervention despite having intermediate cancer prognostic factors. These findings illustrate how efforts to compare outcomes across real-world settings must account for measured and unmeasured risk factors and patient preferences.

Previous studies have integrated complementary databases, supplementing SEER-derived data with treatment details from Medicare claims[27, 28] and HMOs.[29, 30] This study's novelty lies in linking data from the EMRs of nearby yet independent healthcare systems, anchored by data from the CCR, a SEER component. We assessed data quality by reviewing several hundred de-identified patient records, and evaluating agreement between all sources; rare conflicts were adjudicated by physician review.[21, 22] The three-way linkage identified the most informative source for each variable, with the CCR most informative about treatment utilization, and EMRs the only source of diagnostic test data. Missing data were reduced by the three-way linkage, with "Both" patients having the most data available.

We encountered limitations in extracting research data from EMRs. We extracted structured data from billing, drug ordering and administration records, and performed simple natural language processing of diagnostic reports, but many important concepts remain buried in the unstructured paragraphs of clinicians' notes. These include nuances of decision-making which lack representation elsewhere, notably physicians' recommendations and patients' preferences. EMRs also promise a wealth of clinical detail that cannot be obtained from administrative databases or registries, including the images and reports of radiologic exams and genomic sequencing tests. Some of this information can be extracted and encoded as discrete data elements (for example, BI-RADS scores for mammogram and breast MRI), whereas identifying the determinants of treatment choices may require advances in natural language processing. The accurate retrieval of such specific patient information from unstructured, free-text EMR notes remains an active area of research.[31, 32] Given the EMR's unique potential to enhance understanding of cancer outcomes, studies to optimize the clinical and research uses of EMRs should remain a high priority.[33, 34] Some limitations may be addressed through EMR changes, with structured fields facilitating data extraction; others require new data sources, including patient-reported information.[8, 35] Bridging such gaps should be a priority of emerging data integration initiatives.[36, 37] Health information

technology is developing rapidly, and the decade of 2000–2010 witnessed the implementation of EMRs and complementary databases. EMR modules for clinical data exchange between University and Community (Care Everywhere: Epic, Verona WI) and between patients and physicians (Patient Portal: Epic, Verona WI) were activated in 2012, and should enhance both clinical care and research. In the future, standardized data representation models will facilitate the interoperability of digital health data between institutions.

The “Both” patients offer an intriguing glimpse across healthcare systems. This category comprised 16% of patients, disproportionately representing top-quintile SES and intermediate cancer prognostic factors. Without information about physician referrals and patient preferences, we do not know why patients accessed both systems, but the over-representation of sicker Community patients in the “Both” category suggests tertiary center consultation on challenging cases. The “Both” patients are remarkable for their significantly greater utilization of every intervention studied, including mastectomy, chemotherapy, radiation, MRI, PET, and genetic testing. One explanation might be that “University-only” and “Community-only” patients actually accessed other healthcare systems, leading us to underestimate their test use; however, such potential under-ascertainment cannot explain treatment differences recorded in the CCR, which aggregates statewide cancer data comprehensively because of mandated reporting. Previous studies reported rising mastectomy rates,[38–42] despite a lack of survival benefit,[4, 43, 44] and found correlations with an increase in diagnostic testing.[39, 45, 46] The “Both” patients’ high SES might explain their greater use of interventions which are usually optional, such as MRI and bilateral mastectomy,[25, 47–50] but we lack information about other factors that may drive utilization, including family cancer history and clinical trial participation. Assessing the value added by specific interventions[51–53] will require a deeper understanding of the patient, physician and healthcare factors that shape the care patterns we observed.

Integrating breast cancer data from two EMRs and the state registry proved feasible and informative, broadening our understanding of care beyond what could be achieved from just one or two data sources. This approach offers insight about real-world treatment across healthcare systems, which can advance comparative effectiveness and outcomes research in oncology.

ACKNOWLEGMENTS

The authors thank Robert W. Carlson, M.D., Wei-Nchih Lee, M.D., Ph.D., and Sandra Wilson, Ph.D. for their critical review of this manuscript.

Funding: Susan and Richard Levy Gift Fund; Regents of the University of California’s California Breast Cancer Research Program (#16OB-0149); Stanford University Developmental Research Fund; and the National Cancer Institute’s Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California. The collection of cancer incidence data used in this study was supported by the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute’s Surveillance, Epidemiology, and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention’s National Program of Cancer Registries, under agreement #1U58 DP000807-01 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the authors, and endorsement by the University or State of California, the California Department of Health Services, the National Cancer Institute, or the

Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred.

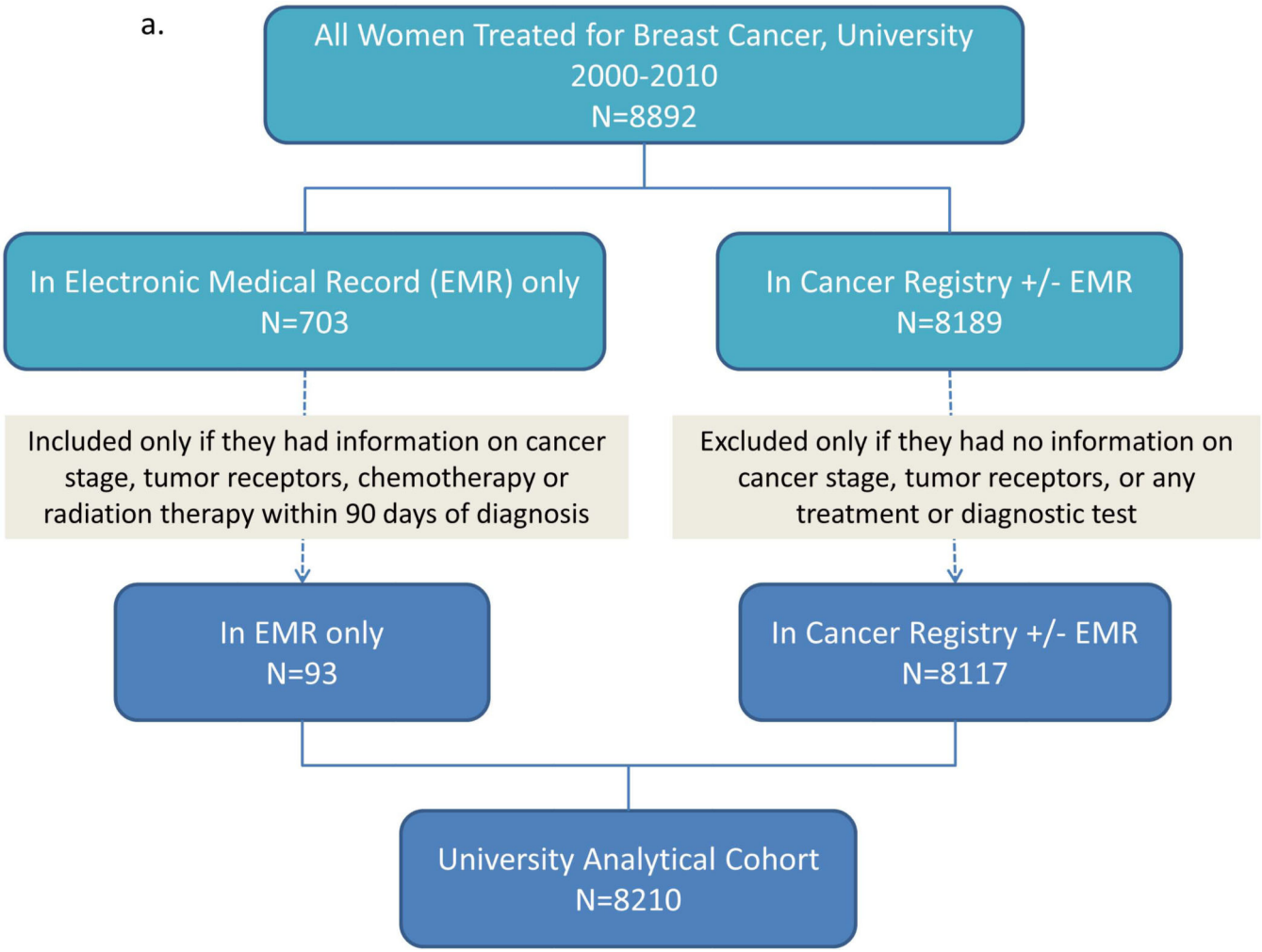
REFERENCES

1. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet. 2000; 355:1757–1770. [PubMed: 10832826]
2. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005; 365:1687–1717. [PubMed: 15894097]
3. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med. 2005; 353:1784–1792. [PubMed: 16251534]
4. Veronesi U, Cascinelli N, Mariani L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med. 2002; 347:1227–1232. [PubMed: 12393819]
5. Katz SJ, Lantz PM, Janz NK, et al. Patient involvement in surgery treatment decisions for breast cancer. J Clin Oncol. 2005; 23:5526–5533. [PubMed: 16110013]
6. Katz SJ, Morrow M. The challenge of individualizing treatments for patients with breast cancer. JAMA. 2012; 307:1379–1380. [PubMed: 22474200]
7. Morrow M, Jagsi R, Alderman AK, et al. Surgeon recommendations and receipt of mastectomy for treatment of breast cancer. JAMA. 2009; 302:1551–1556. [PubMed: 19826024]
8. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. JAMA. 2012; 307:1583–1584. [PubMed: 22511682]
9. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med. 2009; 151(3):203–205. [PubMed: 19567618]
10. VanLare JM, Conway PH, Sox HC. Five next steps for a new national program for comparative-effectiveness research. N Engl J Med. 2010; 362:970–973. [PubMed: 20164480]
11. Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. JAMA. 2012; 307:1636–1640. [PubMed: 22511692]
12. Hershman DL, Wright JD. Comparative effectiveness research in oncology methodology: observational data. J Clin Oncol. 2012; 30:4215–4222. [PubMed: 23071228]
13. Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. J Clin Oncol. 2012; 30:4243–4248. [PubMed: 23071233]
14. Bickell NA, McAlearney AS, Wellner J, Fei K, Franco R. Understanding the challenges of adjuvant treatment measurement and reporting in breast cancer: cancer treatment measuring and reporting. Med Care; electronic publication ahead of print. 2011
15. Lodrigues W, Dumas J, Rao M, Lilley L, Rao R. Compliance with the commission on cancer quality of breast cancer care measures: self-evaluation advised. Breast J. 2011; 17:167–171. [PubMed: 21294806]
16. California Cancer Registry. Available from URL: http://www.ccrcal.org.
17. Lowe, HJ.; Ferris, TA.; Hernandez, PM.; Weber, SC. STRIDE--An integrated standards-based translational research informatics platform; AMIA Annu Symp Proc; 2009. p. 391-395.
18. Hernandez, PN.; Podchiyska, T.; Weber, SC.; Ferris, TA.; Lowe, HJ. Automated mapping of pharmacy orders from two electronic health record systems to RxNorm within the STRIDE clinical data warehouse; AMIA Annual Symposium; 2009.
19. [accessed June 7, 2013] National Cancer Institute Enterprise Vocabulary Services. Available from URL: http://evs.nci.nih.gov/
20. Yost K, Perkins C, Cohen R, Morris C, Wright W. Socioeconomic status and breast cancer incidence in California for different race/ethnic groups. Cancer Causes Control. 2001; 12(8):703–711. [PubMed: 11562110]
21. Weber SC, Lowe H, Das A, Ferris T. A simple heuristic for blindfolded record linkage. J Am Med Inform Assoc. 2012; 19:e157–e161. [PubMed: 22298567]

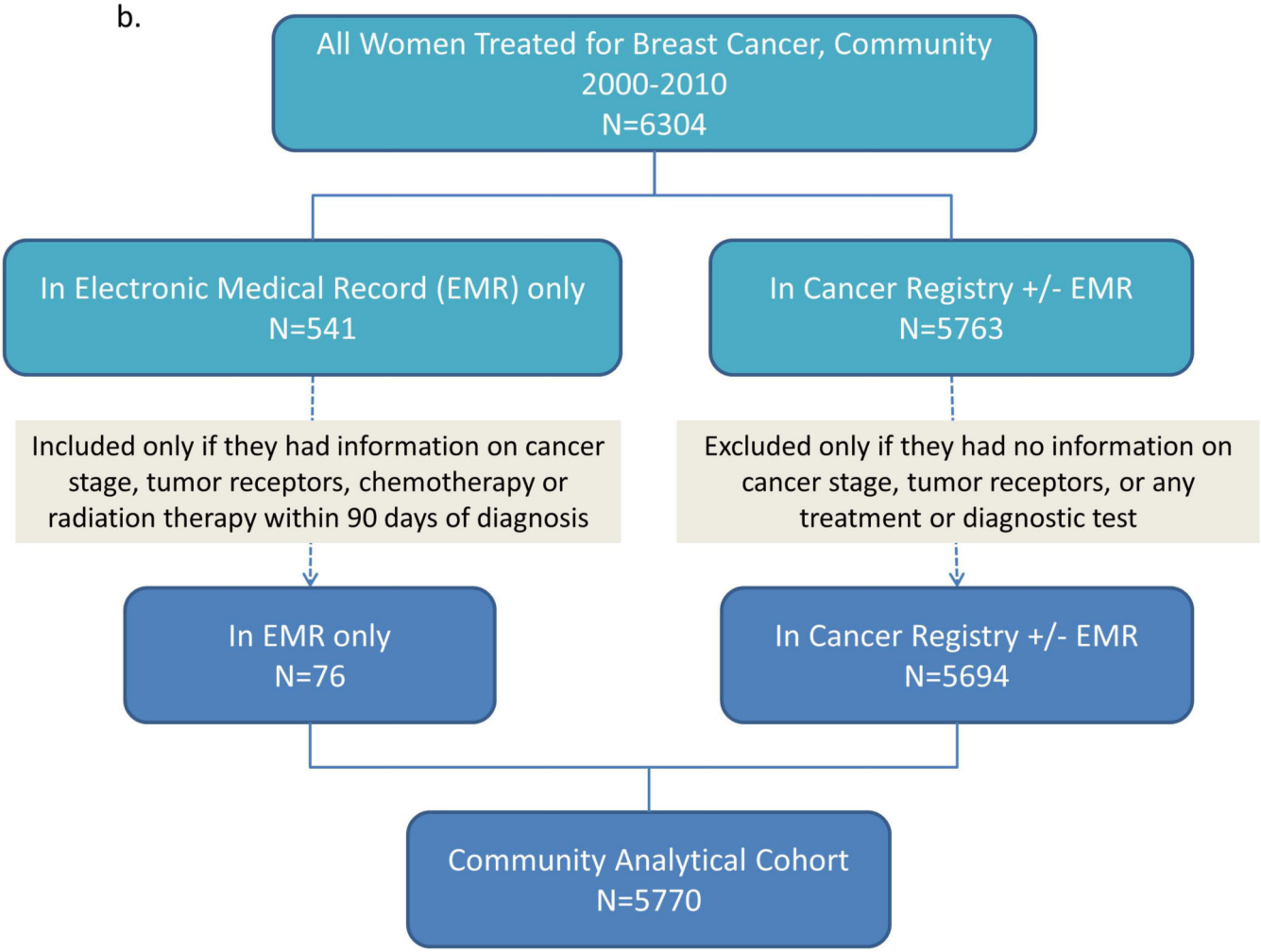
22. Weber, SC.; Seto, T.; Olson, C.; Kenkare, P.; Kurian, AW.; Das, AK. Oncoshare: lessons learned from building an integrated multi-institutional database for comparative effectiveness research; AMIA Annu Symp Proc; 2012. p. 970-978.
23. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006; 295:2492–2502. [PubMed: 16757721]
24. Polychemotherapy for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet. 1998; 352:930–942. [PubMed: 9752815]
25. Carlson RW, Allred DC, Anderson BO, et al. Invasive breast cancer. J Natl Compr Canc Netw. 2011; 9:136–222. [PubMed: 21310842]
26. Darby S, McGale P, Correa C, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet. 2011; 378:1707–1716. [PubMed: 22019144]
27. Guadagnolo BA, Liao KP, Elting L, Giordano S, Buchholz TA, Shih YC. Use of radiation therapy in the last 30 days of life among a large population-based cohort of elderly patients in the United States. J Clin Oncol. 2013; 31:80–87. [PubMed: 23169520]
28. Snyder CF, Frick KD, Herbert RJ, et al. Quality of care for comorbid conditions during the transition to survivorship: differences between cancer survivors and noncancer controls. J Clin Oncol. 2013; 31:1140–1148. [PubMed: 23401438]
29. Hershman DL, Kushi LH, Shao T, et al. Early discontinuation and nonadherence to adjuvant hormonal therapy in a cohort of 8,769 early-stage breast cancer patients. J Clin Oncol. 2010; 28:4120–4128. [PubMed: 20585090]
30. Kurian AW, Lichtensztajn DY, Keegan TH, et al. Patterns and predictors of breast cancer chemotherapy use in Kaiser Permanente Northern California, 2004–2007. Breast Cancer Res Treat. 2013; 137:247–260. [PubMed: 23139057]
31. Edinger, T.; Cohen, AM.; Bedrick, S.; Ambert, K.; Hersh, W. Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC Medical Records Track; AMIA Annu Symp Proc; 2012. p. 180-188.
32. Ohno-Machado L. Realizing the full potential of electronic health records: the role of natural language processing. J Am Med Inform Assoc. 2011; 18:539. [PubMed: 21846784]
33. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical Care. 2013; 51:S30–S37. [PubMed: 23774517]
34. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc. 2013; 20:144–151. [PubMed: 22733976]
35. Basch E, Abernethy AP, Mullins CD, et al. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. J Clin Oncol. 2012; 30:4249–4255. [PubMed: 23071244]
36. Stewart AK, McNamara E, Gay EG, Banasiak J, Winchester DP. The Rapid Quality Reporting System--a new quality of care tool for CoC-accredited cancer programs. J Registry Manag. 2011; 38:61–63. [PubMed: 22097705]
37. Kurian AW, Edge SB. Information technology interventions to improve cancer care quality: a report from the American Society of Clinical Oncology Quality Care Symposium. Journal of Oncology Practice. 2013; 9:142–144. [PubMed: 23942493]
38. Gomez SL, Lichtensztajn D, Kurian AW, et al. Increasing mastectomy rates for early-stage breast cancer? Population-based trends from California. J Clin Oncol. 2010; 28:e155–e157. [PubMed: 20159812]
39. Katipamula R, Degnim AC, Hoskin T, et al. Trends in mastectomy rates at the Mayo Clinic Rochester: effect of surgical year and preoperative magnetic resonance imaging. J Clin Oncol. 2009; 27:4082–4088. [PubMed: 19636020]
40. Tuttle TM, Habermann EB, Grund EH, Morris TJ, Virnig BA. Increasing use of contralateral prophylactic mastectomy for breast cancer patients: a trend toward more aggressive surgical treatment. J Clin Oncol. 2007; 25:5203–5209. [PubMed: 17954711]

41. Tuttle TM, Jarosek S, Habermann EB, et al. Increasing rates of contralateral prophylactic mastectomy among patients with ductal carcinoma in situ. J Clin Oncol. 2009; 27:1362–1367. [PubMed: 19224844]
42. Collins ED, Moore CP, Clay KF, et al. Can women with early-stage breast cancer make an informed decision for mastectomy? J Clin Oncol. 2009; 27:519–525. [PubMed: 19114703]
43. Hwang ES, Lichtensztajn DY, Gomez SL, Fowble B, Clarke CA. Survival after lumpectomy and mastectomy for early stage invasive breast cancer: The effect of age and hormone receptor status. Cancer. 2013; 119:1402–1411. [PubMed: 23359049]
44. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med. 2002; 347:1233–1241. [PubMed: 12393820]
45. Tuttle TM. Magnetic resonance imaging and contralateral prophylactic mastectomy: the "no mas" effect? Ann Surg Oncol. 2009; 16:1461–1462. [PubMed: 19290489]
46. King TA, Sakr R, Patil S, et al. Clinical management factors contribute to the decision for contralateral prophylactic mastectomy. J Clin Oncol. 2011; 29:2158–2164. [PubMed: 21464413]
47. Bedrosian I, Hu CY, Chang GJ. Population-based study of contralateral prophylactic mastectomy and survival outcomes of breast cancer patients. J Natl Cancer Inst. 2010; 102:401–409. [PubMed: 20185801]
48. Daly MB, Axilbund JE, Buys S, et al. Genetic/familial high-risk assessment: breast and ovarian. J Natl Compr Canc Netw. 2010; 8:562–594. [PubMed: 20495085]
49. Mainiero MB, Lourenco A, Mahoney MC, et al. ACR Appropriateness Criteria Breast Cancer Screening. J Am Coll Radiol. 2013; 10:11–14. [PubMed: 23290667]
50. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin. 2007; 57:75–89. [PubMed: 17392385]
51. Blayney DW, McNiff K, Hanauer D, Miela G, Markstrom D, Neuss M. Implementation of the Quality Oncology Practice Initiative at a university comprehensive cancer center. J Clin Oncol. 2009; 27:3802–3807. [PubMed: 19487377]
52. Neuss MN, Desch CE, McNiff KK, et al. A process for measuring the quality of cancer care: the Quality Oncology Practice Initiative. J Clin Oncol. 2005; 23:6233–6239. [PubMed: 16087948]
53. Schnipper LE, Smith TJ, Raghavan D, et al. American Society of Clinical Oncology identifies five key opportunities to improve care and reduce costs: the top five list for oncology. J Clin Oncol. 2012; 30:1715–1724. [PubMed: 22493340]

a.



b.



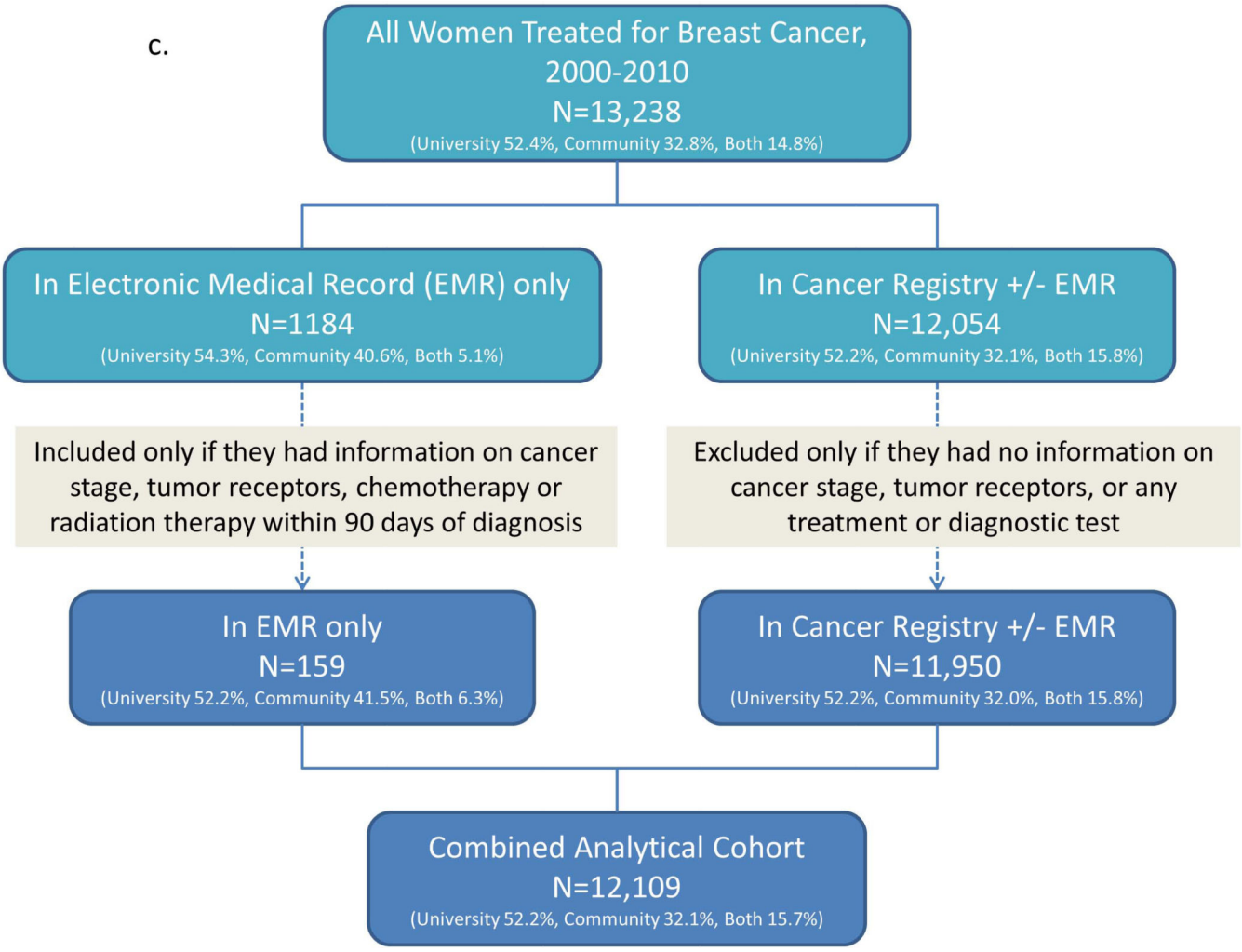


Figure 1.

Patient identification and inclusion in analytical cohorts for **a)** University, **b)** Community, and **c)** University and Community combined.

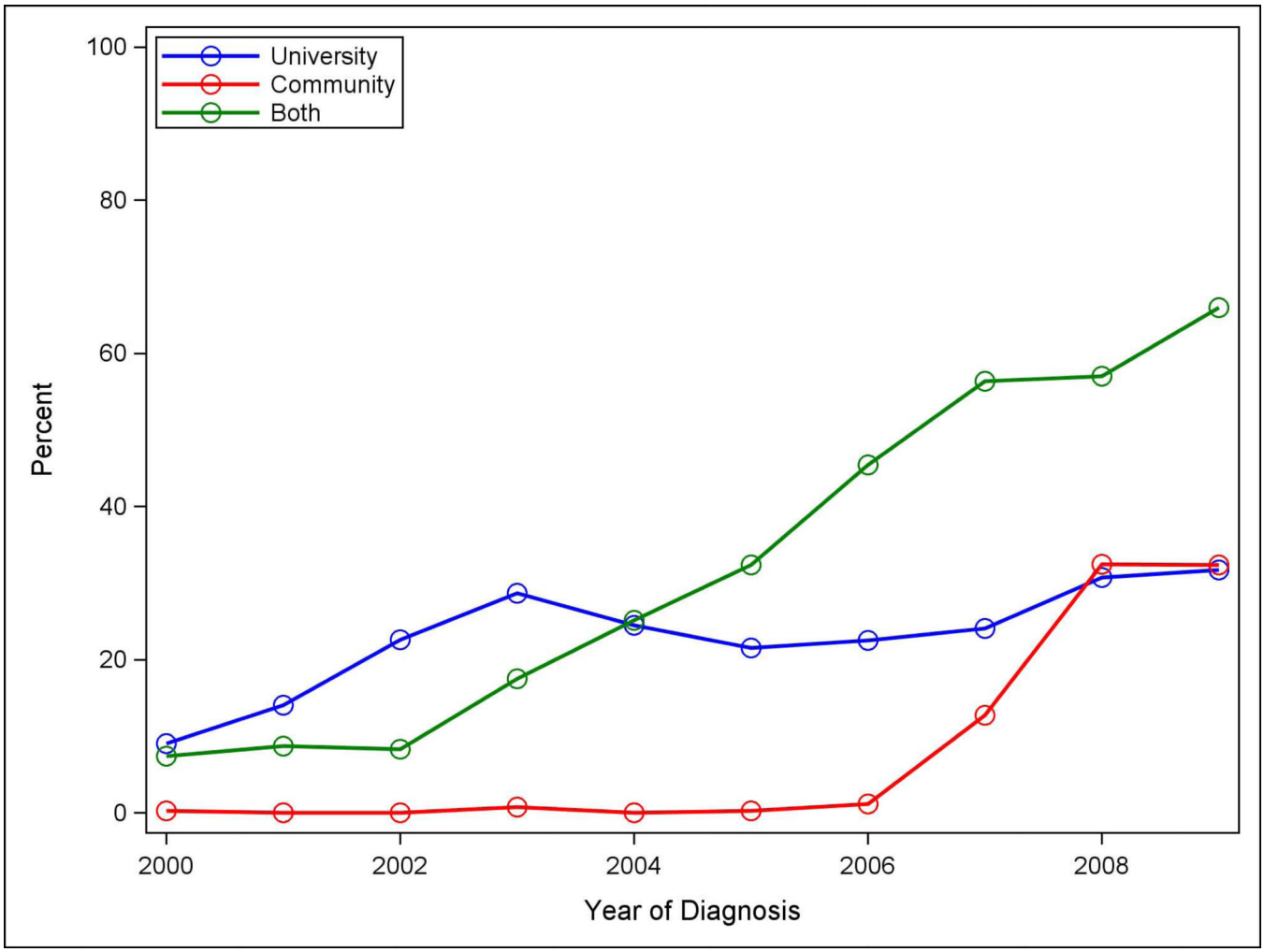


Figure 2.
Use of breast magnetic resonance imaging by year and treating institution: University, Community, and Both.

Table 1

Patient characteristics, ascertained before and after linking University and Community electronic medical record (EMR) data

	Before Linking Data				After Linking Data							
	University		Community		University-only		Community-only		"Both"		Proportion in "Both"	
	N	%	N	%	N	%	N	%	N	%	University	Community
Total	8210	100%	5770	100%	6321	52.2%	3886	32.1%	1902	15.7%	23.1%	32.9%
Age at Diagnosis, years[a]												
<40	880	10.7%	332	5.8%	689	10.9%	142	3.7%	191	10.0%	21.7%	57.4%
40–49	2247	27.4%	1289	22.3%	1717	27.2%	758	19.5%	534	28.1%	23.7%	41.3%
50–64	3211	39.1%	2211	38.3%	2493	39.4%	1498	38.5%	723	38.0%	22.5%	32.6%
≥65	1872	22.8%	1938	33.6%	1422	22.5%	1488	38.3%	454	23.9%	24.2%	23.4%
Year of Breast Cancer Diagnosis[a]												
2000–2003	3003	36.6%	2439	42.3%	2279	36.1%	1721	44.3%	733	38.5%	24.3%	29.9%
2004–2006	2780	33.9%	1672	29.0%	2121	33.6%	1012	26.0%	662	34.8%	23.8%	39.5%
2007–2009	2427	29.6%	1659	28.8%	1921	30.4%	1153	29.7%	507	26.7%	20.9%	30.5%
Race[a]												
Missing	164	2.0%	76	1.3%	156	2.5%	68	1.7%	8	0.4%	4.9%	10.5%
White	6495	79.1%	4714	81.7%	4978	78.8%	3201	82.4%	1525	80.2%	23.5%	32.3%
Black	251	3.1%	82	1.4%	218	3.4%	49	1.3%	34	1.8%	13.5%	41%
Asian	1208	14.7%	862	14.9%	882	14%	539	13.9%	328	17.2%	27.1%	37.8%
Other	92	1.1%	36	0.6%	87	1.4%	29	0.7%	7	0.4%	7.4%	19.4%
Ethnicity[a]												
Missing	155	1.9%	96	1.7%	148	2.3%	89	2.3%	7	0.4%	4.5%	7.3%
Non-Hispanic	7541	91.9%	5430	94.1%	5712	90.4%	3607	92.8%	1841	96.8%	24.4%	33.8%
Hispanic	514	6.3%	244	4.2%	461	7.3%	190	4.9%	54	2.8%	10.5%	22.1%
Socioeconomic Status[a]												
Missing	323	3.9%	339	5.9%	295	4.7%	311	8.0%	28	1.5%	8.7%	8.3%
Lowest quintile	299	3.6%	37	0.6%	293	4.6%	31	0.8%	6	0.3%	2.0%	16.2%
Second quintile	652	7.9%	162	2.8%	603	9.5%	112	2.9%	51	2.7%	7.8%	31.3%
Third quintile	916	11.2%	300	5.2%	825	13.1%	207	5.3%	93	4.9%	10.1%	31.0%

	Before Linking Data				After Linking Data							
	University		Community		University-only		Community-only		"Both"		Proportion in "Both"	
	N	%	N	%	N	%	N	%	N	%	University	Community
Fourth quintile	1487	18.1%	1002	17.4%	1193	18.9%	714	18.4%	294	15.5%	19.8%	29.2%
Highest quintile	4533	55.2%	3930	68.1%	3112	49.2%	2511	64.6%	1430	75.2%	31.5%	36.3%
Stage[a]												
Missing	554	6.7%	529	9.2%	453	7.2%	433	11.1%	114	6%	20.1%	20.8%
Stage 0	1581	19.3%	1077	18.7%	1214	19.2%	710	18.3%	367	19.3%	23.2%	34.1%
Stage I	2536	30.9%	2050	35.5%	1908	30.2%	1422	36.6%	628	33%	24.8%	30.6%
Stage II	2489	30.3%	1657	28.7%	1890	29.9%	1058	27.2%	599	31.5%	24.1%	36.1%
Stage III	721	8.8%	349	6%	574	9.1%	202	5.2%	147	7.7%	20.4%	42.1%
Stage IV	329	4%	108	1.9%	282	4.5%	61	1.6%	47	2.5%	14.3%	43.5%
Tumor Receptor Subtype (Stages I–IV)[a]												
Missing data for any receptor	2349	32.1%	2300	44%	1346	26.4%	1450	45.7%	334	21.8%	19.9%	18.7%
HR-positive, HER2-negative[b]	2070	42.1%	2070	39.6%	2275	44.6%	1266	39.9%	804	52.4%	26.1%	38.8%
HER2-positive[b]	565	15.7%	565	10.8%	889	17.4%	304	9.6%	261	17%	22.7%	46.2%
HR- and HER2-negative (triple-negative)[b]	292	10%	292	5.6%	597	11.7%	156	4.9%	136	8.9%	18.6%	46.6%
Grade[a]												
Missing	1330	16.2%	1589	27.5%	1070	16.9%	1329	34.2%	273	14.4%	20.3%	17%
1	1365	16.6%	917	15.9%	1039	16.4%	591	15.2%	326	17.1%	23.9%	35.6%
2	2915	35.5%	1934	33.5%	2173	34.4%	1195	30.8%	742	39%	25.5%	38.3%
3	2600	31.7%	1330	23.1%	2039	32.3%	771	19.8%	561	29.5%	21.6%	42.1%
Histology[a]												
Missing	47	0.6%	70	1.2%	42	0.7%	65	1.7%	5	0.3%	10.6%	7.1%
Ductal	6613	80.5%	4696	81.4%	5059	80%	3145	80.9%	1566	82.3%	23.6%	33.2%
Lobular	733	8.9%	525	9.1%	537	8.5%	329	8.5%	197	10.4%	26.8%	37.5%
Other	817	10%	479	8.3%	683	10.8%	347	8.9%	134	7%	16.4%	27.9%

[a] p value using Chi-square statistic <0.001, for comparison between University, Community and Both patients after EMR data linkage

[b] HR: hormone receptor (estrogen and progesterone receptors, ER and PR). HR-positive tumors have ER and/or PR positive; HR-negative tumors have ER and PR both negative. Receptor subtype is not available for Stage 0, because HER2 was not tested.

Table 2

Diagnostic test and treatment utilization, ascertained before and after linking University and Community electronic medical record (EMR) data

	Before University-Community EMR Data Linkage				After University-Community EMR Data Linkage					
	University N (%)	**Users identified by data source N (%)**	**Community N (%)**	**Users identified by data source N (%)**	**University-only N (%)**	**Users identified by data source N (%)**	**Community-only N (%)**	**Users identified by data source N (%)**	**"Both" N (%)**	**Users identified by data source N (%)**
Total	8210		5770		6321		3886		1902	
Mastectomy[a]	3545 (43.2%)		2172 (37.6%)		2510 (39.7%)		1187 (30.5%)		1062 (55.8%)	
EMR: physician billing records		904 (25.5%)		821 (37.8%)		732 (29.2%)		409 (34.5%)		581 (54.7%)
EMR: facility billing records		1845 (52%)		1000 (46%)		1115 (44.4%)		499 (42%)		904 (85.1%)
California Cancer Registry (CCR)		3367 (95%)		2076 (95.6%)		2390 (95.2%)		1137 (95.8%)		983 (92.6%)
Unilateral Mastectomy[b]	2615 (31.9%)		1637 (28.4%)		1887 (29.9%)		935 (24.1%)		731 (38.4%)	
Bilateral Mastectomy[b]	752 (9.2%)		439 (7.6%)		503 (8%)		202 (5.2%)		252 (13.2%)	
Chemotherapy[a]	3426 (41.7%)		2021 (35%)		2624 (41.5%)		1169 (30.1%)		897 (47.2%)	
EMR: facility billing records		133 (3.9%)		404 (20%)		114 (4.3%)		229 (19.6%)		188 (21%)
EMR: drug administration records		822 (24%)		1115 (55.2%)		659 (25.1%)		662 (56.6%)		596 (66.4%)
CCR		3235 (94.4%)		1707 (84.5%)		2468 (94.1%)		951 (81.4%)		778 (86.7%)
Radiation Therapy[a]	4284 (52.2%)		2661 (46.1%)		3340 (52.8%)		1748 (45%)		1028 (54%)	
EMR: facility billing records		2022 (47.2%)		1468 (55.2%)		1653 (49.5%)		1008 (57.7%)		802 (78%)
CCR		3845 (89.8%)		2377 (89.3%)		2972 (89%)		1556 (89%)		877 (85.3%)
Magnetic Resonance Imaging[a, c]	2402 (29.3%)		576 (10%)		1777 (28.1%)		414 (10.7%)		740 (38.9%)	

	Before University-Community EMR Data Linkage				After University-Community EMR Data Linkage					
	University N (%)	**Users identified by data source N (%)**	**Community N (%)**	**Users identified by data source N (%)**	**University-only N (%)**	**Users identified by data source N (%)**	**Community-only N (%)**	**Users identified by data source N (%)**	**"Both" N (%)**	**Users identified by data source N (%)**
Diagnostic (<1 year from diagnosis)	1944 (23.7%)		412 (7.1%)		1438 (22.7%)		306 (7.9%)		601 (31.5%)	
Screening (>1 year from diagnosis)	930 (11.3%)		217 (3.8%)		692 (10.9%)		147 (3.8%)		299 (15.7%)	
Positron Emission Tomography[a, c]	440 (5.4%)		296 (5.1%)		353 (5.6%)		163 (4.2%)		216 (11.4%)	
***BRCA1/2* Genetic Testing**[a, c]	755 (9.2%)		145 (2.5%)		585 (9.3%)		101 (2.6%)		208 (10.9%)	

[a] p value <0.001 for comparison between University-only, Community-only, and “Both” patients after EMR data linkage

[b] Available from CCR only

[c] Available from EMR only