Published in final edited form as:

Eval Program Plann. 2021 October ; 88: 101967. doi:10.1016/j.evalprogplan.2021.101967.

Utility of linking survey and registry data to evaluate interventions and policies to address disparities in breast cancer survivorship among young women

Sujha Subramanian[a,*], **Madeleine Jones**[a], **Florence K.L. Tangka**[b], **Patrick Edwards**[a], **Tim Flanigan**[a], **Jenya Kaganova**[a], **Kevin Smith**[a], **Temeika Fairley**[b], **Nikki A. Hawkins**[b], **Juan L. Rodriguez**[b], **Gery P. Guy Jr.**[b], **Cheryll C. Thomas**[b]

[a]RTI International, 307 Waverly Oaks Road, Waltham, MA, 0245, USA

[b]Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta, GA, USA

Abstract

Purpose: There is limited research linking data sources to evaluate the multifactorial impacts on the quality of treatment received and financial burden among young women with breast cancer. To address this gap and support future evaluation efforts, we examined the utility of combining patient survey and cancer registry data.

Patient and Methods: We administered a survey to women, aged 18–39 years, with breast cancer from four U.S. states. We conducted a systematic response-rate analysis and evaluated

---

[*] Corresponding author at: RTI International, 307 Waverly Oaks Road, Waltham, MA, 0245, USA, ssubramanian@rti.org (S. Subramanian).

CRediT authorship contribution statement

Sujha Subramanian: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration. **Madeleine Jones:** Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Florence K.L. Tangka:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - review & editing, Supervision, Project administration. **Patrick Edwards:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Tim Flanigan:** Software. **Jenya Kaganova:** Conceptualization, Methodology, Validation, Software, Formal analysis, Investigation, Resources, Data curation, Visualization. **Kevin Smith:** Validation, Formal analysis, Methodology. **Temeika Fairley:** Writing - review & editing, Project administration, Supervision, Funding acquisition, Supervision. **Nikki A. Hawkins:** Writing - review & editing, Project administration, Supervision, Funding acquisition, Supervision. **Juan L. Rodriguez:** Writing - review & editing, Project administration, Supervision, Funding acquisition, Supervision. **Gery P. Guy:** Writing - review & editing, Project administration, Supervision, Funding acquisition, Supervision. **Cheryll C. Thomas:** Writing - review & editing, Project administration, Supervision, Funding acquisition, Supervision.

Ethical approval

"All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards."

Informed consent

Informed consent was obtained from all individual participants included in the study.

Statement on the welfare of animals

This article does not contain any studies with animals performed by any of the authors.

Declaration of Competing Interest

The authors have no conflicts to report.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.evalprogplan.2021.101967.

differences between racial groups. Survey responses were linked with cancer registry data to assess whether surveys could reliably supplement registry data.

Results: A total of 830 women completed the survey for a response rate of 28.4 %. Blacks and Asian/Pacific Islanders were half as likely to respond as white women. Concordance between survey and registry data was high for demographic variables (Cohen's kappa [k]: 0.879 to 0.949), moderate to high for treatments received (k: 0.467 to 0.854), and low for hormone receptor status (k: 0.167 to 0.553). Survey items related to insurance status, employment, and symptoms revealed racial differences.

Conclusion: Cancer registry data, supplemented by patient surveys, can provide a broader understanding of the quality of care and financial impacts of breast cancer among young women.

Keywords

Breast cancer; State cancer registry; Survey data; Race/ethnicity; Insurance status; Late-stage diagnosis; Multifactorial research

1. Introduction

In 2017, over 40,000 women living in the United States were diagnosed with breast cancer before the age of 40, and it is estimated that more than 10,500 breast cancers are diagnosed each year (American Cancer Society, 2017). Young women face unique challenges, as their cancers are often diagnosed at a later stage than breast cancers in older women, and they undergo more aggressive treatments (Ademuyiwa, Cyr, Ivanovich, & Thomas, 2016; Anders, Johnson, Litton, Phillips, & Bleyer, 2009; Johnson, Chien, & Bleyer, 2013; Rosenberg, Newman, & Partridge, 2015). Currently, limited research focuses on the multifactorial causes of disparities in breast cancer outcomes among young women, such as tumor biology, access to care, quality of care, social support, and financial resources. In this manuscript, we explore opportunities to link data sources to address multifactorial drivers of health disparity.

Recent evidence suggests that, among their peers, nonwhite women are particularly disadvantaged when it comes to disease progression, prevalence, and screening. Breast cancer incidence is highest among young black women, as is the proportion of more aggressive breast cancers (Chollet-Hinton et al., 2017). Compared with non-Hispanic white women, non-Hispanic black and Hispanic women exhibit higher rates of diagnoses at later stages (Shoemaker, White, Wu, Weir, & Romieu, 2018). Although breast cancer mortality has decreased across all racial and ethnic groups over the past three decades, not all groups have experienced the same rate of decline in mortality and incidence—particularly non-Hispanic black, Hispanic, and Asian/Pacific Islander women and those younger than age 40 (Ademuyiwa et al., 2016; Anders et al., 2009; Johnson et al., 2013; Rosenberg et al., 2015).

There are many different types of breast cancer, each associated with a multitude of genetic, behavioral, and environmental factors, giving rise to the term *multifactorial disease* (Ritchie et al., 2001). This model provides a lens through which to view many of the disparities in health outcomes between young breast cancer survivors and older breast cancer survivors—

particularly when it comes to stage at diagnosis. Disparities in breast cancer outcomes can be explained partially by tumor characteristics, but other factors, including access to care, quality of care, social support, and financial resource availability, also contribute (Anders et al., 2009; Johnson et al., 2013; Letourneau et al., 2012; Rosenberg et al., 2015). In fact, recent evidence suggests that many young breast cancer survivors experience substantial financial decline as a result of their diagnosis and treatment and must make employment decisions solely to maintain health insurance coverage (Tangka et al., 2020). Psychosomatic factors, such as stress related to breast cancer treatment, along with other disparities in breast cancer outcomes across race/ethnicity, may contribute to the productivity losses that can occur due to a breast cancer diagnosis (Ekwueme et al., 2014; Maunsell, Brisson, Dubois, Lauzier, & Fraser, 1999)

Therefore, there is a need to explore a comprehensive set of underlying factors, including racial disparities, related to outcomes differences between young and older women. A limited number of studies have used cancer registry data, claims databases, and survey responses to assess breast cancer disparities but these have not specially addressed the unique issues facing younger women (Griggs et al., 2012; Hassett et al., 2016; Jagsi et al., 2014). To address this gap in the literature, we evaluated the utility of linking patient surveys with cancer registry databases to capture multifactorial data elements that impact health outcomes and financial burden among young breast cancer patients. We report on the methodology of a survey of young breast cancer patients that was administered in four U.S. states, systematically comparing differences in responses and response rates across racial/ethnic groups. We also assess the value of conducting patient surveys to supplement information not available in cancer registry data. Furthermore, we use linked cancer registry and survey responses to evaluate completeness and concordance between the information available from both sources. The findings from this study can help guide future methodological approaches to perform comprehensive and multifactorial evaluations of interventions and policies to improve outcomes and the overall well-being of young women diagnosed with breast cancer.

2. Methods

2.1. Study cohort and data collection overview

We included women with breast cancer who were diagnosed between the ages of 18 and 39 years from four state cancer registries with relatively large minority populations to ensure adequate sample size among all races and ethnicities. The central cancer registries in California, Florida, Georgia, and North Carolina were selected because new cancer cases in these states tended to be more evenly distributed among non-Hispanic whites and other racial/ethnic groups. Women were eligible to participate if they were (1) diagnosed with ductal carcinoma in situ (D05.90) or invasive breast cancer (C50) between January 1, 2013, and December 31, 2014, and (2) alive at the time of data extraction, as determined by central cancer registries, state death records, and the National Death Index, a national database compiled by the National Center for Health Statistics (2018 ICD-10-CM). All women who met the inclusion criteria were selected for the study to allow for population-based assessments. Information was collected from the eligible young women using a survey and these responses were then linked with data elements available in the cancer registry database.

Once all necessary approvals had been obtained from the states, and ethical approvals were obtained from RTI International, CDC, and the states' institutional review boards, along with data collection approval from the Office of Management and Budget (No. 0920–1123), we worked with each registry to compile the contact information for the breast cancer survivors who meet the study selection criteria. In Georgia, because of state privacy laws governing the conduct of the central cancer registry, the survey mailings were sent directly by staff at the Georgia State Cancer Registry. For other states, a centralized process was initiated to mail and track survey responses.

2.2. Patient survey

The survey consisted of dichotomous and multiple response questions on concepts identified through a literature search as potential sources of barriers and feedback from breast cancer advocacy groups. Subsections of the survey addressed insurance status, employment, access to care, out-of-pocket costs, and quality of care. Additional questions addressed the subject's quality of life; demographic information, such as race/ethnicity; and cancer history. Respondents completed the validated FBSI-8 (Functional Assessment of Chronic Illness Therapy (FACIT) Breast Symptom Index (FBSI), a list of eight common symptoms experienced by breast cancer patients (Yost et al., 2005). The respondents selected appropriate categories to indicate whether they experience each symptom "not at all," "a little bit," "somewhat," "quite a bit," or "very much." The survey was designed to be completed in about 20 min and priority was given to questions that had been previously tested and fielded for low-income and low-literacy populations. Before sending the mailings, we performed cognitive testing in English (with nine women) and Spanish (with eight women) to ensure that the questions were appropriate for respondents with low levels of literacy (Arora, Reeve, Hays, Clauser, & Oakley-Girvan, 2011; Jagsi et al., 2014; Malin et al., 2010; Smith et al., 2007).

The first mailings were sent in March 2017 and included the paper survey in English and Spanish, along with a cover letter that discussed passive consent and instructions for the web-based version of the survey. All mailings included both the Spanish- and English-language materials so respondents could select their preferred language. Participants could choose to respond using the paper survey or the web-based version. We offered a $10 gift card incentive to participants who completed the survey. Reminder postcards were sent to individuals who did not respond within 2 weeks, and a second survey instrument was mailed to those who did not return a completed survey within 8 weeks of the initial mailing. We tracked all mail that was returned because of an incorrect address. In addition, we logged phone and mail correspondence indicating whether an individual we were trying to reach had died.

All surveys completed online were automatically uploaded into a dataset within RTI's Hatteras Survey System. The paper survey responses were scanned, interpreted, and verified using TELEform software (Cardiff Software, San Marcos, CA). When the survey operations were completed, the web-based and mail responses were combined into a single dataset to support analysis and reporting.

2.3. Central cancer registry data

We linked the data from our surveys with the central cancer registry data via previously assigned ID numbers. Respondents identified from the California Cancer Registry had to provide consent for linkage with the cancer registry data, as stipulated by the state institutional review board. Nineteen of these respondents did not provide consent, so their information was not linked with the cancer registry data. We obtained demographic and cancer-specific information from the registry data, including the date of diagnosis, stage at diagnosis (Surveillance, Epidemiology, and End Results [SEER] summary stage), and type of initial cancer treatment (e.g., surgery, radiation therapy, chemotherapy, hormone treatment, immunotherapy). Additionally, we obtained details on hormonal status related to estrogen, progesterone, and human epidermal growth factor receptor 2 (HER2).

2.4. Statistical analysis

We created a mutually exclusive race/ethnicity variable to facilitate our analysis. In the survey, participants were instructed to select as many racial/ethnic categories as applied. To help ensure that the race/ethnicity variable remained mutually exclusive, those who selected “Hispanic” in addition to any other race variable(s) were coded as “Hispanic.” All other racial/ethnic combinations that did not include “Hispanic” were considered nonhierarchical and resulted in a coding of “other.” Our final race categories were “non-Hispanic white,” “non-Hispanic black,” “Hispanic,” “non-Hispanic Asian/Pacific Islander,” and “non-Hispanic other.”

As a first step, we evaluated the response rate to the survey stratified by race/ethnicity. We cataloged the number of undeliverable surveys—either due to an incorrect address or because the respondent was no longer alive—to determine the true available sample and calculate the response rate. We also noted the number of responses by mode (web-based vs. mail) and language (English vs. Spanish). We conducted multivariable analysis to assess response-rate differences by race/ethnicity using state-specific logistic regression specifications, as the data available for analysis varied across the states. The dependent variable was set to one if a member of the available sample responded to the survey and zero if she did not. Our primary regression included dichotomous state indicator variables for the four central cancer registry locations and binary variables for each racial/ethnic group (see Table 2). In subsequent regressions, we included age, stage at diagnosis, and treatment variables to assess the impact of including these additional covariates (see supplement Table S1).

To further assess variation among racial/ethnic groups, we compared differences between non-Hispanic white respondents and the combined respondents of all other racial/ethnic groups. We pooled minority groups to ensure sufficient sample sizes for the planned comparisons and measured the concordance between survey responses and registry data for variables that were available from both sources. These data elements included demographic variables related to age and marital status; treatment variables, including surgery, chemotherapy, radiation, and hormone therapy; and HER2 status. Measures of concordance reported are percentage agreement and Cohen's kappa statistic. The kappa statistic measures the extent of exact agreement, adjusting for chance agreement. In general, values greater than 0.75 represent excellent concordance, values of 0.40 to 0.75 represent moderate concordance, and values less than 0.40 represent poor concordance (Gupta et al., 2011). For all the above concordance analyses, we excluded all respondents identified from the Florida Cancer Data System and the aforementioned 19 respondents from California. We did not receive treatment and hormone status information from the Florida Cancer Data System.

The stage at diagnosis information collected through self-report in the survey was based on the American Joint Commission on Cancer (AJCC), as this staging system was considered most relevant to the discussions patients would have with their physicians (Edge & Compton, 2010). The information we received from the cancer registry contained the SEER summary stage. These two staging categories are not directly comparable (Walters et al., 2013), but we stratified both categories by non-Hispanic white and all other racial/ethnic groups to identify potential differences in using one definition vs. another in race/ethnicity-based assessments.

Finally, we reported differences in non-Hispanic white respondents vs. respondents of all other racial/ethnic groups in selected variables available in the survey that are not available in the registry data to evaluate how useful this information is for conducting racial disparity analysis. The survey responses could supplement the information available in the registry data to perform a more comprehensive assessment of multifactorial attributes of disparities. We assess racial differences in insurance at time of diagnosis and at the time of survey response along with the FSBI symptoms. For the FBSI, we examined the proportion who responded either “quite a bit” or “very much” to each of the symptoms included in the index.

All statistical analysis was performed using STATA SE version 15 (StataCorp., College Station, TX).

3. Results

In Table 1, we summarize the response rate for each of the four states in this study. Surveys were initially mailed to 3,659 young breast cancer survivors; however, 733 of these were undeliverable due to the incorrect address, outdated address, or death of the targeted respondent. Therefore, the total available sample was 2,926. In total, 830 young women completed the survey, a response rate of 28.4 %. The response rate was very similar across states, ranging from 27.9%–29.3%. Most responded using paper surveys (702 respondents) rather than the web-based version, and most used English (808 respondents).

Fig. 1 presents the racial/ethnic distribution of the target cohort, those identified as eligible by the cancer registries, the available sample, and the survey respondents. Although the racial/ethnic distribution was similar in the target sample (50.5 % non-Hispanic white) and available sample (51.2 % non-Hispanic white), a higher percentage of non-Hispanic whites responded to the survey. Consequently, there is a larger proportion of non-Hispanic whites than other racial/ethnic groups among survey respondents (58.7 %).

Table 2 contains the results from multivariable logistic regressions to explore the odds of a response by race/ethnicity, controlling for the state in which the women were diagnosed. There are no statistically significant differences by state. Across all states, Hispanic (odds ratio (OR): 0.42; 95 % confidence interval (CI): 0.33–0.54), non-Hispanic black, (OR:0.50; CI:0.39–0.62) and non-Hispanic Asian/Pacific Islander women (OR:0.54; CI:0.37–0.77) were only about half as likely to respond as non-Hispanic white women. All these differences in response to the survey are statistically significant (p-value ≤ 0.001). This finding remains stable across all regression specifications when additional variables, such as age, stage at diagnosis, and treatments, are included (see supplement Table S1).

Among survey respondents, few demographic and clinical characteristics were significantly different between non-Hispanic white women and those of all other racial/ethnic categories women (Table 3). There was a similar proportion of non-Hispanic white women as women from all other racial/ethnic groups between ages 18–34, and they received the same types of nonsurgical treatments, including chemotherapy, radiotherapy, and hormone therapy. According to the survey data, the statistically significant differences ($p < 0.05$) were that non-Hispanic white women were more likely than nonwhite women to be married (74.5 % vs. 55.4 %), undergo surgery (97.9 % vs. 93.7 %), and have hormone receptor–positive breast cancers (33.5 % vs. 25.1 %).

In Table 3, we also report the concordance between survey responses and registry data. We used self-reported race distribution in the survey to stratify non-Hispanic white respondents from all other racial/ethnic categories, as we assumed that self-reported race information would be more accurate than registry-collected race data. Survey and registry data were generally in high agreement on race, with 92.6 % agreement and a kappa statistic (k) of 0.850 (data not shown), but the assignment for other variables differed. The concordance between survey response and registry data for age and marital status was very high among both non-Hispanic whites and non-white women. For example, for marital status, the agreement was 95.7 % ($k = 0.891$) for non-Hispanic white women and 94.0 % ($k = 0.879$) for women of other racial/ethnic groups.

The treatment variables varied more. According to percentage agreement and kappa values, chemotherapy had high concordance, and radiation therapy and hormone therapy had lower concordance. For surgery, the percentage agreement was high (90.0 % or higher); however, the kappa statistics were low (0.203 and 0.217) for both non-Hispanic white respondents and those from all other racial/ethnic groups, as survey respondents reported additional surgical procedures that were not in the registry data. For surgery, radiation, and hormone therapy, overall, the percentage agreement was lower for all other racial/ethnic groups than for non-Hispanic white respondents. Hormone receptor status had the lowest level of concordance between survey and registry information, whereas HER2 had moderate concordance overall. For both hormone receptor and HER2 status, concordance among all other racial/ethnic groups revealed much lower percentage agreement and kappa values than among non-Hispanic white respondents. Overall, compared with the self-reported responses, the registry data indicated more instances where hormone status was positive.

Table 4 presents the stage distributions from the survey and the registry data. Even though the definitions differ, the overall pattern is similar. Stage distributions show that respondents from nonwhite racial/ethnic groups have higher proportions of stage 0 or in situ breast cancer and stage IV or distant cancers than non-Hispanic white respondents. The stage information from the cancer registry shows a lower proportion of young women with missing information: 1.2 % for non-Hispanic white women and 3.3 % for women in other racial/ethnic groups in the cancer registry data, compared with 3.7 % and 4.8 % for those respective groups of women in the survey data.

In Table 5, we compare insurance status and breast cancer symptoms between nonHispanic white respondents and those from all other racial/ethnic groups. These additional variables are only available from the survey data. Non-Hispanic white respondents displayed statistically significant differences in insurance status, both at the time of breast cancer diagnosis and at the time of survey response, from women of all other racial/ethnic groups. Compared with the proportion of non-Hispanic whites, a higher proportion of respondents from nonwhite racial/ethnic groups were enrolled in Medicaid or uninsured, and a lower proportion had private insurance or were self-insured ($p < 0.001$). Insurance status also changed between diagnosis and survey response, with higher Medicaid enrollment and a lower proportion of uninsured individuals across both groups. At the time of survey response, 8.0 % of non-Hispanic white respondents were enrolled in Medicaid, compared with 23.9 % of the women from other ethnic/racial groups. The symptom index reported in Table 5 also shows statistically significant differences between the two groups in the mean FBSI score and most of the specific symptoms, including higher reports of lack of energy, nausea, and shortness of breath. Non-Hispanic white respondents did not differ from those of other racial/ethnic groups in terms of their satisfaction with their overall quality of life and well-being.

4. Discussion

In this study, we examined the utility of supplementing cancer registry data with survey data to assess multifactorial aspects impacting outcomes and financial toxicity among young breast cancer survivors. Survey data can serve as a rich source of information, as many more domains and concepts can be explored through self-report than through cancer registry data alone. Survey data can capture patient-reported symptoms and outcomes and key areas related to insurance status and employment, which can directly affect access to quality care.

We found the concordance between survey and registry data to be high for demographic variables and chemotherapy. As in prior studies, we found some differences in treatment reported for surgery and radiation therapy (Cooper et al., 2002; Noone et al., 2016; Virnig et al., 2002). Additional information about treatments received was identified via self-report; therefore, the use of both data sources may be appropriate. One reason for using both data sources is the difference in the timeframe captured, as the cancer registry data only reflect the initial course of treatment. Hormone receptor status and stage information appear to be underreported in the survey data relative to registry data—this may suggest an area for further exploration to ensure the accuracy of the information provided in both data sources.

The insurance status information available from the survey reveals racial differences between white and minority women. This can be an important source of disparities (Wharam et al., 2018) and highlights the benefits of supplementing cancer registry data with patient survey information. Additionally, the details on insurance status at the time of the survey allow for a systematic assessment of changes in insurance status after cancer diagnosis. The proportion of young women enrolled in Medicaid increased among both non-Hispanic white women and non-white women, but the increase was much more pronounced for non-white women. The differences and changes in insurance coverage can have meaningful impacts on access to care and quality of the services received. Additionally, more women became self-insured after the cancer diagnosis and may experience significant financial burden which can impact their breast cancer outcomes and overall wellbeing. Further research is needed to understand the role of health insurance coverage, financial stress, and breast cancer outcomes among young women.

The symptom index completed by the survey respondents revealed that young women from racial/ethnic groups experience more detrimental symptoms after breast cancer diagnosis than non-Hispanic white young women. These symptoms may be a result of underlying disparities (Patel et al., 2020) and the drivers of these racial differences should be further investigated at the individual, provider, and health system level factors. A comprehensive assessment of the barriers can help develop interventions to ameliorate disparities faced by young women diagnosed with breast cancer.

Although this study provided important knowledge on the utility of survey and cancer registry data, there are a few limitations. One primary limitation in the use of patient surveys among young breast cancer survivors is the low response rate. However, our response rate of 28 % is comparable to, and often higher than, rates reported by other surveys (9 %–21 %) targeting young women diagnosed with breast cancer (Tathanhlong, Bristow, & McGuffin, 2015; Tercyak et al., 2015). Furthermore, we found that non-Hispanic black, Hispanic, non-Hispanic Asian/Pacific Islander, and non-Hispanic-other women are less likely to respond than non-Hispanic white women (George, Duran, & Norris, 2014). This lower response rate poses a challenge for racial disparity analysis, as adequate sample sizes across all racial/ethnic groups are required for meaningful comparisons. Additional research should identify approaches to increase the response rate among minorities in the young female breast cancer population including modality of initial contact, reminder procedures and optimal incentives.

Both survey and cancer registry data may present inherent challenges with regard to collection and analysis (Malin et al., 2002). For example, clinical information may be missing in the cancer registry data, and self-reported clinical data is subject to a variety of information biases—particularly recall and measurement biases (Althubaiti, 2016). These data could be supplemented with clinical information available from electronic medical records, or enhanced with information from all-state discharge databases when available to provide details on treatment received.

The variety of clinical, demographic, and self-reported variables that are available through the linkage of survey and cancer registry data, supplemented by other sources when feasible, can enhance our understanding of the multiple factors that determine breast cancer outcomes. For instance, cancer registries do not routinely collect data on variables that could be used to thoroughly evaluate racial/ethnic disparities (e.g., changes in insurance status over the treatment period). This study on breast cancer among young women demonstrates the utility of partnering with cancer registries to recruit participants, validate self-reported survey responses, and obtain detailed clinical data. The resulting linked database may provide a foundational model for multifactorial research, not only on outcomes and financial consequences faced by young women with breast cancer, but also for many types of chronic diseases with multiple underlying factors that impact health and economic outcomes.

4.1. Lessons Learned

Evaluators and planners seeking to enhance survey data with variables obtained from cancer registry databases can draw several key in-sights from this study. First, we would like to emphasize the areas that researchers can potentially strengthen their understanding of the of data—and thereby of the factors that determine breast cancer outcomes—by linking the two datasets. Concordance between the two data sources appeared high for demographic variables and for the chemotherapy treatment variable, but the divergence widened across the surgery and radiation therapy variables. Hormone receptor status and stage information were underreported in the survey data as opposed to the cancer registries.

However, the low survey response rate that is often observed among young breast cancer survivors was a primary limitation of this study (Tathanhlong et al., 2015; Tercyak et al., 2015), and minorities were even less likely to respond. Thus, we did not have an adequate sample size to perform individual analysis by racial/ethnic category. Our inability to further stratify the analysis by race/ethnicity could have masked potential differences between Hispanic, non-Hispanic black, non-Hispanic Asian/Pacific Islander, and other nonwhite respondents. Thus, further research is necessary to increase minority response enough to enable racial disparity assessments.

Another direction for further research could be to supplement the survey and cancer registry data with information from all-state discharge databases to provide details on treatment received. The variety of clinical, demographic, and self-reported variables that are available through the linkage of multiple data sources can enhance our understanding of the factors that determine breast cancer outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

Funding support for RTI staff was provided by the Centers for Disease Control and Prevention (CDC) (Contract N0. 200-2008-27958 Task 48, to RTI International).

The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Biography

Sujha Subramanian, PhD, MA, is a Fellow in Health Economics and Policy Research at RTI International and has extensive experience performing economic burden assessments and evaluations of noncommunicable disease (NCD) screening programs both in the United States and in international settings. Over the past two decades, she has directed several program evaluations, including the assessment of the cost and effectiveness of the National Breast and Cervical Cancer Early Detection Program (NBCCEDP) and the Colorectal Cancer Control Program (CRCCP). Dr. Subramanian received her master's in Applied Econometrics & Public Finance and PhD in Health Economics from Clark University.

Madeleine Jones, BS, is an economist in RTI's Health Care Financing and Payment Program, where she applies quantitative and qualitative methods to support health economics and policy research. Her current work is primarily in the areas of health care payment systems and noncommunicable disease screening program evaluation. Ms. Jones received her bachelor's degree in Agricultural and Applied Economics from the University of Wisconsin, Madison.

Florence K.L. Tangka, PhD, MS, is a health economist in the Division of Cancer Prevention and Control's Epidemiology and Applied Research Branch. She is the principal investigator for a number of CDC cancer economics studies. Her research focuses on the economics of cancer, the economics of CDC's National Program of Cancer Registries and Colorectal Cancer Control Program, and the use of breast and cervical cancer screening services. Dr. Tangka received her master's from Rutgers University in agricultural economics, her doctoral degree from the University of Florida, Department of Food and Resource Economics, and completed a two-year postdoctoral fellowship in Prevention Effectiveness at CDC.

Patrick Edwards, MS, is an economist in RTI International's Health Care Financing and Payment program where he uses quantitative and qualitative methods to assess public health programs and health policies. Mr. Edwards' research focuses primarily on the health economics, evaluation, and effectiveness of public health programs and services. He received his master's in Global Health Science from the University of Oxford, United Kingdom.

Tim Flanigan, MA, is a survey methodologist with more than 22 years of professional experience in survey design, questionnaire pretesting, and data management. Mr. Flanigan has experience supervising field staff, training new interviewers, field material design, database management, and data analysis. Mr. Flanigan received his Master's in Sociology, with emphasis on Applied Social Research and Statistics, from West Virginia University, Morgantown.

Jenya Kaganova, PhD, MS, is a senior research systems programmer/analyst at RTI International with more than 15 years of experience in health care data analysis, including creating data sets and databases for statistical analysis, carrying out analyses, and contributing to the dissemination of findings via technical reports and peer-reviewed journal articles. Dr. Kaganova received her master's in Mathematics and her PhD in Mathematical Statistics and Probability Theory from Moscow University.

Kevin Smith, MA, is a behavioral scientist with a strong background in survey methods, psychometric assessment, evaluation research, sampling methodology, and multivariate statistical analysis. Over the past 20 years, he has been involved in a variety of studies of chronic diseases and population health. He served as the principal investigator on National Institutes of Health (NIH)-funded studies of heart disease risk, mammography use, pediatric brain injury, and the measurement of socioeconomic status. Mr. Smith received his master's in Sociology from Tufts University, focusing on applied research methods.

Temeika Fairley, PhD, is a Senior Health Scientist with the Centers for Disease Control and Prevention's Division of Cancer Prevention and Control. Dr. Fairley directs public health research, program support, and communications. She puts her interest and expertise to work in the fields of health disparities and breast cancer survivorship, especially for women diagnosed under the age of 40. Dr. Fairley earned her PhD in Biology from the University of Vermont.

Nikki A. Hawkins, PhD, is a Lead Health Scientist in the Epidemiology Branch in CDC's Office on Smoking and Health. She has expertise in research and evaluation methods, especially within the areas of psychosocial and systems-level factors that influence health behavior and chronic disease prevention. She has co-authored over 40 peer-reviewed articles on preventive health behavior, care delivery, health literacy, and psychosocial outcomes associated with chronic disease. Dr. Hawkins received her PhD from the University of California, Irvine.

Juan L. Rodriguez Juan Rodriguez, MPH, MS, is at the Centers for Disease Control and Prevention (CDC) in the Division of Cancer Prevention and Control. His research spans the cancer continuum, from primary cancer prevention and screening, to cancer survivorship. Mr. Rodriguez is also the Program Director for the CDC Public Health Cancer Genomics Program. His programmatic work has focused on the implementation of cancer genomics into public health settings, specifically state health departments. Mr. Rodriguez received his master's in Behavioral Science and Health Education from Emory University.

Gery P. Guy Jr., PhD, MPH, is a health scientist and serves as the team lead for the Applied Research, Evaluation, and Integration team in the Division of Overdose Prevention (DOP). His work at CDC has focused on opioid and naloxone prescribing in the United States, the impact of the CDC Guideline for Prescribing Opioids for Chronic Pain, and the impact of policies and interventions on drug overdoses. Dr. Guy received his MPH in Health Policy from the Rollins School of Public Health at Emory University, his PhD in Health Services Research and Health Policy from Emory University, and completed a post-doctoral fellowship in Prevention Effectiveness at the CDC.

Cheryll C. Thomas, MS, is the current Deputy Associated Director for Science, a position she has held since September 2010. She worked for six years on a variety of projects related to the National Program of Cancer Registries. Prior to coming to CDC, Ms. Thomas worked at the American Cancer Society's National Home Office in cancer surveillance. She developed surveillance publications for data dissemination, provided and assisted with the interpretation of incidence, mortality, and behavioral risk factor data for cancer control planning, and analyzed data for peer-reviewed publications. Ms. Thomas received her master's in Public Health (Epidemiology) from Emory University's Rollins School of Public Health.

References

Ademuyiwa FO, Cyr A, Ivanovich J, & Thomas MA (2016). Managing breast cancer in younger women: Challenges and solutions. Breast Cancer (Dove Medical Press), 8, 1–12. 10.2147/BCTT.S68848 [PubMed: 26730210]

Althubaiti A (2016). Information bias in health research: Definition, pitfalls, and adjustment methods. Journal of Multidisciplinary Healthcare, 9, 211–217. 10.2147/JMDH.S104807 [PubMed: 27217764]

American Cancer Society. (2017). Breast Cancer facts and figures: 2017–2018. Retrieved April 1, 2020, from https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf.

Anders CK, Johnson R, Litton J, Phillips M, & Bleyer A (2009). Breast cancer before age 40 years. Seminars in Oncology, 36(3), 237–249. 10.1053/j.seminoncol.2009.03.001 [PubMed: 19460581]

Arora NK, Reeve BB, Hays RD, Clauser SB, & Oakley-Girvan I (2011). Assessment of quality of cancer-related follow-up care from the cancer survivor's perspective. Journal of Clinical Oncology, 29(10), 1280–1289. 10.1200/JCO.2010.32.1554 [PubMed: 21357781]

Chollet-Hinton L, Olshan AF, Nichols HB, Anders CK, Lund JL, Allott EH, et al. (2017). Biology and etiology of young-onset breast cancers among premenopausal African American women: Results from the AMBER Consortium. Cancer Epidemiology, Biomarkers & Prevention, 26(12), 1722–1729. 10.1158/1055-9965.Epi-17-0450

Cooper GS, Virnig B, Klabunde CN, Schussler N, Freeman J, & Warren JL (2002). Use of SEER-Medicare data for measuring cancer surgery. Medical Care, 40(8 Suppl). 10.1097/01.MLR.0000020943.21850.F1. IV-43–48.

Edge SB, & Compton CC (2010). The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Annals of Surgical Oncology, 17(6), 1471–1474. 10.1245/s10434-010-0985-4 [PubMed: 20180029]

Ekwueme DU, Guy GP Jr., Rim SH, White A, Hall IJ, Fairley TL, et al. (2014). Health and economic impact of breast cancer mortality in young women, 1970–2008. American Journal of Preventive Medicine, 46(1), 71–79. 10.1016/j.amepre.2013.08.016 [PubMed: 24355674]

George S, Duran N, & Norris K (2014). A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. American Journal of Public Health, 104(2), E16–E31. 10.2105/Ajph.2013.301706

Griggs JJ, Hawley ST, Graff JJ, Hamilton AS, Jagsi R, Janz NK, et al. (2012). Factors associated with receipt of breast cancer adjuvant chemotherapy in a diverse population-based sample. Journal of Clinical Oncology, 30(25), 3058–3064. 10.1200/JCO.2012.41.9564 [PubMed: 22869890]

Gupta V, Gu K, Chen Z, Lu W, Shu XO, & Zheng Y (2011). Concordance of self-reported and medical chart information on cancer diagnosis and treatment. BMC Medical Research Methodology, 11, 72. 10.1186/1471-2288-11-72 [PubMed: 21592352]

Hassett MJ, Schymura MJ, Chen K, Boscoe FP, Gesten FC, & Schrag D (2016). Variation in breast cancer care quality in New York and California based on race/ethnicity and Medicaid enrollment. Cancer, 122(3), 420–431. 10.1002/cncr.29777 [PubMed: 26536043]

Jagsi R, Pottow JA, Griffith KA, Bradley C, Hamilton AS, Graff J, et al. (2014). Long-term financial burden of breast cancer: Experiences of a diverse cohort of survivors identified through population-based registries. Journal of Clinical Oncology, 32(12), 1269–1276. 10.1200/JCO.2013.53.0956 [PubMed: 24663041]

Johnson RH, Chien FL, & Bleyer A (2013). Incidence of breast cancer with distant involvement among women in the United States, 1976 to 2009. JAMA, 309(8), 800–805. 10.1001/jama.2013.776 [PubMed: 23443443]

Letourneau JM, Smith JF, Ebbel EE, Craig A, Katz PP, Cedars MI, et al. (2012). Racial, socioeconomic, and demographic disparities in access to fertility preservation in young women diagnosed with cancer. Cancer, 118(18), 4579–4588. 10.1002/cncr.26649 [PubMed: 22451228]

Malin JL, Diamant AL, Leake B, Liu Y, Thind A, Kahn KL, et al. (2010). Quality of care for breast cancer for uninsured women in California under the breast and cervical cancer prevention treatment act. Journal of Clinical Oncology, 28(21), 3479–3484. 10.1200/JCO.2009.27.7491 [PubMed: 20530279]

Malin JL, Kahn KL, Adams J, Kwan L, Laouri M, & Ganz PA (2002). Validity of cancer registry data for measuring the quality of breast cancer care. Journal of the National Cancer Institute, 94(11), 835–844. 10.1093/jnci/94.11.835 [PubMed: 12048271]

Maunsell E, Brisson C, Dubois L, Lauzier S, & Fraser A (1999). Work problems after breast cancer: An exploratory qualitative study. Psychooncology, 8(6), 467–473. [PubMed: 10607979]

Noone AM, Lund JL, Mariotto A, Cronin K, McNeel T, Deapen D, et al. (2016). Comparison of SEER treatment data with Medicare claims. Medical Care, 54(9), e55–e64. 10.1097/MLR.0000000000000073 [PubMed: 24638121]

Patel MI, Lopez AM, Blackstock W, Reeder-Hayes K, Moushey EA, Phillips J, et al. (2020). Cancer disparities and health equity: A policy statement from the American Society of Clinical Oncology. Journal of Clinical Oncology, 38(29), 3439–3448. 10.1200/JCO.20.00642 [PubMed: 32783672]

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. American Journal of Human Genetics, 69(1), 138–147. 10.1086/321276 [PubMed: 11404819]

Rosenberg SM, Newman LA, & Partridge AH (2015). Breast cancer in young women: rare disease or public health problem? JAMA Oncology, 1(7), 877–878. 10.1001/jamaoncol.2015.2112 [PubMed: 26204453]

Shoemaker ML, White MC, Wu MX, Weir HK, & Romieu I (2018). Differences in breast cancer incidence among young women aged 20–49 years by stage and tumor characteristics, age, race, and ethnicity, 2004–2013. Breast Cancer Research and Treatment, 169(3), 595–606. 10.1007/s10549-018-4699-9 [PubMed: 29445940]

Smith T, Stein KD, Mehta CC, Kaw C, Kepner JL, Buskirk T, et al. (2007). The rationale, design, and implementation of the American Cancer society’s studies of cancer survivors. Cancer, 109(1), 1–12. 10.1002/cncr.22387 [PubMed: 17146781]

Tangka FKL, Subramanian S, Jones M, Edwards P, Flanigan T, Kaganova Y, et al. (2020). Insurance coverage, employment status, and financial well-being of young women diagnosed with breast cancer. Cancer Epidemiology, Biomarkers & Prevention, 29(3), 616–624. 10.1158/1055-9965.EPI-19-0352

Tathanhlong L, Bristow B, & McGuffin ML (2015). Providing supportive care for young breast cancer patients: Exploring the comfort level and educational needs of the interprofessional health care team at a comprehensive cancer centre. Journal of Medical Imaging and Radiation Sciences, 46(3), S69–S74. [PubMed: 31052111]

Tercyak KP, Silber E, Johnson AC, Fleischmann A, Murphy SE, Mays D, et al. (2015). Survey on addressing the information and support needs of Jewish women at increased risk for or diagnosed with breast cancer: The Sharsheret experience. Healthcare (Basel), 3(2), 324–337. 10.3390/healthcare3020324 [PubMed: 27417765]

Virnig BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, & Freeman J (2002). Studying radiation therapy using SEER-Medicare-linked data. Medical Care, 40(8 Suppl). 10.1097/01.MLR.0000020940.90270.4D. IV-49–54.

Walters S, Maringe C, Butler J, Brierley JD, Rachet B, & Coleman MP (2013). Comparability of stage data in cancer registries in six countries: Lessons from the International cancer benchmarking partnership. International Journal of Cancer, 132 (3), 676–685. 10.1002/ijc.27651 [PubMed: 22623157]

Wharam JF, Zhang F, Lu CY, Wagner AK, Nekhlyudov L, Earle CC, et al. (2018). Breast cancer diagnosis and treatment after high-deductible insurance enrollment. Journal of Clinical Oncology, 36(11), 1121–1127. 10.1200/JCO.2017.75.2501 [PubMed: 29489428]

Yost KJ, Yount SE, Eton DT, Silberman C, Broughton-Heyes A, & Cella D (2005). Validation of the functional assessment of cancer therapy-breast symptom index (FBSI). Breast Cancer Research and Treatment, 90(3), 295–298. 10.1007/s10549-004-5024-3 [PubMed: 15830143]

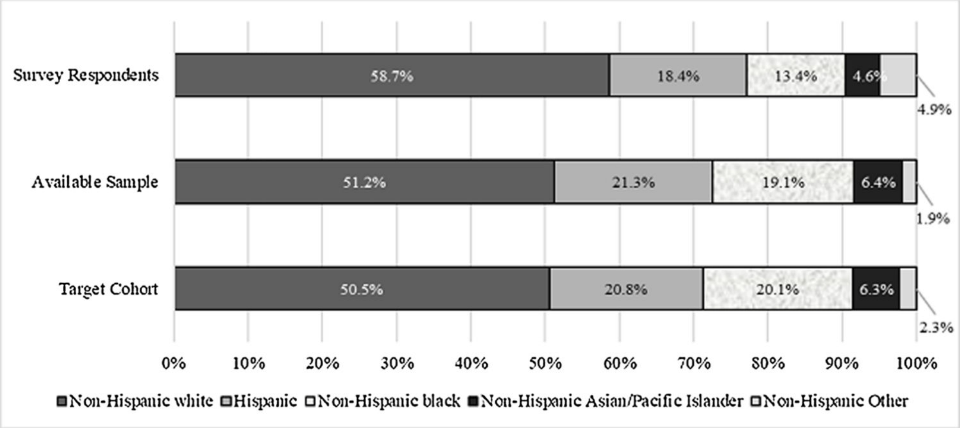

Fig. 1.

Racial/ethnic distribution of the target cohort, available sample, and survey respondents.

Note: The target cohort includes all breast cancer survivors who were identified via state cancer registry and were mailed surveys. In the available sample, individuals identified as deceased or as having an undeliverable address have been dropped.

Alt text: This bar graph displays the racial/ethnic distribution across three populations: the target cohort, available sample, and actual survey respondents. Non-Hispanic whites responded at the highest rates across all three populations. Hispanics responded the next most frequently, followed by non-Hispanic Asian/Pacific Islanders and other non-Hispanic minority groups. Compared with the target cohort, survey respondents tended to skew non-Hispanic white at the expense of all other racial/ethnic groups except for “non-Hispanic other.”

Note to the Editors: This figure was created with Microsoft Excel 2016 and is attached separately as a TIFF file.

Table 1

Survey response rate assessment by state.

State[a]	Target Cohort	Undeliverable/Deceased	Available Sample	Responded	Response Rate	Mode of Response		Language of Respondent	
						Web	**Paper**	**English**	**Spanish**
California	1,110	148	962	268	27.9 %	37	231	255	13
Florida	1,146	273	873	246	28.2 %	40	206	237	9
Georgia	701	170	531	152	28.6 %	26	126	152	0
North Carolina	702	142	560	164	29.3 %	25	139	164	0
Total	3,659	733	2,926	830	28.4 %	128	702	808	22

[a] The *State* variable indicates that the respondents were drawn from each state's respective central data registry: the Florida Cancer Data System, the North Carolina Central Cancer Registry, the California Cancer Registry, and the Georgia Comprehensive Cancer Registry.

Table 2

Variables associated with survey participation.

Covariate	**Modeling Response to Survey (Available sample = 2,926)**		
	OR	**95 % CI**	***P***
State[a]			
North Carolina (reference)	1.00		
California	1.13	0.88 to 1.44	0.882
Florida	1.05	0.83 to 1.33	0.699
Georgia	1.05	0.81 to 1.37	0.703
Race/Ethnicity			
Non-Hispanic white (reference)	1.00		
Hispanic	0.42	0.33 to 0.54	< 0.001
Non-Hispanic black	0.50	0.39 to 0.62	< 0.001
Non-Hispanic Asian/Pacific Islander	0.54	0.37 to 0.77	0.001
Non-Hispanic Other	0.92	0.52 to 1.61	0.763

OR odds ratio, *CI* confidence interval.

[a] The *State* variable indicates that the respondents were drawn from each state's respective central data registry: the Florida Cancer Data System, the North Carolina Central Cancer Registry, the California Cancer Registry, and the Georgia Comprehensive Cancer Registry.

Table 3

Concordance between survey responses and registry data for non-Hispanic white respondents vs. all other racial/ethnic groups.

	Characteristics of Survey Respondents		Concordance Between Survey Responses and Registry Data			
			Non-Hispanic White		All Other Racial/Ethnic Groups	
	Non-Hispanic White Freq (%)	All Other Racial/Ethnic Groups Freq (%)	Agreement Freq (%)	Kappa Statistic (CI)	Agreement Freq (%)	Kappa Statistic (CI)
Demographics						
Aged 18–34	124 (36.5 %)	93 (41.7 %)	332 (97.6 %)	0.949 (0.914–0.984)	213 (95.5 %)	0.907 (0.850–0.963)
Married	251 (74.5 %)	119 (55.4 %)*	291 (95.7 %)	0.891 (0.833–0.949)	189 (94.0 %)	0.879 (0.813–0.946)
Treatments						
Surgery	332 (97.9 %)	209 (93.7 %)*	320 (94.4 %)	0.217 (−0.006–0.440)	198 (89.6 %)	0.203 (−0.010–0.415)
Chemotherapy	263 (77.6 %)	166 (74.4 %)	315 (92.9 %)	0.814 (0.743–0.885)	208 (94.1 %)	0.854 (0.778–0.930)
Radiation therapy	203 (59.9 %)	146 (65.5 %)	275 (81.1 %)	0.635 (0.559–0.710)	165 (74.7 %)	0.520 (0.423–0.616)
Hormone therapy	214 (63.1 %)	136 (61.0 %)	251 (74 %)	0.499 (0.416–0.582)	161 (72.9 %)	0.467 (0.357–0.577)
Hormone Receptor or HER2 Status						
Hormone receptor–positive	114 (33.5 %)	56 (25.1 %)*	205 (60.3 %)	0.300 (0.229–0.370)	108 (48.4 %)	0.167 (0.099–0.235)
HER2-positive	76 (22.4 %)	42 (18.8 %)	287 (84.4 %)	0.553 (0.447–0.660)	172 (77.1 %)	0.285 (0.135–0.435)

CI confidence interval.

Note: Only includes those eligible in North Carolina, Georgia, and California. The cancer registry in Florida did not provide treatment data.

[*] Differences between non-Hispanic white and all other racial/ethnic groups are statistically significant at the 5 % level.

Table 4

Stage at diagnosis from survey responses and registry data.

	Non-Hispanic White Freq (%)	All Other Racial/Ethnic Groups Freq (%)	*P*
Stage from survey (AJCC staging)			0.016
0	53 (10.9 %)	38 (11.3 %)	
I	119 (24.4 %)	61 (18.2 %)	
II	179 (36.8 %)	112 (34.4 %)	
III	97 (19.9 %)	76 (22.7 %)	
IV	21 (4.3 %)	32 (9.6 %)	
Unknown or missing	18 (3.7 %)	16 (4.8 %)	
Stage from registry data (SEER summary staging)			0.099
In situ	49 (10.1 %)	42 (12.5 %)	
Local	220 (45.2 %)	142 (42.4 %)	
Regional	190 (39.0 %)	118 (35.2 %)	
Distant	22 (4.5 %)	22 (6.6 %)	
Unknown or missing	6 (1.2 %)	11 (3.3 %)	

AJCC American Joint Committee on Cancer, *SEER* Surveillance, Epidemiology, & End Results Program. The AJCC stage categories cannot be directly compared to the SEER categories. Although they are derived from similar source data elements, they use difference assumptions and algorithms to assign cases to the stage categories. We did not have access to the source data elements and therefore did not attempt to compare the values. Our goal was to highlight the percentage missing from self-report data compared to registry data.

Table 5

Insurance status and breast cancer symptoms from patient survey.

	Non-Hispanic White	**All Other Racial/Ethnic Groups**	***P***
Insurance status at diagnosis [a]			
Private insurance	362 (74.3 %)	208 (62.1 %)	< 0.001
Self-insured	46 (9.5 %)	17 (5.1 %)	
Medicaid (some Medicare)	28 (5.8 %)	52 (15.5 %)	
Other	23 (4.7 %)	15 (4.5 %)	
Uninsured	26 (5.3 %)	37 (11.0 %)	
Insurance status at time of survey response [b]			
Private insurance	360 (73.9 %)	191 (57.0 %)	< 0.001
Self-insured	56 (11.5 %)	28 (8.4 %)	
Medicaid (some Medicare)	39 (8.0 %)	80 (23.9 %)	
Other	22 (4.5 %)	18 (5.4 %)	
Uninsured	7 (1.4 %)	11 (3.3 %)	
Mean Functional Assessment of Chronic Illness Therapy (FACIT) Breast Symptom Index (FBSI) Score [c]	22.1	20.8	0.002
FBSI Items: "Quite a bit" or "very much" response			
I have a lack of energy	174 (36.0 %)	125 (38.4 %)	0.021
I have pain	92 (28.0 %)	92 (19.1 %)	0.051
I have nausea	32 (6.6 %)	37 (11.3 %)	< 0.001
I have certain parts of my body where I experience pain	132 (27.4 %)	119 (35.8 %)	0.073
I have been short of breath	35 (7.3 %)	36 (10.9 %)	0.002
I worry that my condition will get worse	154 (32.0 %)	106 (32.1 %)	0.011
Because of my physical condition, I have trouble meeting the needs of my family	53 (11.0 %)	60 (18.1 %)	0.006
I am content with the quality of my life right now	253 (52.4 %)	190 (57.8 %)	0.375

[a] Information was missing for two non-Hispanic white respondents and six respondents from all other racial/ethnic groups.

[b] Information was missing for three non-Hispanic white respondents and seven respondents from all other racial/ethnic groups.

[c] Where the lowest possible score, zero, indicates that the patient is severely symptomatic and the highest possible score, 32, indicates that the patient is asymptomatic.