



Comparison of treatment of early-stage breast cancer among Nurses' Health Study participants and other Medicare beneficiaries

Andrea M. Austin¹  · Nirav S. Kapadia^{1,2,3} · Gabriel A. Brooks^{1,2,3} · Tracy L. Onega^{1,3} · A. Heather Eliassen^{4,5} · Rulla M. Tamimi^{4,5} · Michelle Holmes⁵ · Qianfei Wang¹ · Francine Grodstein^{4,5} · Anna N. A. Tosteson^{1,2,3}

Received: 14 September 2018 / Accepted: 12 December 2018 / Published online: 3 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Purpose Increasingly epidemiological cohorts are being linked to claims data to provide rich data for healthcare research. These cohorts tend to be different than the general United States (US) population. We will analyze healthcare utilization of Nurses' Health Study (NHS) participants to determine if studies of newly diagnosed incident early-stage breast cancer can be generalized to the broader US Medicare population.

Methods Analytic cohorts of fee-for-service NHS–Medicare-linked participants and a 1:13 propensity-matched SEER–Medicare cohort (SEER) with incident breast cancer in the years 2007–2011 were considered. Screening leading to, treatment-related, and general utilization in the year following early-stage breast cancer diagnosis were determined using Medicare claims data.

Results After propensity matching, NHS and SEER were statistically balanced on all demographics. NHS and SEER had statistically similar rates of treatments including chemotherapy, breast-conserving surgery, mastectomy, and overall radiation use. Rates of general utilization include those related to hospitalizations, total visits, and emergency department visits were also balanced between the two groups. Total spending in the year following diagnosis were statistically equivalent for NHS and SEER (\$36,180 vs. \$35,399, $p=0.70$).

Conclusions NHS and the general female population had comparable treatment and utilization patterns following diagnosis of early-stage incident breast cancers with the exception of type of radiation therapy received. This study provides support for the larger value of population-based cohorts in research on healthcare costs and utilization in breast cancer.

Keywords Breast cancer · Epidemiology · Nurses' Health Study · Generalizability study

Introduction

It is important to understand the spectrum of care before and after diagnosis of breast cancer, which one in eight women in the United States (US) will experience in their lifetimes [1]. Categorizing patterns of care may provide greater insight into the disease and treatment course, allowing for improved quality, cost, or survival outcomes. Increasingly, investigators are utilizing linkages of population-based cohorts and administrative claims data to efficiently consider multiple aspects of health and health services with regard to complex diseases, providing a tremendous range of data—from behavioral, lifestyle, and quality of life data before and after cancer diagnosis, to care utilization, treatments, and expenditures. However, many population-based cohorts focus on highly specific populations and usually include participants who may differ in several ways from the general population;

✉ Andrea M. Austin
andrea.m.austin@dartmouth.edu

¹ The Dartmouth Institute for Health Policy and Clinical Practice at the Geisel School of Medicine at Dartmouth, Hanover, NH, USA

² Department of Medicine, Dartmouth-Hitchcock Medical Center, Lebanon, NH, USA

³ Norris Cotton Cancer Center, Lebanon, NH, USA

⁴ Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

⁵ Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA

this may influence their healthcare utilization patterns and potentially influence the generalizability of health services research within such populations.

Thus, to address breast cancer care in a cohort study compared to the general population, we utilized the linkage of the long-standing Nurses' Health Study (NHS) cohort with Medicare claims data to understand the differences and similarities in healthcare utilization between NHS and the general female Medicare fee-for-service (FFS) population with early-stage breast cancer. We focused here on breast cancer since it is the most common cancer in US women, after skin cancer. We identified women with incident early-stage breast cancer in the Surveillance, Epidemiology, and End Results-18-Medicare (SEER) database and compared their utilization near diagnosis to women in the NHS cohort. Patient preference plays a key role in treatment decision making for early-stage breast cancer in particular. We hypothesized that utilization patterns would be similar between the two groups, except with respect to screening, where previous work has shown NHS have higher adoption [2].

Methods

Data sources

Nurses' Health Study (NHS)

The NHS is based at Brigham and Women's Hospital in Boston, in the Channing Division of Network Medicine. The cohort was established when 121,700 female registered nurses, aged 30–55 years, returned a mailed questionnaire in 1976. Women were identified from 11 states around the US, including all primary geographic regions. Follow-up of the original cohort is ongoing via biennial mailed questionnaires to update lifestyle and health information. To date, follow-up of participants is over 90%. In 2000, NHS were asked to provide social security numbers (SSNs) to allow for linkage of participant data with other datasets, and SSNs were identified for all but 2000 participants.

Medicare linkage

Prior to initiating the Medicare linkage, all NHS were given the opportunity to opt out of the linkage; 390 women indicated they did not want the linkage done and were not included. The NHS dataset was linked to the Medicare claims using a crosswalk from an NHS identifier to a Medicare beneficiary identifier. There were 97,729 women from the NHS who were linked to the 2006–2012 Medicare claims. Of these, 1524 women had a primary incident diagnosis of breast cancer between the years 2007 and 2011, where the cancer was self-reported and subsequently confirmed with a medical record review [3].

SEER–Medicare-linked database

The SEER program of the National Cancer Institute consists of population-based tumor registries serving 18 geographic areas, encompassing 28% of the population, in the US, including San Francisco–Oakland, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle–Puget Sound, Utah, Georgia (Metropolitan Atlanta, Rural and Greater Georgia), San Jose–Monterey, Los Angeles, Alaska, Greater California (excluding SF/SJM/LA), Kentucky, Louisiana, and New Jersey. These cancer registries routinely collect cancer incidence and survival data including patient demographics and clinical factors. SEER data and Medicare data are then linked based on an algorithm involving SSN, name, sex, and date of birth, allowing for the analysis of healthcare utilization with in-depth clinical measures for cancer patients [4]. We were thus able to identify women from the general population with a breast cancer diagnosis, and detailed clinical information.

Cohort definition

The eligible NHS participants were between 66 and 90 years of age, diagnosed with breast cancer between 2007 and 2011, and had staging information available from medical records that NHS routinely collects. We excluded women with a prior history of another primary cancer (including breast). To ensure we could capture their complete care during the year prior to and following their breast cancer diagnosis, we further limited our query to participants living in the US enrolled in Medicare Parts A and B without concurrent enrollment in an HMO (FFS) both for the year prior to and the year following their diagnosis, or until death. This resulted in an NHS cohort of 759 women with incident breast cancer of any stage. To increase the homogeneity of therapy types, as middle and advance stage therapy patterns are more complex and varied, we limited the subset to women with early-stage cancer. Focusing on early-stage cancer also allowed us to analyze differences in preference-sensitive treatments of breast-conserving surgery (BCS) and mastectomy. Thus, we excluded 44 women with stage 3 or 4 disease, resulting in 715 NHS with early-stage (stages 0–2) breast cancer for the final analytic cohort.

Similarly, we identified a cohort of women from the SEER–Medicare database diagnosed with early-stage breast cancer, between 2007 and 2011. Women were included if they were 66 years or older at the time of their diagnosis and had a pathologically confirmed diagnosis as defined by SEER [5]. Also, we required the women to be FFS and living in the US during the study period. We again excluded women with a prior history of another primary cancer. As there were few women (< 11) from the NHS cohort identified as Medicare–Medicaid dual eligible, we excluded all dually eligible women from both cohorts.

Breast cancer identification and staging definitions

We chose to use the American Joint Committee on Cancer (AJCC) TNM staging system as the basis for stage classification, which takes into account the size of the tumor (T), spread to regional lymph nodes (N), and the presence of metastatic disease (M). The AJCC staging definition uses both clinical and pathological classification including information from the physical exam, imaging, diagnostic biopsies, operative findings, and pathology report from definitive surgery. Because SEER and NHS encoded these variables slightly differently, we used the classification algorithm shown in [Appendix Table 4](#) to harmonize staging classifications across the two sources of data.

Demographic covariates

For both cohorts, we obtained demographic information including age at diagnosis (categorized as 66–74, 75–84, 85+), race (collapsed to white and non-white due to sample size constraints), death in year following diagnosis (mortality), diagnosis year, and zip code from their Medicare claims. We measured comorbidities using the Klabunde modification of the Charlson Index applied to the Medicare claims [6–8]. The Klabunde Index is a validated comorbidity index applied to cohorts of cancer patients [6]. Zip code was paired to US Census data to obtain a proxy for socioeconomic status, defined as the median household income of the woman's zip code, and to define their geographic region. Marital status was obtained from the NHS questionnaire closest to year of diagnosis and the SEER Registry for the respective cohorts.

Outcome measures

We characterized general and treatment-related healthcare use. Additionally, we examined women's overall use of inpatient and outpatient services, and the total costs accrued in this same time period. To characterize the utilization patterns that may have led to early diagnosis, we also studied the 12 months prior to diagnosis to obtain information on screening practices.

Cancer-related healthcare utilization

As cancer diagnosis date from the NHS and SEER databases is defined by month and year, we assigned the date of diagnosis to the first day of each month. The outcomes of interest prior to a breast cancer diagnosis were having at least one screening or diagnostic mammogram in the year prior to diagnosis or breast MRI or biopsy at the time of diagnosis.

The treatment/post-diagnosis window was defined as the year following the date of diagnosis. We considered surgical treatments including mastectomy and BCS. We also considered

non-surgical treatments including chemotherapy and radiation therapy. Radiation therapy was further sub-classified into three groups: (1) hypofractionated (short course) external beam radiation, (2) standard (long course) radiation, and (3) brachytherapy, the internal placement of temporary radiation sources [9]. To account for various dose and fractionation regimens, we defined hypofractionated radiation therapy as a course of radiotherapy consisting of treatment with 22 fractions or fewer [10, 11]. See [Appendix Table 5](#) for definitions.

General utilization and costs

Physician services were derived from the Carrier/Part B data and included services as categorized by the Berenson–Eggers Type of Service (BETOS) code classification system. This included claims for durable medical equipment (DME), evaluation and management (E&M) visits (where a patient directly interacts with a clinician), visits for imaging, labs, procedures, and all others. We additionally considered emergency department (ED) visits not leading to a hospitalization. We used claims from the Medicare Provider Analysis and Review (inpatient services, MedPAR) to determine inpatient utilization including total discharges, discharges by type (medical or surgical), ambulatory care sensitive condition (ACSC) hospitalizations, and the mean length of stay for each hospitalization.

Costs were evaluated by BETOS code in Part B and overall by Medicare file type including total, hospital (MedPAR), physician (Part B total), outpatient facility, hospice, home health (HHA), and DME. We used the actual payment from each claim, without standardizing the payments to preserve consistency between the typical Medicare claims used to obtain the outcomes for the NHS and those from the SEER–Medicare database.

Risk-adjustment and statistical analysis

First, propensity scores were generated to match NHS to the SEER–Medicare population. We used multivariate logistic regression to regress cohort status (NHS vs. SEER) on age, race, census region, marital status, mortality, zip code level median household income, Klabunde comorbidity index, year of diagnosis, and stage at diagnosis.

From the model, we obtained the predicted probability, or propensity score, performed a transformation using the logit function, and matched patients from each cohort who had “close” logit-transformed propensity scores. An optimal caliper (maximum distance allowed between the transformed propensity scores from each cohort) was used to obtain matches; it was equal to 0.2 times the standard deviation of the logit of the propensity scores [12–14]. A nearest-neighbor greedy algorithm was used to obtain matches, without replacement, between the NHS and SEER cohorts.

Since there were tens of thousands of women who met the cohort criteria in SEER, in order to obtain as large and complete a comparative sample as possible, we performed a 1:*N* propensity matching where *N* was determined by the first iteration in which we were unable to obtain a 100% match. Covariate imbalance after matching was assessed with a cutoff of a standardized difference greater than (or less than for negative standardized differences) or equal to ± 0.10 to determine if the cohorts remained unbalanced. After propensity score matching was complete, outcomes of interest for the NHS and SEER-matched sample of women were compared using a two-sided *t* test (for continuous variables) or χ^2 test (for categorical variables).

Analyses were conducted using SAS software v14.2 and STATA v15.

Results

For the SEER–Medicare FFS population, there were 37,373 women identified with early-stage breast cancer (Table 1) and 715 NHS from 2007 to 2011. Before propensity matching, NHS were not statistically different than the general SEER population on many characteristics including age, comorbidity level, and zip code median household income. NHS were more likely to be white and be concentrated in the Northeast. NHS were more likely to be diagnosed with stage 1 breast cancer than SEER, with lower rates of stages 0 and 2 diagnoses.

Propensity matching yielded a 1:13 match, with a 100% match rate, between the NHS and SEER cohorts, with each NHS participant assigned 13 similar SEER cases. We ended

Table 1 Characteristics of Nurses' Health Study (NHS) and SEER–Medicare cohorts, before and after propensity score matching

Characteristics	NHS		SEER		Standardized difference: SEER versus NHS	SEER matched		
	<i>N</i> /mean	%/STD	<i>N</i> /mean	%/STD		<i>N</i> /mean	%/STD	Standardized difference: SEER matched versus NHS
<i>N</i>	715		37,373			9295		
Age at diagnosis	74.96	5.87	75.42	6.85	0.07	75.01	6.24	0.00
66–74	355	49.65	18,924	50.64	0.02	4653	50.06	0.02
75–84	312	43.64	14,135	37.82	–0.12	3997	43.00	–0.02
85+	48	6.71	4314	11.54	0.17	645	6.94	0.01
Median household income of zip code	\$67,435	\$25,439	\$65,218	\$26,769	–0.08	\$67,520	\$28,025	–0.02
Klabunde comorbidity score	0.41	0.87	0.50	0.92	0.10	0.41	0.81	0.00
0	516	72.17	25,473	68.16	–0.09	6721	72.31	0.00
1–2	184	25.73	10,733	28.72	0.07	2381	25.62	0.00
3+	15	2.10	1167	3.12	0.06	193	2.08	0.00
Geographical region								
Northeast	316	44.20	8025	21.47	–0.50	4108	44.20	0.00
South	194	27.13	9033	24.17	–0.07	2604	28.02	0.02
Midwest	92	12.87	4666	12.48	–0.01	1187	12.77	0.00
West	113	15.80	15,649	41.87	0.60	1396	15.02	–0.02
White	694	97.06	33,717	90.22	–0.28	8986	96.68	–0.02
1-year mortality	11	1.54	946	2.53	0.07	152	1.64	0.01
Married	422	59.02	18,545	49.62	–0.19	5397	58.06	–0.02
Stage at diagnosis								
0	128	17.90	7460	19.96	0.05	1665	17.91	0.01
1	395	55.24	19,051	50.98	–0.09	5183	55.76	0.01
2	192	26.85	10,862	29.06	0.05	2447	26.33	–0.02
Year of diagnosis								
2007	131	18.32	7332	19.62	0.03	1689	18.17	–0.01
2008	145	20.28	7469	19.99	–0.01	1904	20.48	0.01
2009	147	20.56	7471	19.99	–0.01	1902	20.46	–0.01
2010	153	21.40	7425	19.87	–0.04	2026	21.80	0.01
2011	139	19.44	7676	20.54	0.03	1774	19.09	0.00

up with a total of 10,010 women with early-stage breast cancer, 715 of whom were NHS, and the remaining 9295 in SEER. With the propensity matching, we obtained statistical balance as determined by the standardized differences in all observed covariates (Table 1).

Healthcare utilization by the NHS with incident breast cancer and SEER was similar for most measures associated with breast cancer staging and treatment (Table 2). Specifically, no statistically significant differences were seen in ED use (1.5 vs. 1.5 visits, $p=0.74$) or E&M visits (22.1 vs. 21.5, $p=0.30$). Inpatient hospitalizations for ACSC (0.03 vs. 0.04, $p=0.82$), medical discharges (0.17 vs. 0.19, $p=0.30$), and surgical discharges (0.29 vs. 0.28, $p=0.58$) were also statistically similar. Treatment-related utilization, including rates of breast biopsy (97.1% vs. 96.6%, $p=0.55$), mastectomy (26.4% vs. 25.4%, $p=0.55$), BCS (71.1% vs. 70.5%, $p=0.77$), radiation therapy (61.1% vs. 59.5%, $p=0.39$), and chemotherapy (16.1% vs. 15.0%, $p=0.43$), had similar rates of usage among both cohorts.

We observed modest differences in utilization between the NHS and SEER cohorts in use of mammography (diagnostic and screening), MRI, and in radiation course length. A greater proportion of NHS received screening (61.5% vs. 52.8%, $p<0.0001$) and diagnostic (71.5% vs. 60.4%, $p<0.0001$) mammogram and had any mammogram in the year prior to diagnosis (77.5% vs. 65.4%, $p<0.0001$), and slightly more received breast MRI (31.8% vs. 27.1%, $p=0.01$) than the SEER population. Among women who received radiotherapy as part of their treatment, substantially more NHS received standard course radiotherapy (32.0% vs. 9.6%) and fewer received hypofractionated radiotherapy (54.9% vs. 79.6%, $p<0.0001$). This was reflected by more procedural visits, on average (33.5 vs. 28.2, $p<0.0001$) and higher overall per capita procedural costs (\$5633 vs. \$4485, $p<0.0001$) for NHS compared to the matched SEER cohort.

Overall payments in the year following diagnosis were similar for NHS, \$36,180, and the matched SEER cohort, \$35,399 ($p=0.70$). However, the costs associated with payments from the Carrier/Part B file were somewhat higher for NHS (\$12,200 vs. \$10,825, $p=0.002$) and when broken down by BETOS code, this difference was mostly accounted for by higher procedural costs, as noted above (Fig. 1), among the NHS than in the SEER-matched cohort.

Costs by radiation course are shown in Table 3. The difference in Part B spending between NHS and SEER is in part accounted for by differences in costs among women without radiation. The NHS cohort had higher costs for those without radiation in the year following diagnosis (\$9330 vs. \$7814, $p=0.02$). The difference in spending among women without radiation is caused by higher spending on both procedures (\$2694 vs. \$2306, $p=0.004$) and laboratory-related claims (\$1814 vs. \$1441, $p=0.001$) for NHS compared to the matched SEER cohort. Average spending among women

receiving standard radiation courses was balanced between the NHS and SEER (\$20,507 vs. \$21,643, $p=0.35$).

Discussion

We compared healthcare utilization between NHS and women in the general SEER–Medicare-linked registry with early-stage incident breast cancer. We found no meaningful differences in utilization between the two cohorts in most measures examined, including the use of hospital services and most procedures, including biopsy, mastectomy, BCS, chemotherapy, and overall radiotherapy use. These findings, in a cohort of older nurses, provide reassuring evidence that population studies may provide an important resource for health services research, especially in combination with administrative databases such as Medicare.

We observed greater utilization of mammography, both screening and diagnostic, among NHS, consistent with previously described differences in utilization patterns between NHS and the general Medicare population [2]. In addition, we observed differences in the length of radiation therapy course. Our findings suggest NHS participants may not be early adopters of treatment innovations. We posit there is a great likelihood that nurses may seek out providers they know who may be older. Research has shown more recently trained physicians are more likely to quickly adopt new standards of care [15]. However, overall utilization of radiation of any course is balanced among the NHS and SEER cohorts. Among NHS who received external beam radiation treatments, 32% received standard course radiation as compared to 10% of SEER. This difference in utilization translated into higher average outpatient procedural costs and number of procedures, although overall payments in the year following diagnosis were similar in the NHS and SEER cohorts. The differences in procedural costs between the NHS and SEER cohorts were most largely driven by just two factors: (1) radiation fractionation schedules in NHS and (2) spending on procedures and laboratory tests in NHS without radiation.

Many studies have suggested an association between high use of mammography and an increase in early-stage breast cancer rates [16–18]. Perhaps not surprisingly then, the NHS had both greater mammography use and greater stage 1 diagnoses than SEER. The uptake and costs associated with hypofractionated breast radiation compared to standard length therapy during a similar timeframe, 2008–2013, have been previously studied showing adoption rates up to 35% in the time period [19]. We saw similar rates of overall usage among NHS (34%) and higher adoption among SEER (47%). The cost of each regimen showed patterns resembling those observed in our study, with standard course radiation yielding greater total healthcare costs [19]. This study highlights our

Table 2 Utilization of healthcare by NHS participants and matched SEER–Medicare beneficiaries for those diagnosed with incident early-stage breast cancer

	NHS		SEER matched		<i>p</i> values
	<i>N</i> /mean	%/StdDev	<i>N</i> /mean	%/StdDev	
Screening and diagnosis					
Screening or diagnostic mammography	554	77.48	6082	65.43	<0.0001
Breast biopsy	694	97.06	8983	96.64	0.55
Breast MRI	227	31.75	2489	26.78	0.004
Treatment					
Chemotherapy	115	16.08	1393	14.99	0.43
Breast-conserving surgery	508	71.05	6555	70.52	0.77
Mastectomy	189	26.43	2363	25.42	0.55
Radiation course					
No	278	38.88	3765	40.51	0.39
Yes ^a	437	61.12	5530	59.49	
Hypofractionated	240	54.92	4402	79.60	<0.0001 ^b
Standard	140	32.04	533	9.64	
Brachytherapy	57	13.04	595	10.76	
ACSC hospitalizations	0.03	0.19	0.04	0.23	0.82
Length of hospitalization ^c	4.13	6.28	5.09	7.40	0.02
Number of ED visits ^d	1.46	0.88	1.48	1.17	0.74
Number of visits					
Durable medical equipment	7.57	15.69	8.29	18.97	0.25
E&M	22.12	15.29	21.51	13.63	0.30
Imaging	15.14	10.32	14.28	8.32	0.03
Lab	26.89	22.61	30.53	25.81	<0.0001
Procedures	33.53	28.37	28.24	24.86	<0.0001
Hospital discharges					
Medical	0.17	0.52	0.19	0.58	0.30
Surgical	0.29	0.51	0.28	0.51	0.58
Total	0.47	0.79	0.48	0.85	0.79
Cost by Medicare file source					
DME	\$121	376.13	\$217	691.71	<0.0001
HHA	\$462	1577.55	\$572	1811.18	0.08
Hospice	\$21	267.12	\$97	1479.45	<0.0001
Outpatient	\$8708	9657	\$7941	7989	0.04
MedPAR	\$14,669	49,963	\$15,748	43,160	0.57
Part B/Carrier	\$12,200	11,268	\$10,825	10,385	0.002
Total	\$36,180	53,504	\$35,399	47,311	0.70
Part B costs by BETOS ^e code					
DME	\$2135	7386	\$2137	7456	0.99
E&M	\$1588	1243	\$1593	1137	0.91
Imaging	\$1000	1111	\$931	929	0.10
Lab	\$1843	1833	\$1679	1605	0.02
Procedures	\$5634	5505	\$4485	4021	<0.0001

^aThe percentage of radiation that is intensity-modulated radiation therapy is 11.75% among NHS participants and 11.72% among the matched SEER–Medicare cohort

^bAmong those with some radiation

^cAmong those with at least one hospitalization

^dAmong those with at least one emergency department visit

^eBerenson–Eggers Type of Service classification for Health Care Financing Administration Procedure Coding System (HCPCS) codes

Fig. 1 Average spending on physician services in the year following diagnosis, by BETOS code, for NHS and the matched SEER–Medicare cohorts



Table 3 Spending on physician services by radiation course, by BETOS code, for NHS and the matched SEER–Medicare cohorts

	Radiation course											
	None			Hypofractionated			Standard			Brachytherapy		
	NHS	SEER	<i>p</i> values	NHS	SEER	<i>p</i> values	NHS	SEER	<i>p</i> values	NHS	SEER	<i>p</i> values
<i>N</i>	278	3765		240	4402		140	533		57	595	
Part B costs by BETOS, mean												
Procedures	\$2694	\$2306	0.004	\$4227	\$4924	<0.0001	\$13,041	\$13,685	0.31	\$7696	\$6780	0.23
DME	\$2355	\$1786	0.24	\$1466	\$2457	0.01	\$2609	\$3027	0.58	\$2708	\$1192	0.25
E&M	\$1648	\$1532	0.27	\$1545	\$1664	0.04	\$1594	\$1535	0.45	\$1464	\$1512	0.66
Imaging	\$820	\$750	0.16	\$887	\$1010	0.06	\$1556	\$1492	0.68	\$991	\$989	0.98
Lab	\$1814	\$1441	0.001	\$1876	\$1833	0.73	\$1706	\$1904	0.19	\$2183	\$1841	0.25

findings that radiation therapy duration has an influence on medical costs during treatment of early-stage breast cancer.

One limitation of our study is that the SEER–Medicare database has encrypted beneficiary identifiers that differ from the beneficiary identifiers in the standard Medicare analytic files. Thus, there is no way to know if an NHS participant is also present in the SEER Registry. Thus, we constructed a 1:13 match to obtain a large enough comparator sample to reduce the influence of any women who were present in both the NHS and matched SEER cohorts. Indeed, since NHS represent a very small proportion of the overall population, it is highly unlikely that they have any meaningful representation in the comparison group we selected from SEER. Another limitation is the somewhat limited distribution of socioeconomic status in NHS (although median household income was similar in NHS and SEER–Medicare even prior to propensity matching) and the small number of minority participants; thus, we cannot provide information on breast cancer care in those groups. Additionally, there is some degree of misclassification with splitting mammography use into screening and diagnostic; however,

we expect the misclassification to be similar between the two cohorts. Another limitation of combining the data sources for the two cohorts is the modification of the standard TNM staging to make accurate comparisons between the two groups of women, losing some granularity of the stage in the process. Finally, a limitation of both the NHS and SEER data sources is with the date of diagnosis. Each source only provides the month of diagnosis. For consistency, we placed the diagnosis date on the first day of the diagnosis month.

Overall, our study generally provides support for the larger value of population-based cohorts in research on healthcare costs and utilization in breast cancer. This reinforces previous studies showing that healthcare utilization for NHS is largely similar to the general Medicare FFS female population [2, 3]. While we observed some higher utilization of radiation therapy and breast cancer screening among NHS, overall differences were small. The ability to leverage the extensive work done to characterize the health and behaviors of the NHS around breast cancer risk, treatment, and utilization [20–28] to the general US female population and augment this work with extensive utilization detail

from the Medicare claims is invaluable to continuing health services research. More importantly, these findings strongly suggest that linkage of administrative claims with population-based cohorts can provide meaningful information on breast cancer care and survivorship for the broader population. Future studies may address similar research in cohorts with larger minority populations.

Acknowledgements We would like to thank the participants and staff of the Nurses’ Health Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

Author Contributions All authors contributed to the study conception and design; AMA and QW analyzed the data. AMA, ANAT, and FG prepared the manuscript. All authors provided critical revisions to the manuscript.

Funding Supported by the National Cancer Institute (Grant Number UM1CA186107).

Compliance with ethical standards

Conflict of interest All the authors declare that they have no conflict of interest.

Research involving human and animal participants This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent This study utilizes retrospective, de-identified information. Informed consent was not necessary.

Appendix

See Tables 4 and 5.

Table 4 Stage classification algorithm for NHS and SEER–Medicare databases for early-stage breast cancer

Stages	Cohort	Tumor size (cm)	Node involvement
0	NHS	Not described/in situ	0 (No reported nodes involved)
	SEER	Main tumor cannot be found	0 Nodes
1	NHS	0.1–2.0	0 Nodes
	SEER	≤2.0	0 Nodes
2	NHS	0.1–2.0	1–3 Nodes
		2.1–4.0	0–3 Nodes
		4.0+	0 Nodes
	SEER	≤2.0	1–3 Nodes
		2.0–4.0	0–3 Nodes
		4.0+	0 Nodes

Table 5 Medical codes used to identify the procedures and treatments in the Medicare claims

Measures	Codes
Mammography	Screening—CPT Codes: 77057, G0202 Diagnostic—CPT Codes: 77052, 77055, 77056, G0204, G0206
ED visit	Revenue Center Codes: 0450-0459, 0981
Breast biopsy	ICD9 Codes: 401.1, 402.2, 405.1, 851.1, 851.2, 851.9, 853.1, 853.2, 859.1, 859.9, 873.5, 403, 851, 852 CPT Codes: 10022, 19000, 19001, 19030, 19105, 19110, 19112, 19260, 19271, 19272, 19499, 38500, 38505, 38510, 38525, 38530, 38740, 38792, 76360, 76393, 76942, 77012, 77021, 77031, 77032, 77053, 77054, 0046T, 0047T, 19100–19103, 19120–19126, 19290, 19295, 76087–76089, 76095, 76098
BCS	ICD9 Codes: 852.2, 852.3 CPT Codes: 19120, 19160, 19162, 19301, 19302
Mastectomy	ICD9 Codes: V510, 853.3, 853.4, 853.5, 853.6, 854 CPT Codes: 19180, 19182, 19200, 19220, 19240, 19303–19306
Chemotherapy	ICD9 Codes: V58.1, 9925 CPT Codes: 36640, 51720, 96400, 96405, 96406, 96408, 96410, 96412, 96414, 96420, 96422, 96423, 96425, 96440, 96445, 96450, 96501, 96504, 96505, 96508, 96520, 96524, 96530, 96538, 96540, 96542, 96545, 96549, 96555, 96510–96512, Q0083–Q0085, I9000–I9999 Revenue Center Codes: 0331, 0332, and 0335
Breast MRI	CPT Codes: C8903–C8908, 76093, 76094, 76498, 77058, 77059, 0159T
IMRT	CPT Code: 77418
Radiation	CPT Codes: 77280, 77285, 77290, 77295, 77299, 77300, 77305, 77310, 77315, 77321, 77336, 77370, 77399, 77416, 77417, 77419, 77420, 77470, 77499, 76960, 55859, 55860, 55862, 55865, 77750, 77781, 77782, 77784, 77789, 77790, 77799, 79200, 79300, 79400, 79440, 79900, 79999, 77261–77263, 77326–77328, 77331–77334, 77401–77404, 77406–77409, 77411–77414, 77430–77432
Brachytherapy	CPT Codes: 19296, 19297, 19298, 77761, 77762, 77763, 77785, 77786, 77787, 77776, 77777, 77778, 77424, 77425

References

- National Breast Cancer Foundation, Inc. (2016) Breast Cancer Facts. <https://www.nationalbreastcancer.org/breast-cancer-facts>. Accessed 22 Aug 2018
- Kapadia NS, Austin AM, Carmichael DQ, Wang Q, Ereksion ED, Feskanich D, Grodstein F, Tosteson AN, Bynum JPW (2018) Medicare utilization and spending among nurses compared to the general U.S. population. *J. Women's Health*. <https://doi.org/10.1089/jwh.2017.6802>
- Bronson MR, Kapadia NS, Austin AM, Wang Q, Feskanich D, Bynum JPW, Grodstein F, Tosteson ANA (2018) Leveraging linkage of cohort studies with administrative claims data to identify individuals with cancer. *Med Care*. <https://doi.org/10.1097/MLR.0000000000000875>
- Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF (2002) Overview of the SEER–Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care* 40(8 Suppl):IV-3-18
- National Cancer Institute (2018) Cancer Diagnosis. <https://training.seer.cancer.gov/disease/diagnosis/>. Accessed 22 Aug 2018
- Klabunde CN, Harlan LC, Warren JL (2006) Data sources for measuring comorbidity: a comparison of hospital records and Medicare claims for cancer patients. *Med Care* 44(10):921–928
- Klabunde CN, Potosky AL, Legler JM, Warren JL (2000) Development of a comorbidity index using physician claims data. *J Clin Epidemiol* 53(12):1258–1267
- Klabunde CN, Legler JM, Warren JL, Baldwin LM, Schrag D (2007) A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. *Ann Epidemiol* 17(8):584–590
- Mayo Clinic (2018) Patient Care and Health Info: Brachytherapy. <https://www.mayoclinic.org/tests-procedures/brachytherapy/about/pac-20385159>. Accessed 22 Aug 2018
- Valle LF, Agarwal S, Bickel KE, Herchek HA, Nalepinski DC, Kapadia NS (2017) Hypofractionated whole breast radiotherapy in breast conservation for early-stage breast cancer: a systematic review and meta-analysis of randomized trials. *Breast Cancer Res Treat* 162(3):409–417
- Takagi R, Komiya Y, Sutherland KL, Shirato H, Date H, Mizuta M (2018) Comparison of the average surviving fraction model with the integral biologically effective dose model for an optimal irradiation scheme. *J Radiat Res* 59(suppl_1):i32–i39
- Austin PC (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res* 46(3):399–424
- Austin PC (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 28(25):3083–3107
- Austin PC (2011) Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat* 10(2):150–161
- Boero IJ, Gillespie EF, Hou J, Paravati AJ, Kim E, Einck JP, Yashar C, Mell LK, Murphy JD (2017) The impact of radiation oncologists on the early adoption of hypofractionated radiation therapy for early-stage breast cancer. *Int J Radiat Oncol Biol Phys* 97(3):571–580
- Lynge E, Ponti A, James T, Májek O, von Euler-Chelpin M, Anttila A, Fitzpatrick P, Frigerio A, Kawai M, Scharpantgen A et al (2014) Variation in detection of ductal carcinoma in situ (DCIS) during screening mammography: a survey within the International Cancer Screening Network (ICSN). *Eur J Cancer (Oxf Engl 1990)*. <https://doi.org/10.1016/j.ejca.2013.1008.1013>
- Braithwaite D, Mandelblatt JS, Kerlikowske K (2013) To screen or not to screen older women for breast cancer: a conundrum. *Future Oncol (Lond Engl)* 9(6):763–766
- Welch HG (2010) Screening mammography—a long run for a short slide? *N Engl J Med* 363(13):1276–1278
- Bekelman JE, Sylwestrzak G, Barron J, Liu J, Epstein AJ, Freedman G, Malin J, Emanuel EJ (2014) Uptake and costs of hypofractionated vs conventional whole breast irradiation after breast conserving surgery in the United States, 2008–2013. *JAMA* 312(23):2542–2550
- Lairson DR, Parikh RC, Cormier JN, Chan W, Du XL (2015) Cost-effectiveness of chemotherapy for breast cancer and age effect in older women. *Value Health J Int Soc Pharmacoecon Outcomes Res* 18(8):1070–1078
- Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, Stolk L, Finucane HK, Sulem P, Bulik-Sullivan B et al (2015) Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet* 47(11):1294–1303
- Liu Y, Tobias DK, Sturgeon KM, Rosner B, Malik V, Cespedes E, Joshi AD, Eliassen AH, Colditz GA (2016) Physical activity from menarche to first pregnancy and risk of breast cancer. *Int J Cancer* 139(6):1223–1230
- Sisti JS, Collins LC, Beck AH, Tamimi RM, Rosner BA, Eliassen AH (2016) Reproductive risk factors in relation to molecular subtypes of breast cancer: results from the nurses' health studies. *Int J Cancer* 138(10):2346–2356
- Eliassen AH, Missmer SA, Tworoger SS, Spiegelman D, Barbieri RL, Dowsett M, Hankinson SE (2006) Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. *J Natl Cancer Inst* 98(19):1406–1415
- Huang Z, Willett WC, Colditz GA, Hunter DJ, Manson JE, Rosner B, Speizer FE, Hankinson SE (1999) Waist circumference, waist:hip ratio, and risk of breast cancer in the Nurses' Health Study. *Am J Epidemiol* 150(12):1316–1324
- Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A et al (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39(7):870–874
- Rice MS, Eliassen AH, Hankinson SE, Lenart EB, Willett WC, Tamimi RM (2016) Breast cancer research in the nurses' health studies: exposures across the life course. *Am J Public Health* 106(9):1592–1598
- Holmes MD, Chen WY, Feskanich D, Kroenke CH, Colditz GA (2005) Physical activity and survival after breast cancer diagnosis. *JAMA* 293(20):2479–2486