Research Article

Accuracy of Self-Reported Breast Cancer Information among Women from the Ontario Site of the Breast Cancer Family Registry

Andriana Barisic,[1, 2] Gord Glendon,[1, 3] Nayana Weerasooriya,[3] Irene L. Andrulis,[1, 4] and Julia A. Knight[1, 2]

[1] *Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada M5G 1X5*
[2] *Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada M5T 3M7*
[3] *Ontario Familial Breast Cancer Registry, Princess Margaret Hospital, Toronto, ON, Canada M5T 2M9*
[4] *Departments of Molecular Genetics and Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada M5T 3H7*

Correspondence should be addressed to Julia A. Knight, knight@lunenfeld.ca

Obtaining complete medical record information can be challenging and expensive in breast cancer studies. The current literature is limited with respect to the accuracy of self-report and factors that may influence this. We assessed the agreement between self-reported and medical record breast cancer information among women from the Ontario site of the Breast Cancer Family Registry. Women aged 20–69 years diagnosed with incident breast cancer 1996–1998 were identified from the Ontario Cancer Registry, sampled on age and family history. We calculated kappa statistics, proportion correct, sensitivity, specificity, and positive and negative predictive values and conducted unconditional logistic regression to examine whether characteristics of the women influenced agreement. The proportions of women who correctly reported having received a broad category of therapy (hormone therapy, chemotherapy, radiation, or surgery) as well as sensitivity and specificity were above 90%, and the kappa statistics were above 0.80. The specific type of hormonal or chemotherapy was reported with low-to-moderate agreement. Aside from recurrence, no factors were consistently associated with agreement. Thus, most women were able to accurately report broad categories of treatment but not necessarily specific treatment types. The finding of this study can aid researchers in the use and design of self-administered treatment questionnaires.

1. Introduction

Medical records are considered the gold standard in regards to obtaining accurate information pertaining to breast cancer diagnosis, treatment, and prognosis. However, this method of data collection can be very time consuming and labor intensive. Obtaining complete longitudinal treatment information for individuals over long periods of time can be challenging and expensive, particularly when conducting large population-based studies of breast cancer patients. With an increasing focus on epidemiological studies of prognosis and survivorship in breast cancer, clinical data collection is a methodological issue of rising importance.

Self-reported treatment questionnaires may be a more feasible option in some cases; however, few studies have examined the accuracy of self-reported breast cancer treatment and fewer have considered factors potentially influencing accuracy. To our knowledge, only four studies have specifically examined the accuracy of self-reported breast cancer treatment to date [1–4]. All four studies demonstrated that breast cancer survivors can accurately recall some important information pertaining to their treatment. However, two of the studies had small sample sizes [2, 3], while one study was restricted to low income women in California [1] and only two of the studies assessed factors potentially associated with reporting accuracy. To add to the limited

literature on this topic, we assessed the agreement between self-reported breast cancer treatment and recurrence and medical record information among women in the Ontario Familial Breast Cancer Registry (OFBCR), the Ontario site of the Breast Cancer Family Registry (BCFR). We also examined characteristics of the women that could potentially influence agreement, including breast cancer family history (the study population was enriched for women with a family history of breast cancer) and menopausal status at the time of diagnosis. We also assessed the agreement between menopausal status in medical records and self-report.

2. Methods

2.1. Study Sample. Breast cancer patients included in this analysis were women recruited into the OFBCR. The OFBCR is one of six registries in the Breast CFR, an international collaboration of academic and research institutions located in the United States, Canada, and Australia designed to examine the genetic, environmental, and lifestyle factors associated with breast cancer [5].

In Ontario, women aged 20 to 69 diagnosed with breast cancer between the calendar years of 1996 and 1998 were identified from the population-based Ontario Cancer Registry and recruited between 1996 and 2000 using a sampling strategy incorporating age, ethnicity, and family history of cancer to enrich the registry with genetically predisposed individuals [5]. All women diagnosed under the age of 36 years were eligible for the OFBCR. For women between the ages of 36 and 54 years, all those meeting the high-risk criteria (in the following) or those of Ashkenazi Jewish heritage were eligible. Furthermore, a sample of 25% of the women in this age group who did not meet the high-risk criteria was also eligible. For women between the ages of 55 and 69 years, a sample of 35% of those meeting the high-risk criteria or those of Ashkenazi heritage and a sample of 8.75% (35%× 25%) of those not meeting the high-risk criteria were eligible. High-risk was defined as women with a previous diagnosis of breast or ovarian cancer; at least one first- or two second-degree relative(s) with breast or ovarian cancer; at least one second- or third-degree relative with breast cancer diagnosed before the age of 36 years, ovarian cancer diagnosed before the age 61 years, multiple breast cancer primaries, both breast and ovarian cancer, or male breast cancer; at least three first-degree relatives with any combination of breast, ovarian, colon, prostate, or pancreatic cancer or sarcoma, with at least one diagnosis before the age of 51 years [5].

Using the sampling strategy described above, physicians gave permission to contact 7,662 women of 8,448 women who were identified with incident breast cancer diagnosed between 1996 and 1998. Of these women, 4,962 women completed the Family History Questionnaire (described in the following), and 2,399 women were eligible and willing to proceed. After excluding women for whom blood samples and permission to access medical records were not available (note that in some cases it was because the women had died rather than refused) and those with a prior cancer diagnosis or distant metastases at diagnosis, 1082 women were eligible

for the current study, of which 939 had a first diagnosis of unilateral breast cancer with medical record and treatment questionnaire data available for analysis. The selection of the study sample has been previously described in detail [6] with the exception of the final number of women with medical records available as some were subsequently obtained. In the nearly completely overlapping study population in the previous publication, we showed that there were no significant differences compared to women without a blood sample and/or permission to retrieve medical records or women whose records were unavailable with the exception that the latter group were slightly less likely to be node positive [6].

2.2. Data Collection

2.2.1. Treatment Questionnaire. The self-administered Treatment Questionnaire gathered information regarding aspects of treatment for breast or ovarian cancer. It consists of questions that address stage (whether the cancer was only in the breast or breast and lymph glands or had spread to other sites), the type of initial breast cancer treatments received, and whether recurrence has occurred. Each item related to treatment first asked whether the broad category of treatment was received (e.g., "did you have hormonal therapy for this breast cancer"), and then for those who responded "yes," more specific questions regarding the details of the treatment were asked (e.g., "please list medicines if known"). The question on recurrence asked "has the cancer come back (recurred) after the treatments listed above?".

2.2.2. Additional Questionnaires. The mailed Family History Questionnaire gathered information regarding previous cancer diagnoses in the women, their parents, siblings, children, and any other relatives, and family history was reviewed over the telephone with a genetic counselor. All cancers, except nonmelanoma skin cancers and cervical carcinoma *in situ*, were recorded, along with dates of all cancer diagnoses and deaths [5]. The mailed Personal History Questionnaire gathered information on the period prior to diagnosis pertaining to the demographic characteristics of the women, personal history of cancers, prior breast and ovarian surgeries, prior radiation exposure, smoking and alcohol behavior, reproductive history, breast-feeding, hormones used, height, weight, and physical activities [5].

2.2.3. Medical Record Abstraction. Clinical and treatment data were abstracted from the patient's medical records at each clinic using standard data collection forms. Data collected included the method and date of diagnosis, clinical stage, type of surgical treatment, whether or not radiation treatment was given, dates and types of hormonal therapies, and dates and types of chemotherapy administration. All data were reviewed, verified, and coded centrally by the research nurse coordinator and an oncologist [6]. Follow-up data from medical records were collected annually for breast cancer recurrence, new primary cancers, and death. Vital status was also ascertained from linkage to the Ontario

Cancer Registry which routinely obtains death certificate information [6].

2.3. Statistical Methods. Agreement analysis was conducted with analysis restricted to the most common types of treatment reported in both the Treatment Questionnaire and in the medical records. Women who had start dates for hormonal therapy, chemotherapy, and surgery were coded as having received the therapy in question, while women without respective start dates were coded as not having received the therapy (based on the assumption that a start date indicated that a prescribed treatment was actually administered). Any treatment that occurred after the completion date of a woman's respective treatment questionnaire was coded as not having occurred (since it would have occurred after self-report). For radiation therapy, no treatment dates were available to us in the study database; accordingly we could not determine whether any treatments occurred after the treatment questionnaire completion date. Thus, analysis regarding radiation therapy was conducted excluding women with breast cancer recurrence at the time of questionnaire completion (based on medical record abstraction) since it is likely that this was the only situation where radiation might have occurred following the questionnaire, which was completed at least a year after diagnosis. In regards to whether women could accurately report the specific type of hormonal therapy or chemotherapy medication they received (alone and/or in combination with other medications), agreement analysis was restricted to only those women who were coded as having received the respective broader category of treatment (i.e., hormonal therapy or chemotherapy).

Agreement between self-reported treatment and breast cancer recurrence and medical records was first assessed by calculating the kappa statistic, which takes chance agreement into account. Next, using medical records as the gold standard, agreement was also analyzed by calculating *the proportion correct* (the percentage of women who correctly identified whether they received a given treatment or not), *sensitivity* (the proportion of women who correctly report having received a given treatment) and *specificity* (the proportion of women who correctly report not having received a given treatment), and *positive predictive values* and *negative predictive values* where appropriate. With respect to menopausal status, the Personal History Questionnaire asked women to identify the age and the reasons why their periods stopped. Based on their responses, women were categorized as being premenopausal, postmenopausal, or as unknown at the time of diagnosis, where unknown status applied to women who had undergone a hysterectomy without a bilateral oophorectomy, or were on hormonal replacement therapy at the time of their diagnosis. As there is no gold standard, only the proportion in agreement and the kappa statistic were calculated. For treatment, two models of agreement analysis were conducted. The first model excluded unknown and missing responses, while the second model was based on the classification system suggested by Landis and Koch (1977) [7], where all self-reported "unknown" and missing responses were recoded

as "no." The first model provided a more conservative estimate of agreement in terms of the proportion correct and accordingly was the model used to report results of the agreement analyses and for all subsequent regression analyses. Results were similar as there were few missing responses. For menopausal status where there is a larger proportion of unknown and missing information, agreement analysis was conducted treating this variable as both 2-level (excluding unknown/missing information) and 3-level variables (including unknown/missing information).

Lastly, unconditional logistic regression analysis was conducted to examine the influence of the following on agreement: age at breast cancer diagnosis (continuous), recall period (continuous), education level, marital status, alcohol consumption (continuous), smoking behavior, first-degree family history of breast or ovarian cancer, menopausal status at diagnosis, breast cancer recurrence, and place of birth. In regards to place of birth, women were categorized as having been born in an English or non-English speaking country which was meant to serve as a proxy for proficiency in English. Logistic regression analysis was conducted to examine the factors associated with agreement for the broad categories of hormonal therapy, chemotherapy, and radiation, and, as suggested by Lipstiz et al. (2003) [8], an offset controlling for agreement due to chance was included in all logistic regression models. Due to the limited number of women who did not have surgery for their breast cancer, regression analysis was not conducted examining the factors associated with whether women could accurately report whether they received surgery or not. All statistical analyses were conducted using SAS Enterprise Guide, version 4.2 OnDemand.

3. Results

Descriptive statistics are displayed in Table 1, and as displayed in Table 2, according to the medical records, approximately half of the women (54.6%, $n = 513$) received hormonal therapy and chemotherapy (53.2%, $n = 500$) respectively, while over three quarters of the women (76.2%, $n = 712$) received radiation therapy, all but three of the women ($n = 936$) had surgery for her breast cancer, and 5.3% of the women experienced breast cancer recurrence ($n = 50$) (prior to the completion of the treatment questionnaire). The overall proportions based on self-reported information were similar (Table 2).

3.1. Results of Agreement Analysis by Broad Category. Agreement between medical records and self-reported treatment is summarized in Table 3.

Overall the proportion of women who correctly reported having received a broad category of therapy (hormone therapy, chemotherapy, radiation, or surgery) was above 90% for each therapy. Sensitivity and specificity along with positive and negative predictive values were also above 90%, and the kappa statistic was above 0.80 for each category, with the exception of surgery, which is likely attributed to the low number of women who did not have surgery for their breast

TABLE 1: Characteristics of the study population of women with breast cancer from questionnaire information.

	Subjects included in analysis ($n = 939$)	
	No.	%
Age at diagnosis (years)		
20–29	13	1.4
30–39	139	14.8
40–49	342	36.4
50–59	315	33.6
60–69	130	13.8
Range	23–69	
Mean (Std)	48.7 (9.19)	
Median	49.0	
Year of breast cancer diagnosis		
1996	282	30.0
1997	329	35.1
1998	328	34.9
Recall period (years)		
1–5	896	95.4
>5	43	4.6
Mean (Std)	3.0 (1.2)	
Median	3.0	
Marital status		
Married/living as married	734	78.2
Widowed/divorced/separated/never married	205	21.8
Educational status		
High school or less	358	38.3
Some vocational/technical/ college/university	334	35.8
Bachelor or graduate degree	242	25.9
Missing	5	n/a
Born in English speaking country		
Yes	825	88.1
No	111	11.9
Missing	3	n/a
Ethnic background		
White	901	96.5
Other	31	3.3
Don't know	2	0.2
Missing	5	n/a
Have you ever consumed alcohol?[1]		
Yes	462	49.4
No	474	50.6
Missing	3	n/a

TABLE 1: Continued.

	Subjects included in analysis ($n = 939$)	
	No.	%
Number of drinks per week		
0	495	52.7
1–4	218	23.2
5–9	145	15.4
10–19	55	5.9
20–29	16	1.7
30–39	9	1.0
40–49	1	0.1
Have you ever smoked?[2]		
Yes	473	50.4
No	465	49.6
Missing	1	n/a
Smoking status[2]		
Current smoker	187	19.9
Former smoker	280	29.8
Never smoker	465	49.5
Don't know	6	0.8
Missing	1	n/a
Breast cancer recurrence		
Yes	50	5.3
No	889	84.7
First degree relative with history of breast or ovarian cancer		
Yes	341	36.3
No	598	63.7
Menopausal status at diagnosis		
Pre-menopausal	520	55.4
Post-menopausal	248	26.4
Unknown—on menopausal hormone therapy	59	6.3
Unknown—hysterectomy	104	11.1
Missing	8	0.8

[1] More than 1 drink per week on average for a period of at least 6 months.
[2] Refers to cigarette smoking.

cancer. Since the kappa statistic takes chance agreement into account, limited variability increases the likelihood that agreement is due to chance alone. Furthermore, there was moderate agreement between medical records and self-reported information regarding whether women experienced breast cancer recurrence or not. Although 94.8% of women correctly reported whether they experienced breast cancer recurrence, sensitivity was only 75.5%, indicating that a quarter of the women experienced breast cancer recurrence according to their medical records but did not report so. In regards to menopausal status, using the more conservative results (including unknown responses and treating menopausal status as a 3-level variable), there was 83.0% agreement between medical records and self-report and

TABLE 2: Comparison of treatment, recurrence, and menopausal information obtained from medical records and self-administered questionnaire in women with breast cancer.

	From medical records			From treatment questionnaire	
	Frequency ($n = 939$)	Percent (%)		Frequency ($n = 939$)	Percent (%)
Hormone given			Hormone taken		
No	426	45.4	No	434	46.8
Yes	513	54.6	Yes	491	52.9
Tamoxifen			Tamoxifen		
Yes	468	91.2	Yes	441	89.8
No	45	8.8	No	50	10.2
Megestrol			Megestrol		
Yes	16	3.1	Yes	5	1.0
No	497	96.9	No	486	99.0
Anastrozole			Anastrozole		
Yes	67	13.1	Yes	16	3.3
No	446	86.9	No	475	96.7
MA12			MA12		
Yes	22	4.3	Yes	5	1.0
No	491	95.7	No	486	99.0
Don't know	0	0	Don't know	3	0.3
Missing	0	0	Missing	11	n/a
Chemotherapy given			Chemotherapy taken		
No	439	46.8	No	408	43.7
Yes	500	53.2	Yes	525	56.3
Epirubicin			Epirubicin		
Yes	188	37.6	Yes	113	21.5
No	312	62.4	No	412	78.5
Doxorubicin			Doxorubicin		
Yes	148	29.6	Yes	84	16.0
No	352	70.4	No	441	84.0
Methotrexate			Methotrexate		
Yes	213	42.6	Yes	121	23.0
No	287	57.4	No	404	77.0
Cyclophosphamide			Cyclophosphamide		
Yes	495	99.0	Yes	307	58.5
No	5	1.0	No	218	41.5
Fluorouracil			Fluorouracil		
Yes	430	86.0	Yes	260	49.5
No	70	14.0	No	265	50.5
Docetaxel			Docetaxel		
Yes	59	11.8	Yes	5	0.9
No	441	88.2	No	520	99.1
Don't know	0	0.0	Don't know	0	0.0
Missing	0	n/a	Missing	6	n/a
Radiation given			Radiation taken		
No	223	23.8	No	209	22.4
Yes	712	76.2	Yes	724	77.6
Don't know	0	0.0	Don't know	0	0.0
Missing	4	n/a	Missing	6	n/a
Surgery given			Received surgery		
No	3	99.9	No	3	99.7
Yes	936	0.1	Yes	933	0.3
Don't know	0	0.0	Don't know	0	0.0
Missing	0	n/a	Missing	3	n/a

TABLE 2: Continued.

| | From medical records | | | From treatment questionnaire | |
	Frequency ($n = 939$)	Percent (%)		Frequency ($n = 939$)	Percent (%)
Mastectomy			Mastectomy		
Yes	289	30.7	Yes	288	30.7
No	650	69.3	No	651	69.3
Lumpectomy			Lumpectomy		
Yes	645	68.7	Yes	695	74.0
No	294	31.3	No	244	26.0
Other treatment			Other treatment		
No	929	99.9	No	215	22.9
Yes	0	0.0	Yes	727	77.1
Don't know	1	0.1	Don't know	0	0.0
Missing	9	n/a	Missing	0	n/a
Breast cancer recurrence			Breast cancer recurrence		
No	889	94.7	No	865	93.4
Yes	50	5.3	Yes	61	6.6
Don't know	0	0.0	Don't know	0	0.0
Missing	0	n/a	Missing	13	n/a
Menopausal status			Menopausal status		
Premenopausal	520	55.4	Premenopausal	518	55.2
Postmenopausal	248	26.4	Postmenopausal	242	25.8
Unknown/missing	171	18.2	Unknown/missing	179	19.0

TABLE 3: Agreement analysis[1] comparing information from self-administered questionnaires with medical record information in 939 women with breast cancer.

	Proportion correct (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Kappa
Hormone therapy	92.6	93.2	95.0	95.7	92.2	0.88
Chemotherapy	93.6	97.2	96.6	92.4	96.6	0.88
Radiation treatment[2]	98.3	99.8	96.4	99.00	99.5	0.97
Surgery	99.0	99.7	—	99.7	—	—
Breast cancer recurrence	94.8	75.5	97.3	60.7	98.6	0.65
Menopausal status[3]	95.6	—	—	—	—	0.89
Menopausal status[4]	83.0					0.71
Hormonal medication						
Tamoxifen	89.9	91.1	80.9	97.3	54.8	0.60
Megestrol	97.9	31.2	100.0	100.0	97.9	0.47
Anastrozole	89.8	17.6	99.1	75.0	89.2	0.25
MA12	95.7	54.5	97.5	48.0	98.9	0.49
Chemotherapy medication						
Cyclophosphamide	55.4	56.6	40.5	91.9	7.3	−0.01
Flurouracil	60.2	55.2	79.8	91.5	31.1	0.22
Methotrexate	77.0	49.3	95.1	86.9	74.2	0.48
Epirubicin	82.4	54.7	97.4	92.0	78.9	0.57
Doxorubicin	83.5	48.3	96.9	85.7	83.1	0.52
Docetaxel	89.4	6.7	99.8	80.0	89.5	0.11
Surgery						
Mastectomy	99.8	96.2	99.5	96.5	98.3	0.95
Lumpectomy	93.4	99.1	97.5	91.9	97.5	0.84

[1] Agreement analysis was conducted excluding unknown and missing responses.
[2] Agreement analysis for radiation therapy was conducted excluding women with breast cancer recurrence (according to medical records).
[3] Calculated excluding unknown and missing information (i.e. treating menopausal status as a 2-level variable of premenopausal and postmenopausal).
[4] Calculated including unknown and missing information (i.e. treating menopausal status as a 3-level variable of premenopausal, postmenopausal, and unknown/missing. Unknown/missing information that match between medical records and treatment questionnaire data are treated as in agreement).

a kappa statistic of 0.71 (95% CI: 0.67, 0.75). No other agreement statistics are applicable as there is no clear gold standard.

3.2. Results of Agreement Analysis by Type of Treatment Received within the Broad Categories. Women were able to report with low-to-moderate agreement the specific type of hormonal therapy or chemotherapy they received. The proportion of women who correctly reported whether or not they received the specific type of hormonal therapy was above 89% for each medication, while specificity was above 97% for each medication. Sensitivity was high for tamoxifen, however, approximately 69%, 82%, and 46% of women who received megestrol, anastrozole and those who were part of the MA12 trial, respectively, received these medications according to their medical records but did not self-report so. The kappa statistics also demonstrated low-to-moderate agreement for all medications.

In regards to chemotherapy, women could report with low-to-moderate agreement the specific type of medication they received. Specificity was high for all medications with the exception of cyclophosphamide (40.5%), while sensitivity was approximately 50% for each medication with the exception of docetaxel (6.7%), indicating that approximately half of the women who received a given medication according to their medical records did not self-report receiving the medication in question. The kappa statistic also demonstrated low-to-moderate agreement for each medication. In general, it appears as though agreement was lower for the medications commonly used in combination therapies (i.e., cyclophosphamide, methotrexate, and fluorouracil), indicating that women may experience greater difficulty identifying a medication they received in combination with other drugs in comparison to medications they received alone. And lastly, women were able to accurately report whether they received a mastectomy or a lumpectomy with excellent agreement on all statistics calculated.

3.3. Results of Regression Analysis. As demonstrated by Tables 4 and 5, generally speaking, with the exception of breast cancer recurrence, no factors were consistently associated with agreement. In regards to the broad categories of treatment, women who experienced breast cancer recurrence during the study period (up to questionnaire completion) were approximately 80% less likely to correctly report whether or not they received hormonal therapy and chemotherapy (OR = 0.19, 95% CI: 0.09, 0.34, and OR = 0.20, 95% CI: 0.09, 0.42 for hormonal therapy and chemotherapy, resp.). In regards to specific types of medications, women who experienced breast cancer recurrence were also approximately 72%–90% less likely to correctly report the specific types of hormonal therapy medications received and were 58% and 92% less likely to correctly report having received the chemotherapy medications doxorubicin and docetaxel, respectively (OR = 0.42; 95% CI: 0.20, 0.92, and OR = 0.08, 95% CI = 0.03, 0.17 for doxorubicin and docetaxel, resp.). Having a family history of breast or ovarian cancer was not associated with accuracy of reporting.

Furthermore, although no other factors were consistently associated with agreement pertaining to accurately reporting broad categories of treatment, or specific types of hormonal therapy medications received, older age at diagnosis, longer periods of recall, being born in a non-English-speaking country (in comparison to a non-English-speaking country), increased alcohol consumption, having a high-school education (in comparison to a bachelor degree), and being a current smoker (in comparison to a nonsmoker) were associated with worse agreement for some specific types of chemotherapy medications. In regards to agreement between menopausal status from a subsequent questionnaire and medical records, agreement was lower among older women, married women, and among current smokers (OR = 0.92, 95% CI: 0.91, 0.94, OR = 0.69, 95% CI: 0.40, 0.87, and OR = 0.63, 95% CI: 0.42, 0.96 for age at diagnosis, being married, and current smoking, resp.).

4. Discussion

This study demonstrates that accuracy of self-reported treatment compared to medical records was high for most broad categories of treatment received (i.e., hormonal therapy, chemotherapy, and radiation). The low specificity and kappa statistic for surgery reporting are likely due to the fact that very few women did not receive surgery for their breast cancer. In regards to menopausal status at time of diagnosis, agreement between medical records and self-report information was also relatively high. Women were only able to report whether they experienced breast cancer recurrence or not with moderate accuracy. We also observed that agreement was moderate for some, but not all, types of treatment received within the broad categories of treatment (i.e., the type of hormonal therapy and chemotherapy received). The results of this study are generally consistent with previous research which also identified that women are able to accurately recall some important details pertaining to their treatment [1–4].

We also found that women had the greatest difficulty reporting the type of chemotherapy received, specifically for medications commonly used in combination therapies (i.e., cyclophosphamide, methotrexate, and fluorouracil). Furthermore, agreement for the specific type of therapy given within the broad categories of treatment appeared to be higher for questions including prompts. For example, in regards to surgery, women were asked to specify whether they received a lumpectomy or a mastectomy, whereas in questions pertaining to the details of hormonal therapy and chemotherapy received, the open-ended instructions of "please list medicines if known" were given. Accordingly, future questionnaires designed to assess self-reported treatment history may want to consider adding prompts to aid recall, specifically when women are being asked to report on treatments that may include combination therapies. Although there are disadvantages associated with adding prompts (i.e., they may deter respondents from considering responses that are not already presented), at the very least further research is warranted to examine which approach

TABLE 4: Odds ratios (ORs) and 95% confidence intervals (CIs) for factors potentially associated with agreement between questionnaire and medical record information for broad categories of treatment, menopausal status and specific types of hormonal therapy in 939 women with breast cancer.

	Hormonal therapy OR[1] (95% CI)	Chemotherapy OR[1] (95% CI)	Radiation therapy[2] OR[1] (95% CI)	Menopuasal status OR[1] (95% CI)	Tamoxifen OR[1] (95% CI)	Megestrol OR[1] (95% CI)	Anastrozole OR[1] (95% CI)
Age at diagnosis[1]	1.02 (0.98, 1.06)	1.02 (0.98, 1.06)	1.04 (0.91, 1.20)	**0.92 (0.91, 0.94)**	1.04 (0.99, 1.09)	1.03 (0.91, 1.16)	1.01 (0.96, 1.05)
Recall period[1]	0.88 (0.73, 1.06)	0.93 (0.77, 1.14)	1.12 (0.71, 1.76)	0.96 (0.85, 1.08)	1.04 (0.82, 1.31)	1.79 (0.82, 3.92)	1.17 (0.91, 1.50)
Marital status (yes versus no)	**1.95 (1.12, 3.40)**	1.37 (0.72, 2.60)	0.47 (0.10, 2.16)	**0.69, (0.40, 0.87)**	1.02 (0.51, 2.06)	0.46 (0.05, 4.21)	1.35 (0.71, 2.58)
English speaking (yes versus no)	1.26 (0.60, 2.64)	1.55 (0.74, 3.25)	1.99 (0.52, 7.61)	0.03 (0.57, 1.51)	0.92 (0.35, 2.42)	1.26 (0.12, 12.98)	1.37 (0.58, 3.22)
Breast Cancer Recurrence (yes versus no)	**0.19 (0.09, 0.34)**	**0.20 (0.09, 0.42)**	—	1.50 (0.71, 3.20)	**0.28 (0.12, 0.65)**	**0.10 (0.02, 0.44)**	**0.24 (0.11, 0.54)**
Family history (yes versus no)	0.88 (0.51, 1.50)	0.86 (0.49, 1.52)	3.64 (0.79, 16.75)	1.11 (0.81, 1.52)	1.13 (0.60, 2.14)	1.45 (0.30, 6.94)	1.40 (0.74, 2.62)
Alcohol consumption[1]	1.01 (0.96, 1.06)	1.00 (0.96, 1.05)	1.04 (0.91, 1.20)	0.99 (0.96, 1.01)	0.98 (0.93, 1.03)	0.97 (0.87, 1.08)	1.01 (0.95, 1.06)
Education level							
Bachelor/graduate	1.00	1.00	1.00	1.00	1.00	1.00	1.00
High school	1.42 (0.75, 2.68)	0.52 (0.24, 1.12)	1.27 (0.30, 5.42)	0.84 (0.56, 1.27)	1.29 (0.60, 2.79)	0.86 (0.12, 5.97)	1.14 (0.52, 2.50)
Some vocational/college	1.64 (0.85, 2.68)	0.57 (0.27, 1.21)	1.07 (0.28, 4.08)	0.74 (0.49, 1.12)	1.14 (0.53, 2.43)	1.30 (0.19, 9.15)	0.77 (0.37, 1.62)
Smoking status							
Never	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Current	0.86 (0.44, 1.68)	1.27 (0.58, 2.76)	0.72 (0.16, 3.20)	**0.63 (0.42, 0.96)**	0.98 (0.46, 2.11)	0.14 (0.01, 1.70)	0.74 (0.34, 1.62)
Former	1.25 (0.67, 2.32)	0.77 (0.42, 1.41)	0.51 (0.16, 1.68)	0.77 (0.54, 1.10)	1.80 (0.84, 3.83)	0.11 (0.01, 1.05)	0.69 (0.35, 1.33)
Menopausal status							
Premenopausal	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Postmenopausal	0.62 (0.27, 1.43)	1.21 (0.50, 2.94)	0.24 (0.05, 1.27)	—	0.35 (0.13, 0.95)	0.74 (0.06, 8.93)	0.54 (0.21, 1.41)
Unknown	0.61 (0.28, 1.32)	1.27 (0.55, 2.92)	1.50 (0.16, 14.00)		0.30 (0.12, 0.70)	1.07 (0.08, 13.73)	0.63 (0.26, 1.52)

[1] Variables were treated as continuous variables.

[2] Odds ratio (OR) adjusted for all other variables included in the table.

[3] Agreement analysis for radiation therapy was conducted excluding women with breast cancer recurrence (according to medical records).

TABLE 5: Odds ratios (ORs) and 95% confidence intervals (CIs) for factors potentially associated with agreement between questionnaire and medical record information for specific types of chemotherapy medication in 939 women with breast cancer.

	Cyclophosphamide OR[1] (95% CI)	Fluorouracil OR[1] (95% CI)	Methotrexate OR[1] (95% CI)	Epirubicin OR[1] (95% CI)	Doxorubicin OR[1] (95% CI)	Docetaxel OR[1] (95% CI)
Age at diagnosis[1]	**0.95 (0.92, 0.97)**	**0.95 (0.93, 0.98)**	**0.96 (0.92, 0.99)**	0.99 (0.96, 1.03)	0.96 (0.92, 1.00)	1.01 (0.96, 1.06)
Recall period[1]	**0.72 (0.62, 0.84)**	**0.80 (0.69, 0.93)**	0.88 (0.75, 1.03)	0.92 (0.77, 1.11)	0.84 (0.71, 1.00)	1.31 (0.97, 1.76)
Marital status (yes versus no)	0.74 (0.46, 1.20)	0.78 (0.48, 1.27)	0.78 (0.45, 1.38)	0.84 (0.45, 1.57)	0.84 (0.45, 1.56)	1.64 (0.79, 3.42)
English speaking (yes versus no)	1.98 (1.05, 3.73)	1.18 (0.97, 2.34)	1.47 (0.76, 2.85)	**2.27 (1.16, 4.43)**	0.76 (0.33, 1.75)	1.59 (0.59, 4.31)
Breast cancer recurrence (yes versus no)	0.81 (0.40, 1.64)	0.59 (0.29, 1.19)	0.88 (0.40, 1.93)	0.63 (0.29, 1.37)	**0.42 (0.20, 0.92)**	**0.08 (0.03, 0.17)**
Family history (yes versus no)	0.95 (0.63, 1.44)	1.09 (0.72, 1.65)	1.02 (0.64, 1.64)	1.10 (0.65, 1.88)	0.88 (0.52, 1.48)	1.42 (0.68, 2.98)
Alcohol consumption[1]	1.01 (0.97, 1.05)	1.02 (0.98, 1.06)	1.05 (1.00, 1.11)	1.03 (0.97, 1.08)	1.00 (0.96, 1.05)	**0.95 (0.90, 0.99)**
Education level						
Bachelor/graduate	1.00	1.00	1.00	1.00	1.00	1.00
High school	**0.56 (0.33, 0.95)**	**0.52 (0.30, 0.87)**	0.87 (0.47, 1.58)	0.56 (0.28, 1.12)	0.75 (0.39, 1.48)	0.91 (0.36, 2.27)
Some vocational/college	0.68 (0.42, 1.11)	0.67 (0.41, 1.10)	1.02 (0.58, 1.82)	0.56 (0.29, 1.05)	0.99 (0.51, 1.90)	0.56 (0.25, 1.27)
Smoking status						
Never	1.00	1.00	1.00	1.00	1.00	1.00
Current	0.66 (0.39, 1.11)	0.83 (0.50, 1.39)	**0.54 (0.30, 0.96)**	1.50 (0.75, 3.02)	**0.35 (0.19, 0.65)**	0.64 (0.27, 1.48)
Former	1.02 (0.65, 1.60)	1.19 (0.76, 1.87)	0.95 (0.56, 1.62)	1.19 (0.68, 2.09)	0.76 (0.42, 1.38)	0.75 (0.36, 1.53)
Menopausal status						
Premenopausal	1.00	1.00	1.00	1.00	1.00	1.00
Postmenopausal	0.73 (0.36, 1.49)	0.96 (0.47, 1.94)	0.84 (0.38, 1.84)	0.86 (0.35, 2.14)	1.05 (0.43, 2.53)	0.62 (0.19, 2.03)
Unknown	1.60 (0.88, 2.91)	1.63 (0.89, 2.97)	1.48 (0.73, 2.99)	1.09 (0.51, 2.33)	1.64 (0.75, 3.59)	1.54 (0.53, 4.71)

[1] Variables were treated as continuous variables.
[2] Odds ratio (OR) adjusted for all other variables included in the table.

encourages more accurate reporting of specific therapies received.

In regards to the factors associated with agreement, women with breast cancer recurrence were less likely to report having received hormonal therapy, chemotherapy, all types of specific hormonal therapy medications, and the chemotherapy medications doxorubicin and docetaxel. This finding is consistent with a previous study conducted by Phillips et al. (2005), who found women with breast cancer recurrence at the time of questionnaire administration were less likely to report correctly on most questions (aside from chemotherapy treatment) [4]. Furthermore, a study conducted by Liu et al. (2010) also found that breast cancer recurrence was significantly associated with lower agreement when reporting having received chemotherapy [1]. This finding is likely attributed to women reporting treatment for the breast cancer recurrence instead of the original diagnosis.

No other factors were consistently associated with better agreement; however, older age at diagnosis, longer recall period, being born in a non-English-speaking country (in comparison to an English-speaking country), increased alcohol consumption, having a high-school education (in comparison to a bachelor degree), and being a current smoker (in comparison to a nonsmoker) were all associated with worse accuracy when reporting some specific types of chemotherapy medication. In regards to education level, alcohol consumption, and smoking status, it is possible these factors may indicate lower socioeconomic status.

Strengths of this study include a large sample size of women recruited from a sample of women initially identified from a population-based cancer registry. Furthermore, although four studies have previously examined the accuracy of self-reported breast cancer treatment, only three of these studies used medical records as the gold standard when assessing agreement, and only two of these studies examined potential factors associated with agreement. To our knowledge, this study is also the first study to examine agreement between menopausal status recorded in medical records at the time of diagnosis and retrospective assessment of menopausal status from a structured questionnaire.

There are several limitations to this study. Medical records may not always be an appropriate gold standard. For example, a prescription may have been given but was not retrieved in the current collection of medical records or the medication may not have actually been taken, although this is less likely with treatments given in a clinic setting such as radiation or most chemotherapy. However, agreement is still high despite this possibility. Furthermore, a Swedish prospective cohort study also confirmed the validity of self-reporting current hormonal therapy using a 7-day personal diary as the gold standard [9]. Accordingly, taking into consideration that self-reported treatment demonstrates high consistency with both medical records and a personal diary and the limitations of medical record abstraction described above, this study suggests that self-reported treatment is a reasonable option for gathering treatment data in broad categories. Additionally, the women in the study are largely Caucasian (96.5%), which limits the generalizability of the results beyond this group. Also, the study sample primarily included women with early-stage breast cancer; women with late-stage breast cancer, who are more likely to be severely ill, may have a more difficult time recalling the details of their treatment. Furthermore, women recruited were oversampled for family history, and it is possible that having a family history of breast cancer may make women more knowledgeable regarding treatment options, which in turn may influence agreement. However, we did not find that family history was associated with measures of agreement. More research is warranted to examine agreement between self-reported treatment and medical records among average-risk women, and particularly among a more ethnically diverse population. Given that we found that most characteristics, including age, education, and family history, were not associated with reporting accuracy, it is likely that the levels of agreement we observed would be similar in other nonimmigrant breast cancer patient populations. The self-reported treatment questionnaire also did not collect information on dates treatments were received, or on adherence, accordingly the accuracy of this self-reported information could not be examined, which may be an important consideration for some studies (e.g., those examining the association between treatment delay and survival). Finally, since we conducted many statistical tests, individual significant results could have occurred by chance. Consequently, the focus was placed on overall consistent patterns rather than individual results.

In conclusion, the results of this study demonstrate that women can accurately report important details regarding their breast cancer treatment, even, in some cases, for more specific questions pertaining to the details of treatment within the broader categories of hormonal therapy, chemotherapy, and surgery. Accordingly, self-report appears to be an inexpensive method of data collection pertaining to broad categories of cancer treatment, if not always specific drugs, and a feasible alternative to medical record abstraction in large epidemiological studies of breast cancer patients where intensive medical record abstraction may not be possible.

Conflict of Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] Y. Liu, A. L. Diamant, A. Thind, and R. C. Maly, "Validity of self-reports of breast cancer treatment in low-income, medically underserved women with breast cancer," *Breast Cancer Research and Treatment*, vol. 119, no. 3, pp. 745–751, 2010.

[2] E. Maunsell, M. Drolet, N. Ouhoummane, and J. Robert, "Breast cancer survivors accurately reported key treatment and prognostic characteristics," *Journal of Clinical Epidemiology*, vol. 58, no. 4, pp. 364–369, 2005.

[3] M. Schootman, D. B. Jeffe, M. M. West, and R. Aft, "Self-report by elderly breast cancer patients was an acceptable alternative to surveillance, epidemiology, and end results (SEER) abstract data," *Journal of Clinical Epidemiology*, vol. 58, no. 12, pp. 1316–1319, 2005.

[4] K. A. Phillips, R. L. Milne, S. Buys et al., "Agreement between self-reported breast cancer treatment and medical records in a population-based Breast Cancer Family Registry," *Journal of Clinical Oncology*, vol. 23, no. 21, pp. 4679–4686, 2005.

[5] E. M. John, J. L. Hopper, J. C. Beck et al., "The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer," *Breast Cancer Research*, vol. 6, no. 4, pp. R375–R389, 2004.

[6] J. C. Figueiredo, M. Ennis, J. A. Knight et al., "Influence of young age at diagnosis and family history of breast or ovarian cancer on breast cancer outcomes in a population-based cohort study," *Breast Cancer Research and Treatment*, vol. 105, no. 1, pp. 69–80, 2007.

[7] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.

[8] S. R. Lipsitz, M. Parzen, G. M. Fitzmaurice, and N. Klar, "A two-stage logistic regression model for analyzing inter-rater agreement," *Psychometrika*, vol. 68, no. 2, pp. 289–298, 2003.

[9] J. Merlo, G. Berglund, E. Wirfält et al., "Self-administered questionnaire compared with a personal diary for assessment of current use of hormone therapy: an analysis of 16,060 women," *American Journal of Epidemiology*, vol. 152, no. 8, pp. 788–792, 2000.