Agreement between self-reported and register-based cardiovascular events among Danish breast cancer survivors

Rikke Langballe[1], Esther M John[2,3], Kathleen E Malone[4], Leslie Bernstein[5], Julia A Knight[6,7], Charles F Lynch[8], Rebecca M Howell[9], Roy Shore[10], Meghan Woods[11], Patrick Concannon[12], the WECARE Study Collaborative Group[13], Jonine L Bernstein[11], and Lene Mellemkjær[1]

[1]Danish Cancer Society Research Center, Copenhagen, Denmark

[2]Cancer Prevention Institute of California, Fremont, CA, USA

[3]Department of Health Research and Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

[4]Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[5]Beckman Research Institute of the City of Hope, Duarte, CA, USA

[6]Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Canada

[7]University of Toronto, Dalla Lana School of Public Health, Toronto, Canada

[8]University of Iowa, Iowa City, IA, USA

[9]Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[10]New York University School of Medicine, NY, USA

[11]Memorial Sloan Kettering Cancer Center, New York, NY, USA

[12]Genetics Institute and Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL, USA

Abstract

PURPOSE—We examined the degree of over- and under-reporting of cardiovascular diseases (CVDs) among female breast cancer survivors comparing self-reports to diagnostic codes from the Danish National Patient Register (NPR).

METHODS—The study comprised 357 Danish breast cancer patients from the WECARE Study who completed a telephone interview concerning CVDs. Disease diagnoses for these women were obtained from the NPR. Agreement was calculated as the number of diagnoses that were both self-reported and in the NPR divided by 1) number of self-reported diagnoses (over-reporting) or 2) number of diagnoses in the NPR (under-reporting).

RESULTS—In total, 68 women reported 96 specific cardiovascular outcomes of which 56 (58%) were found in the NPR. Ninety cardiovascular diagnoses were found in the NPR of which 56 (62%) were specifically reported at the interview. There was 80% agreement as to the occurrence of a cardiovascular diagnosis overall. Of 289 women reporting no CVD, 273 (94%) had no diagnoses in the NPR.

CONCLUSIONS—Breast cancer survivors seem to report absence of CVD accurately, but they both over-report and under-report specific cardiovascular diagnoses. Using a broader definition of CVDs improves the agreement between self-reported and NPR data.

IMPLICATIONS FOR CANCER SURVIVORS—Determining how cancer treatments affect the risk of cardiovascular morbidities is essential and the development of high quality methods for collecting such data is critical. While self-reported data are adequate for assessing the presence of any CVD condition, medical record review will yield higher quality data on specific CVD conditions.

INTRODUCTION

Disease outcomes are often assessed by self-report in epidemiological studies. An assessment of the quality of such self-reported data is important for conclusions obtained from the findings. Several factors are likely to influence the accuracy and validity of the self-reported diagnoses such as characteristics of the population interviewed (sex, age and educational level), communication between patient and physician, data collection methods (e.g., interview or self-administered questionnaire and wording of the questions) and type of diagnosis (severity, chronicity and presence of distinct disease criteria).

Alternative sources for identifying disease outcomes for such studies are nationwide or regional administrative databases with diagnostic information. Except for type of diagnosis, these data are independent of the factors influencing self-reports mentioned above; however, accuracy and validity are likely to depend on factors such as coding practices, inclusion rules and completeness of ascertainment.

In previous validation studies self-reported diagnoses of cardiovascular disease (CVD) have been compared either to medical records alone [1–7], administrative database diagnoses alone [8–12] or both medical records and administrative data [13, 14]. The overall conclusion drawn from these studies is that study participants tend to over-report CVDs [1–14], whereas the accuracy of reporting the absence of these conditions appears to be high [8–13].

No validation studies comparable to those described above have been carried out for breast cancer patients. Growing research in cancer survivorship and late-effects of cancer therapies have elevated interest in the collection of information on non-cancer conditions in cancer

survivors. It is essential to improve methods for collecting such data. In the current report, we compare CVDs reported by Danish breast cancer patients in the Women's Environmental Cancer and Radiation Epidemiology (WECARE) Study with their diagnoses recorded in the Danish National Patient Register (NPR).

METHODS

Study population

The WECARE Study includes as cases, women diagnosed with contralateral breast cancer and as controls, women diagnosed with unilateral breast cancer, who are from the United States, Canada or Denmark and who had their first breast cancer diagnosis before age 55 years during the years 1985–2008. Cases had their contralateral breast cancer diagnosis during 1986–2011 [15]. The participants were recruited in two phases during 2000–2004 and 2009–2012. In 2014, information on CVDs was collected from WECARE participants to study late-effects following breast cancer treatment.

Women included in the present study were WECARE participants from Denmark. Of the 735 Danish WECARE participants, 562 (76%) were alive in 2014 and were re-contacted concerning participation in the CVD study. Subsequently, 357 (64%) of the contacted women agreed to participate in a telephone interview concerning CVDs and related conditions. The women were asked to report CVDs by responding to the following question "Have you ever been told by a doctor that you had ….?". If they responded "Yes", the next question was "How old were you when this first occurred?" For angina pectoris and irregular heartbeat, it was required that the condition had been treated (Online Resource, Figure 1). Show cards listing the CVDs were mailed to the participants to assist them during the telephone interview. The interviewer also asked the participants to report the type of health care facility where they were diagnosed with CVD.

Danish National Patient Register (NPR)

We used the Danish NPR, a nationwide health Register that includes diagnoses from in-patient hospitalizations since 1977 and outpatient visits since 1995, as a source of validation data [16]. Diagnoses in the NPR have been coded according to a Danish version of the International Classification of Diseases version 8 (ICD-8) during 1977–1993 and version 10 (ICD-10) since 1994. We used the personal identification number to link each study participant to the NPR to obtain CVD diagnoses.

We restricted the validation to CVDs that are likely referred to a hospital department: cardiomyopathy or congestive heart failure, myocardial infarction, coronary heart disease, angina pectoris, irregular heartbeat, pericarditis or pericardial constriction, valvular diseases and blood clots. For each of these conditions, we identified the corresponding ICD-8 and ICD-10 codes (Online Resource, Table 1).

Assessment of accuracy of self-reported CVD conditions

Our objective was to evaluate the extent of agreement by assessing both over-reporting and under-reporting of CVD conditions. First, for each CVD reported by the women, we

searched for the corresponding ICD codes in the NPR. We estimated the degree of over-reporting as the proportion of diagnoses reported in the interview that were also found in the NPR (the number of diagnoses both at interview and in the NPR divided by the number of diagnoses at interview × 100) [1]. We also calculated the following percentages of self-reported diagnoses for which the participant had: 1) another cardiovascular diagnosis in the NPR, 2) the participant had consulted a specialist clinic in relation to the CVD, 3) self-reported diagnosis before 1995, and 4) none of these applied. For women having self-reported diagnoses in the latter category, we searched for any unspecified cardiac and other non-cardiovascular diagnoses occurring in the NPR around the same time as the self-reported event.

Next, we used the NPR as the primary source of CVD diagnosis information and subsequently looked in the self-reported CVD diagnosis data to assess the degree of under-reporting, estimated as the proportion of diagnoses in the NPR that were also reported in the interview (the number of diagnoses both at interview and in the NPR divided by the number of diagnoses in the NPR × 100) [1]. We also calculated the percentages of diagnoses in the NPR for which the participant had reported 1) another CVD, or 2) no CVD. To assess the validity of absence of self-reporting, the number of women in the latter category was also divided by the total number of women *not* reporting any CVD during the interview × 100.

RESULTS

The mean age at interview of the 357 women included in the study was 67 years (range 44–81). Thirty-three percent of women were 70 years or older at the time of the interview (Table 1). Of the 357 women, 68 (19%) reported at least one CVD diagnosis. Forty-four (65%) of these 68 women reported only one CVD diagnosis, 21 (31%) reported two such diagnoses, while three (4%) women reported three or four diagnoses.

The 68 women reported 96 CVD diagnoses, and the same diagnosis was found in the NPR for 56 (58%) reported diagnoses (Table 2). Another 17 (18%) diagnoses were within the investigated spectrum of CVDs. There were 11 (11%) self-reported diagnoses not identified in the NPR for which the woman had reported consulting a specialist. These diagnoses are not captured by the NPR. Two women reported having been diagnosed before 1995 when outpatient visits were not included in the NPR. For the remaining 10 (10%) diagnoses reported by 9 women, we found no obvious explanation for why the diagnosis was not in the NPR (Table 2). Of these, three women had been hospitalized for suspicion of unspecified cardiac disease and two for an illness not further specified. The CVDs were reported to be diagnosed before the breast cancer diagnosis in six women (five with irregular heart beat and one with valvular disease). Omitting these women from Table 2 did not change the pattern described above. The distributions for total number of self-reported diagnoses did not differ considerably by time between CVD diagnosis and interview (Table 2).

Among the 357 women, a total of 90 cardiovascular diagnoses were found in the NPR (Table 3). Of these diagnoses, 56 (62%) were reported at the interview, while 17 (19%) were reported as another cardiovascular condition during the interview, and 17 (19%) were not reported at all. Notably, of six women who had a diagnosis of cardiomyopathy or congestive

heart failure in the NPR, only one reported such a diagnosis, whereas five reported another CVD. Of CVDs in the NPR more than 10 years before the interview, 37% were not reported by the women, while the corresponding percentages for CVDs 5–9 years and 1–4 years before the interview were 15% and 13%, respectively. Sixteen (6%) of the 289 participants who reported no cardiovascular events had at least one such diagnosis in the NPR. Thus, 94% reported accurately no history of cardiovascular diagnoses. These 16 women had 17 (19%) CVDs in the NPR (Table 3).

DISCUSSION

Among breast cancer survivors, the overall agreement between self-reported and NPR-based CVDs was about 60%, while the agreement increased to about 80%, if CVDs were considered as a combined category. Furthermore, we found high accuracy (94%) for reporting the absence of CVD.

Previous studies using electronic diagnostic codes as the validation source [8–12] or both diagnostic codes and medical records [13] have, in general, found that study participants tend to over-report. Most studies using medical records as the validation source show a similar picture [4–7], though not all [2, 3]. This tendency towards over-reporting was seen despite differences in characteristics of the study populations, data collection methods (mailed questionnaire versus phone interview and wording of questions), or data sources used for validation.

Previous studies comparing self-reports of CVDs to either hospital registry data [9], medical records data [2, 4, 6, 7] or both [13, 14] have found that self-reported CVDs not verified by the validation source are often associated with a positive history of other related CVDs. Likewise, we observed that when the specific self-reported diagnosis could not be found in the NPR, other CVDs were often found in the NPR. In addition, among the participants for whom we found no explanation for the absence of a CVD in the NPR, approximately half had been hospitalized for observation of an unspecified cardiac disease or other condition suggesting that sometimes a definite cardiac diagnosis is difficult for clinicians to reach. The reason for the lack of agreement between self-reported and NPR-based cardiovascular diagnoses may be a combination of unclear or unspecific diagnostic criteria and closely related symptoms of CVDs. Barr et al. [13] suggested that combining closely related cardiac diseases such as myocardial infarction and angina pectoris may increase validity of self-report. Based on our findings, we suggest that the combined category should be even broader – perhaps just capturing whether the study participants ever had a cardiovascular medical problem. If specific CVDs are of particular interest, self-reports would necessitate medical records review to increase accuracy.

In line with our findings, previous studies have consistently shown high accuracy of recall for participants who did not report a CVD [3, 8–13]. Thus, self-reports are quite useful in determining study eligibility by excluding study participants with prevalent disease. Most previous studies [9, 11–13], although not all [10], have reported a high sensitivity, i.e., a low degree of unreported events. A Canadian study showed a sensitivity of 0.77 for self-reported acute myocardial infarction, whereas the sensitivity for congestive heart failure was only

0.26 (8). In our study, only one case of cardiomyopathy or congestive heart failure was self-reported out of six cases found in the NPR. The authors of the Canadian study argued that the under-reporting of heart failure could be caused by nonspecific symptoms related to the disease, whereas administrative data may be more likely to identify chronic diseases requiring ongoing contact with the healthcare system. When using the NPR as the primary data source for validation, we observed that if the specific diagnosis identified in the NPR was unreported by the participant, she often reported another CVD within the spectrum investigated. Our study further showed that under-reporting decreased by recency of the diagnosis. Thus, combining diagnoses and restricting to recent diagnoses could also reduce under-reporting.

Strengths of our study include the wide spectrum of CVD diagnoses for assessment of over- and under-reporting. Also, we had the ability to search for diagnoses outside the spectrum of CVDs to identify nonspecific diagnoses. Limitations of our study include the small number of self-reported CVDs which prevented us from calculating Kappa statistics [8]. Moreover, we had no true gold standard such as medical records that could have been used to determine whether self-reported or NPR-based information was more accurate. Finally, the study participants represent a selected group of breast cancer survivors, and this may limit the generalization of findings.

CONCLUSIONS

Self-reports of the absence of cardiovascular conditions seem to be accurate among female breast cancer survivors. Complex diagnostic criteria and closely related symptoms for CVDs make it difficult for patients to report specific diagnoses accurately. Therefore, our findings suggest that future epidemiological studies using questionnaire-based information should combine specific CVDs to capture whether the study participants ever had a cardiovascular condition. Studies with interest in specific diagnoses should verify self-reports using data from medical records.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Collaborative Site Staff: Kristina Blackmore M.Sc.; Anh T. Diep; Judy Goldstein B.A.; Irene Harris B.S., C.M.D.; Rikke Langballe M.P.H.; Cecilia O'Brien; Susan Smith M.P.H.; Rita Weathers M.S.; Michele West Ph.D.

References

1. Harlow SD, Linet MS. Agreement between questionnaire data and medical records. The evidence for accuracy of recall. Am J Epidemiol. 1989; 129(2):233–48. [PubMed: 2643301]

2. Tretli S, Lund-Larsen PG, Foss OP. Reliability of questionnaire information on cardiovascular disease and diabetes: cardiovascular disease study in Finnmark county. J Epidemiol Community Health. 1982; 36(4):269–73. [PubMed: 7166682]

3. Haapanen N, Miilunpalo S, Pasanen M, Oja P, Vuori I. Agreement between questionnaire data and medical records of chronic diseases in middle-aged and elderly Finnish men and women. Am J Epidemiol. 1997; 145(8):762–9. [PubMed: 9126003]

4. Colditz GA, Martin P, Stampfer MJ, Willett WC, Sampson L, Rosner B, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study of women. Am J Epidemiol. 1986; 123(5):894–900. [PubMed: 3962971]

5. Klungel OH, de Boer A, Paes AH, Seidell JC, Bakker A. Cardiovascular diseases and risk factors in a population-based study in The Netherlands: agreement between questionnaire information and medical records. Neth J Med. 1999; 55(4):177–83. [PubMed: 10555434]

6. Paganini-Hill A, Chao A. Accuracy of recall of hip fracture, heart attack, and cancer: a comparison of postal survey data and medical records. Am J Epidemiol. 1993; 138(2):101–6. [PubMed: 8342528]

7. Bergmann MM, Byers T, Freedman DS, Mokdad A. Validity of self-reported diagnoses leading to hospitalization: a comparison of self-reports with hospital records in a prospective study of American adults. Am J Epidemiol. 1998; 147(10):969–77. [PubMed: 9596475]

8. Muggah E, Graves E, Bennett C, Manuel DG. Ascertainment of chronic diseases using population health data: a comparison of health administrative data and patient self-report. BMC Public Health. 2013; 13:16.doi: 10.1186/1471-2458-13-16 [PubMed: 23302258]

9. Meisinger C, Schuler A, Lowel H, Group MK. Postal questionnaires identified hospitalizations for self-reported acute myocardial infarction. J Clin Epidemiol. 2004; 57(9):989–92. [PubMed: 15551512]

10. St Sauver JL, Hagen PT, Cha SS, Bagniewski SM, Mandrekar JN, Curoe AM, et al. Agreement between patient reports of cardiovascular disease and patient medical records. Mayo Clin Proc. 2005; 80(2):203–10. [PubMed: 15704775]

11. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. J Clin Epidemiol. 2004; 57(10):1096–103. DOI: 10.1016/j.jclinepi.2004.04.005 [PubMed: 15528061]

12. Navin Cristina TJ, Stewart Williams JA, Parkinson L, Sibbritt DW, Byles JE. Identification of diabetes, heart disease, hypertension and stroke in mid- and older-aged women: Comparing self-report and administrative hospital data records. Geriatr Gerontol Int. 2016; 16(1):95–102. DOI: 10.1111/ggi.12442 [PubMed: 25613080]

13. Barr EL, Tonkin AM, Welborn TA, Shaw JE. Validity of self-reported cardiovascular disease events in comparison to medical record adjudication and a statewide hospital morbidity database: the AusDiab study. Intern Med J. 2009; 39(1):49–53. DOI: 10.1111/j.1445-5994.2008.01864.x [PubMed: 19290982]

14. Bolland MJ, Barber A, Doughty RN, Grey A, Gamble G, Reid IR. Differences between self-reported and verified adverse cardiovascular events in a randomised clinical trial. BMJ Open. 2013; 3(3)doi: 10.1136/bmjopen-2012-002334

15. Langballe R, Mellemkjaer L, Malone KE, Lynch CF, John EM, Knight JA, et al. Systemic therapy for breast cancer and risk of subsequent contralateral breast cancer in the WECARE Study. Breast Cancer Res. 2016; 18(1):65.doi: 10.1186/s13058-016-0726-0 [PubMed: 27400983]

16. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health. 2011; 39(7 Suppl):30–3. DOI: 10.1177/1403494811401482 [PubMed: 21775347]

Table 1

Characteristics of the 357 Danish Breast Cancer Patients from the WECARE Study Included in the Validation Study of Cardiovascular Diseases

Characteristics	N	%
Age at interview (years)		
40–49	3	1
50–59	48	14
60–69	187	52
70	119	33
Age at first breast cancer diagnosis (years)		
<40	59	16
40–44	113	32
45–49	121	34
50–54	64	18
Calendar-year of first breast cancer diagnosis		
1985–89	110	31
1990–94	139	39
1995–99	98	27
2000–04	10	3
Years of education		
12	127	36
13–16	154	43
17	76	21
Case-control status in WECARE Study		
Case (contralateral breast cancer)	128	36
Control (unilateral breast cancer)	229	64
Number of cardiovascular self-reported events[a]		
None	289	81
At least one	68	19
1	44	
2	21	
3	2	
4	1	

[a]Defined as cardiomyopathy or congestive heart failure, myocardial infarction, coronary heart disease, angina pectoris, irregular heartbeat, pericarditis or pericardial constriction, valvular diseases and blood clots.

Table 2

Self-reported Cardiovascular Disease Diagnoses Compared to Diagnoses in the Danish National Patient Register Among the 357 Danish Breast Cancer Patients from the WECARE Study Included in the Validation Study

Cardiovascular diagnoses	Self-reports	National Patient Register		No record of cardiovascular diagnoses in the National Patient Register		
		Same diagnosis	Other cardiovascular diagnosis	Consulted a specialist	Self-reported diagnosis before 1995	No explanation
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Cardiomyopathy or congestive heart failure	2	1 (50)	0	0	0	1
Myocardial infarction	3	2 (67)	1	0	0	0
Coronary heart disease	7	4 (57)	1	2	0	0
Angina pectoris	14	7 (50)	5	0	0	2
Irregular heartbeat	29	20[a] (69)	2	4[b]	0	3
Pericarditis or pericardial constriction	3	2 (67)	1	0	0	0
Valvular diseases	17	8 (47)	2	4	1[c]	2
Blood clots	21	12 (57)	5	1	1	2
Total number of diagnoses N (%)	96 (100)	56 (58)	17 (18)	11 (11)	2 (2)	10 (10)
Total number of patients N (%)	68 (100)	46 (68)*	14 (21)*	9 (13)*	2 (3)*	9 (13)*
Time between self-reported cardiovascular disease and interview, years						
1–4	43 (100)	26 (60)	6 (14)	6 (14)	0	5 (12)
5–9	31 (100)	18 (58)	7 (22)	3 (10)	0	3 (10)
10	22 (100)	12 (55)	4 (18)	2 (9)	2 (9)	2 (9)

*
Numbers do not add up to total number of patients due to overlap.

[a]
Four of these cardiovascular diseases were reported to be diagnosed before the breast cancer diagnosis.

[b]
One of these cardiovascular diseases was reported to be diagnosed before the breast cancer diagnosis.

[c]
This cardiovascular disease was reported to be diagnosed before the breast cancer diagnosis.

Table 3

Cardiovascular Disease Diagnoses in the Danish National Patient Register Compared to the Self-reported Cardiovascular Diagnoses among the 357 Danish Breast Cancer Patients from the WECARE Study Included in the Validation Study

Cardiovascular diagnoses	National Patient Register	Self-report		No self-report
	N (%)	Same diagnosis N (%)	Other cardiovascular diagnosis N (%)	N (%)
Cardiomyopathy or congestive heart failure	6	1 (17)	5	0
Myocardial infarction	2	2 (100)	0	0
Coronary heart disease	8	4 (50)	3	1
Angina pectoris	15	7 (47)	3	5
Irregular heartbeat	29	20 (69)	4	5
Pericarditis or pericardial constriction	2	2 (100)	0	0
Valvular diseases	11	8 (73)	2	1
Blood clots	17	12 (71)	0	5
Total number of diagnoses N (%)	**90 (100)**	**56 (62)**	**17 (19)**	**17 (19)**
Total number of patients N (%)	**65 (100)**	**46 (71)** [*]	**15 (23)** [*]	**16 (25)** [*]
Time between cardiovascular disease in the National Patient Register and interview, years				
1–4	37 (100)	24 (65)	8 (22)	5 (13)
5–9	34 (100)	23 (67)	6 (18)	5 (15)
10	19 (100)	9 (47)	3 (16)	7 (37)

[*] Numbers do not add up to total number of patients due to overlap.