pISSN 1598-2998, eISSN 2005-9256

https://doi.org/10.4143/crt.2024.968 Cancer Res Treat. 2026;58(2):481-491

Original Article

Secondary Cancer Risk in Breast Cancer with and without Radiotherapy: The Observational Health Data Sciences and Informatics (OHDSI) Cohort Study

Seok Kim[1,2], **Dachung Boo**[3], **Sooyoung Yoo**[1], **Borham Kim**[1], **Kyubo Kim**[4], **Kwangsoo Kim**[5,6,7], **Eunhye Song**[8], **Junmo Kim**[9], **Hyun Gee Ryoo**[10], **Jin Chul Paeng**[10], **In Young Choi**[11], **SooJeong Ko**[11], **Ie Ryung Yoo**[12], **Rae Woong Park**[2], **Ho-Young Lee**[1,13,14]

A list of author's affiliations appears at the end of the paper.

Purpose Radiotherapy is used to reduce the risk of breast cancer recurrence after surgery, but it is a potential cause of secondary cancer. We validated the risk of secondary cancer in primary breast cancer who received radiotherapy compared with those who did not from a matched cohort using a large-scale observational study of the Observational Health Data Sciences and Informatics (OHDSI) data network.

Materials and Methods A retrospective comparative cohort study using propensity score-matched cohorts was performed using two Observational Medical Outcome Partnership common data model databases, from tertiary general hospitals in South Korea. Among female patients who underwent surgery after the diagnosis of breast cancer, the risk of secondary primary malignant occurrence after 1:1 matching was analyzed.

Results Among 27,078 patients with breast cancer, there was no significant difference in the risk of secondary cancer following radiotherapy in 4,426 patients after 1:1 propensity-score matching. Further, there were no significant differences in the sensitivity analysis according to age, latency period, and number of radiation treatments.

Conclusion There was no difference in the risk of secondary cancer in the patients diagnosed with breast cancer depending on whether or not radiotherapy was performed after surgery. In the future, it is necessary to analyze including data generated during cancer treatment.

Key words Breast neoplasms, Secondary primary malignant, Radiotherapy, Propensity-score matching, Common data model

Introduction

Radiotherapy is used to treat almost every stage of breast cancer. This is an effective way to reduce the risk of breast cancer recurrence after surgery and alleviate the symptoms of metastatic breast cancer. In the United States, 49% of women with early-stage breast cancer underwent breast-conserving surgery with adjuvant radiation therapy. In addition, 45% of patients with stage III breast cancer underwent mastectomy along with adjuvant chemotherapy and radiation therapy. In comparison, 56% of women diagnosed with stage IV breast cancer often receive radiation and/or chemotherapy alone [1]. In South Korea, radiation therapy is also recommended for women with early-stage breast cancer after breast-conserving surgery as well as for patients with recurrent breast cancer or systemic metastasis [2].

However, radiotherapy has been recognized as a potential cause of secondary cancer, especially lung cancer and sarcoma, and this risk depends on the area treated, the radiation dose biological effectiveness, bone marrow exposure, and personal radiosensitivity. A systematic review of the effects of radiotherapy on the risk of secondary cancer in women with breast cancer reported that radiotherapy was significantly associated with an increased risk of secondary non-breast cancer, especially lung cancer, esophageal cancer, and sarcoma [3]. The risk of secondary cancers was most evident in patients exposed to radiation at a younger age and with an increasing length of follow-up [4]. In a study on French patients with non-metastatic breast cancer, patients who received breast radiotherapy experienced more sarcomas (relative risk [RR], 5.59; confidence interval [CI], 1.35 to 23.17) than those who did not [5].

However, these studies have limitations. First, there were no strategies to reconcile the differences between people with breast cancer who received radiotherapy and those who did not receive radiotherapy or treatment other than

Correspondence: Ho-Young Lee
Department of Nuclear Medicine, Office of eHealth Research and Business, Seoul National University Bundang Hospital, 82 Gumi-ro, 173 beon-gil, Bundang-gu, Seongnam 13620, Korea
Tel: 82-31-787-2938 E-mail: debobkr@snubh.org

Received October 8, 2024 **Accepted** June 4, 2025
Published Online June 5, 2025

Co-correspondence: Rae Woong Park
Department of Biomedical Informatics, Ajou University School of Medicine, 164 Worldcup-ro, Yeongtong-gu, Suwon 16499, Korea
Tel: 82-31-219-5342 E-mail: veritas@ajou.ac.kr

*Seok Kim and Dachung Boo contributed equally to this work.

radiotherapy, such as matching or weighting by propensity scores (PSs). Using PS matching, it was possible to determine whether the treatment and control groups were fully balanced in terms of all identified potential confounders [6,7]. As the effects of confounders were not adjusted in previous studies, care must be taken when interpreting the effects of radiotherapy on second cancers. Second, they did not test a falsification hypothesis using negative controls, which researchers assume cannot plausibly be related to the intervention or treatment of interest [8]. Researchers can reduce residual bias from unmeasured confounding by demonstrating that there is no relationship between the negative control outcomes and the intervention of interest [9] and can insist on an association between specific outcomes and treatment of interest in the study. Third, the sample size was limited, as a single-center database was used to access clinical records and analyze the risk of secondary cancer following radiotherapy [10]. Nationwide databases such as a national cancer registry could provide a large number of study populations but they tend to lack detailed clinical information [4,11,12].

The Observational Medical Outcome Partnership (OMOP) common data model (CDM) converts disparate healthcare data into a standardized format to facilitate large-scale data analysis and reproducible studies [13]. Observational Health Data Sciences and Informatics (OHDSI) provides open-source tools to assist in the design and execution of analyses on standardized observational data in the OMOP CDM-based distributed research network.

The objective of this study was to estimate the risk of secondary cancers in patients with surgically resected breast cancer treated with radiotherapy compared to those who were treated without radiotherapy using OMOP CDM data from two hospitals in South Korea. To improve the validity and reliability of the study, we applied propensity score matching and conducted negative control outcome analysis using falsification testing.

Materials and Methods

This study was a retrospective comparative cohort design study across the OHDSI network study using a population-level-estimation package (OHDSI open-source tool). The study protocol and analysis package for research execution are publicly available at: https://github.com/HIRC-SNUBH/HIRC_Breast2ndCancer.

1. Data sources

We included two databases, which were converted to OMOP CDM ver. 5.3, from tertiary general university hospitals in South Korea: (1) Seoul National University Bundang Hospital (SNUBH) (approximately 1.9 million people, between April 2003 and December 2020) and (2) Seoul National University Hospital (SNUH) (approximately 3.29 million people, between October 2004 and December 2020). The CDM database included electronic health records from the outpatient, inpatient, and emergency departments, and all data were de-identified and maintained at each site. SNUBH developed and distributed the research design and analysis code, and SNUH executed the code and provided analysis results.

2. Target and comparator cohort

This study only included female breast cancer. We defined patients who underwent radiotherapy after breast cancer surgery as the target cohort, and the index date and cohort end date were defined as the first radiotherapy after breast surgery and the index date+180 days, respectively. The comparative cohort consisted of patients who underwent breast cancer surgery without radiotherapy. The index date of comparator cohort was defined as the earliest admission date for breast cancer surgery, and the discharge date was defined as the cohort end date. Both cohorts were diagnosed with only breast cancer, without other cancers, before the index date+30 days (Fig. 1). S1 Table shows the concept sets used to define the cohorts.

3. Outcome and time-at-risk period

The outcome was defined as a diagnosis of any cancer type other than breast cancer according to the definition of secondary primary malignancy (SPM). The time-at-risk (TAR) period for the outcome was defined as the time from 5 years after the end date to the last observation date. Patients with observed outcomes before start of TAR and a shorter observation period than the TAR period were excluded from this study.

The concept sets used to define the outcome cohorts, patient characteristics, and outcome types are shown in S1 Table.

4. Negative controls

We performed empirical calibration using 112 negative control outcomes that were selected by manual review (S2 Table). Negative control outcomes assumed no causal effect with the outcome of interest, and its empirical null distribution shows random and systemic bias [9]. We used 112 negative controls that were not associated with radiotherapy exposure to compute the calibrated p-value.

5. Statistical analysis

This study was conducted based on the population-level estimation methodology from the OHDSI. We performed PS

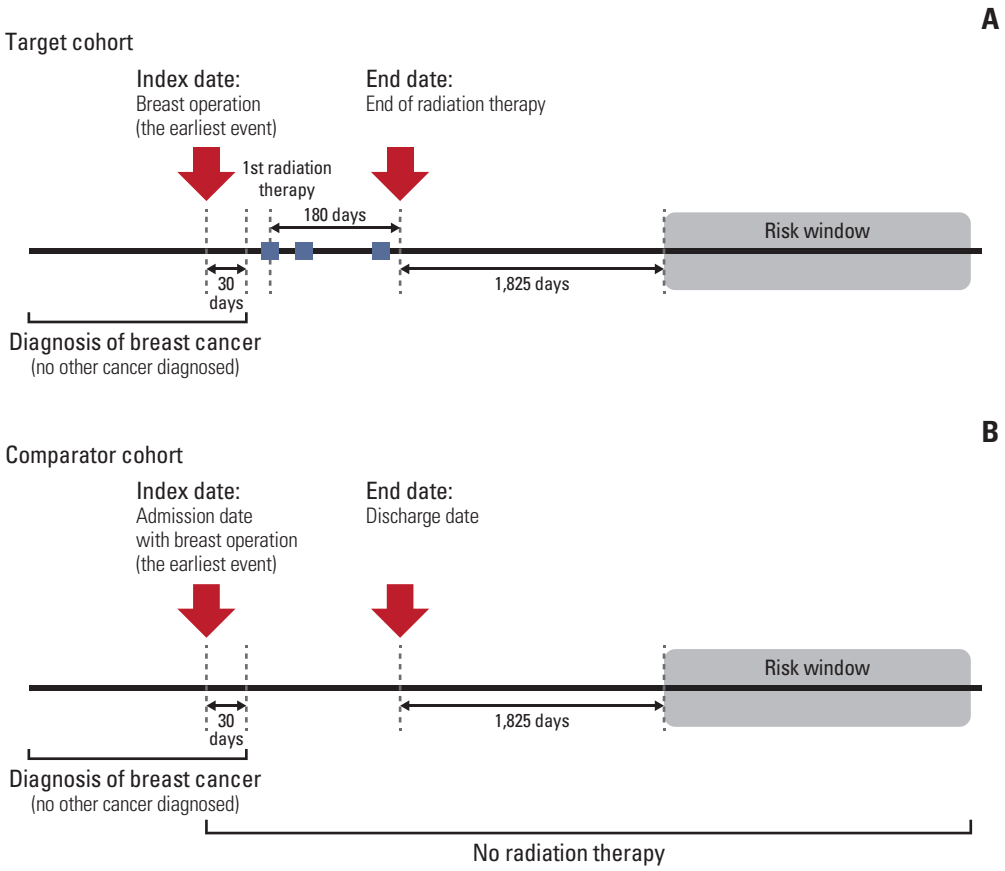

Fig. 1. Definition of target and comparator cohorts. (A) Target cohort is from the date of the first breast cancer surgery (index date) to the date of last radiation treatment (end date). (B) Comparator cohort is the date from admission (index date) to discharge date (end date) for the first breast cancer surgery was defined as each cohort enroll period.

adjustment to improve the characteristic balance between the target and comparator and to adjust the confounder. PSs were estimated using L1 logistic regression tuned by 10-fold validation and removed if a value fell outside the range of 0.25-0.75. Using a caliper of 0.2 on the logit scale of the PS distribution, we matched the target cohort and comparator cohorts in a 1:1 ratio. The covariates selected were demographics (5 groups by age), previous diagnosis (at any time prior), and risk scores (Charlson comorbidity index [CCI], diabetes complications severity index [DCSI], and CHA_2DS_2-VASc Score (CHA_2DS_2VASc), except for the covariates used in the generation of target and comparator cohorts.

The Cox proportional hazard model estimated the hazard ratios (HRs) of the outcome and negative controls for the matched population. Finally, we reported the HRs of the outcomes with 95% CI, nominal p-values, and empirically calibrated p-values. Meta-analysis was used to integrate the estimates for each data source for all analyses. We selected between random- and fixed-effect models according to the homogeneity test results.

In addition, we reviewed potential confounders associated with secondary cancer risk, including other treatments, genetic data, comorbidities, radiation dose, lifestyle data, and cancer stage [14]. Although some of these factors were addressed, other factors, such as genetic factor, lifestyle factor, and cancer stage, could not be included in the analysis due to limitations in the available data, either through 1:1 matching analysis or sensitivity analysis. These limitations are acknowledged and discussed in this manuscript.

6. Sensitivity analyses

Some of the potential confounders such as age at diagnosis, latency period, and radiation dose that were not adequately addressed in the main analysis were further analyzed through sensitivity analyses. Sensitivity analyses were performed according to age at diagnosis (> 50 vs. ≤ 50 years), latency (> 3,650 days vs. ≤ 3,650 days), SPM sites between the target and comparator groups, and radiotherapy fraction number within the target group. We additionally created a target and comparator cohort with the date of the first breast

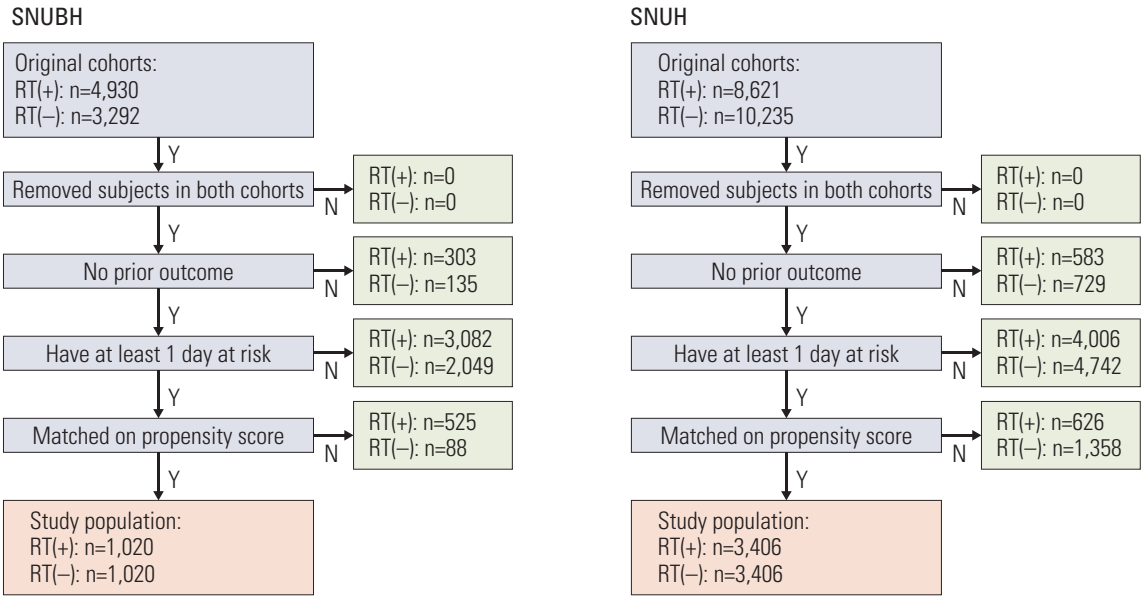

Fig. 2. Attrition of each database to select the final study population to be used for population-level estimation analysis from the target and comparator cohort generated from each medical institution's database. RT, radiation therapy; SNUBH, Seoul National University Bundang Hospital; SNUH, Seoul National University Hospital.

cancer diagnosis set as the index date.

Results

1. Patient population

A total of 27,078 women diagnosed with breast cancer were included in this study (target group, 13,551 patients; comparator group, 13,527 patients). After 1:1 PS matching using the equipoised trim method, 1,020 patients were matched from each group in the SNUBH CDM and 3,406 patients were matched in the SNUH CDM. Fig. 2 shows the attrition of each database. The baseline characteristics of the study population in each group, before and after matching, are shown in Table 1. Fig. 3 shows the standardized mean differences of all covariates before and after the matching. After matching, the median age of both groups was 49 years, and patients in their 40s and 50s accounted for approximately 70.9% of the study population. There were no significant differences in the covariates used for matching, including age group, prior diagnosis, CCI, DCSI, and CHA_2DS_2VASc scores. The follow-up period for each cohort was 3,020 and 3,074 days on median in the target and comparator groups, respectively. The demographics for each SNUBH and SNUH database are presented in S3 and 4 Tables, respectively.

2. Risk of secondary primary malignancy

Table 2 presents the results of each database and meta-analysis. The incidence rates were 12.15 per 1,000 person-years and 15.39 per 1,000 person-years in the target and comparator groups, respectively. The HRs in the SNUBH (HR, 1.31; 95% CI, 0.79 to 2.19; p=0.3) cohort were not significantly different, but the risk of secondary cancer was significantly higher in breast cancer who did not receive radiation treatment in SNUH (HR, 0.74; 95% CI, 0.6 to 0.9; p=0.003). In the meta-analysis, there was no significant difference in SPM between target and comparator groups (HR, 0.93; 95% CI, 0.54 to 1.63; p=0.811).

The HRs for the second cancer site groups are shown in Tables 3 and 4. The most common sites across the databases were the digestive, respiratory, genitourinary system, and endocrine systems. HRs were estimated through meta-analysis for each site, and the frequency of events was generally similar between the target and comparator groups.

3. Sensitivity analyses

Table 5 presents the results of the sensitivity analysis. Age at diagnosis was divided into two groups: > 50 and ≤ 50 years. Although there was no significant difference in the SNUBH cohort (HR, 1.36; 95% CI, 0.63 to 3.04; p=0.445), the SNUH group (HR, 0.64; 95% CI, 0.49 to 0.83; p=0.001) had a significantly higher risk of secondary cancer among patients under the age of 50 who did not receive radiation therapy. The estimates from meta-analysis were significantly different for SPMs in patients aged < 50 years (HR, 0.69; 95% CI, 0.54 to 0.89; p=0.004). Women who were diagnosed with

Table 1. Baseline characteristics of overall database

	Before matching			After matching		
	RT (+) (n=13,551)	RT (–) (n=13,527)	SMD	RT (+) (n=4,426)	RT (–) (n=4,426)	SMD
All women	13,551 (100)	13,527 (100)		4,426 (100)	4,426 (100)	
Age at diagnosis (yr)						
Median (IQR)	50 (44-57)	51 (44-60)	–0.14	49 (43-56)	49 (43-56)	–0.03
Mean±SD	50.8±10.2	52.2±11.5		49.4±9.6	49.6±9.7	
Age at diagnosis (yr)						
< 30	158 (1.2)	148 (1.1)	0.01	68 (1.5)	53 (1.2)	0.02
30-39	1,491 (11.0)	1,430 (10.6)	0.01	546 (12.3)	543 (12.3)	0.00
40-49	4,960 (36.6)	4,605 (34.0)	0.04	1,759 (39.7)	1,752 (39.6)	0.00
50-59	4,255 (31.4)	3,906 (28.9)	0.04	1,380 (31.2)	1,385 (31.3)	0.00
60-69	2,081 (15.4)	2,218 (16.4)	–0.02	555 (12.5)	559 (12.6)	0.00
70-79	550 (4.1)	1,043 (7.7)	–0.13	113 (2.6)	121 (2.7)	–0.01
≥ 80	56 (0.4)	177 (1.3)	–0.09	5 (0.1)	13 (0.3)	–0.04
Year at diagnosis						
2000s	2,709 (20.0)	3,289 (24.3)	–0.09	1,667 (37.7)	2,174 (49.1)	–0.19
2010-2014	3,970 (29.3)	3,713 (27.4)	0.03	2,695 (60.9)	2,015 (45.5)	0.26
2015-2019	5,923 (43.7)	5,490 (40.6)	0.05	64 (1.4)	236 (5.3)	–0.20
2020	949 (7.0)	1,035 (7.7)	–0.01	0	1 (0.0)	–0.02
Surgery						
Lumpectomy	11,914 (87.9)	6,557 (48.5)	0.82	4,048 (91.4)	2,669 (60.3)	0.70
Mastectomy	1,523 (11.2)	6,608 (48.8)	–0.79	320 (7.2)	1,671 (37.7)	–0.71
Excision	116 (0.9)	362 (2.7)	–0.12	59 (1.3)	86 (1.9)	–0.04
Other	0	2 (0.0)	–0.02	0	1 (0.0)	–0.02
Risk score						
CCI						
0	741 (5.5)	1,663 (12.3)	–0.01	273 (6.2)	277 (6.3)	–0.07
1-2	11,971 (88.3)	11,043 (81.6)		3,986 (90.1)	3,922 (88.6)	
3-4	737 (5.4)	710 (5.2)		149 (3.4)	204 (4.6)	
5+	102 (0.8)	111 (0.8)		18 (0.4)	23 (0.5)	
DCSI						
0-1	13,115 (96.8)	12,991 (96.0)	–0.07	4,322 (97.7)	4,282 (96.7)	–0.04
2-3	394 (2.9)	463 (3.4)		95 (2.1)	131 (3.0)	
4+	42 (0.3)	73 (0.5)		9 (0.2)	13 (0.3)	
CHA_2DS_2VASc						
0-1	11,831 (87.3)	11,108 (82.1)	–0.18	4,016 (90.7)	4,004 (90.5)	–0.03
2-3	1,640 (12.1)	2,289 (16.9)		395 (8.9)	406 (9.2)	
4+	80 (0.6)	130 (1.0)		15 (0.3)	16 (0.4)	
Follow-up days (index end date to Obs end date)						
Median (IQR)	1,593 (620-2,768)	1,672 (735-2,898)	–0.09	3,020.5 (2,356-3,939)	3,074 (2,360.25-3,988.5)	–0.05
Mean±SD	1,825.5±1,421.3	1,941.4±1,449.3		3,199.4±998.2	3,259.7±1,055.5	
1st BreastCx Dx to 1st surgery						
Median (IQR)	26 (15-51)	29 (16-51)	–0.08	21 (12-36)	22 (13-37)	–0.08
Mean±SD	70.8±217.8	89.2±349.9		57.2±169	70.6±254.5	

(Continued to the next page)

Table 1. Continued

	Before matching			After matching		
	RT (+) (n=13,551)	RT (–) (n=13,527)	SMD	RT (+) (n=4,426)	RT (–) (n=4,426)	SMD
Death						
No.	134	176		16	32	
Median (IQR, day)	868 (337.75-1,626.75)	1,042 (496-1,977)	–0.21	3,192.5 (2,726.25-3,783.5)	2,998 (2,560-3,625.75)	0.28
Mean±SD (day)	1,176.5±1,127.8	1,382.6±1,139		3,331.3±1,024.4	3,053.9±736.5	

Values are presented as number (%) unless otherwise indicated. BreastCx, breast cancer; CCI, Charlson comorbidity index; CHA_2DS_2VASc, CHA_2DS_2-VASc Score; DCSI, diabetes complication severity index; Dx, diagnosis; IQR, interquartile range; Obs, observation period; RT, radiotherapy; SD, standard deviation; SMD, standard mean difference.

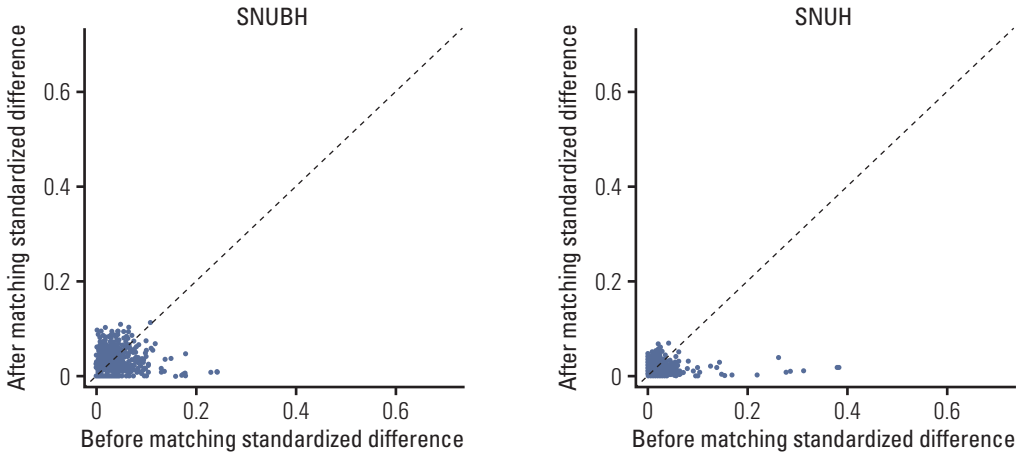

Fig. 3. Standardized difference of the mean before and after matching. Comparison of before and after applying propensity-score matching in the data to be used in population-level estimation analysis. After matching, most of the standardized difference of the mean was distributed below 0.1, confirming that the matching was successful. SNUBH, Seoul National University Bundang Hospital; SNUH, Seoul National University Hospital.

breast cancer under 50 years and with radiation therapy had a lower risk of SPM than those without radiation therapy.

Latency was divided into two groups: > 3,650 days and ≤ 3,650 days. All estimates of both the database and meta-analysis showed no significant differences in latency.

Cumulative radiation therapy fraction number was divided into two groups: > 34 times and ≤ 34 times. Most patients had 33 sessions in both databases, with a mean 200 cGy per session for SNUBH and 180 cGy per time for SNUH. The estimates of both the database and meta-analysis were not significantly different for cumulative radiation therapy.

The cohorts that defined the index date as the breast cancer diagnosis date were significantly different only in the SNUH group (HR, 0.72; 95% CI, 0.59 to 0.88; p=0.001). The estimate of meta-analysis was 1.04 (95% CI, 0.48 to 2.23), which showed no difference in the risk of secondary cancer development according to radiation therapy.

Discussion

In this study, surgically resected people with breast cancer treated with radiotherapy had no risk (HR, 0.93; 95% CI, 0.54 to 1.63; p=0.81) of secondary non-breast cancer compared to PS-matched people with breast cancer who were surgically resected and not treated with radiotherapy. This trend remained in the subgroup analyses according to the age at diagnosis, latency period, and fraction number of radiotherapy.

The risk of secondary cancers related to radiotherapy in patients with breast cancer is controversial. A previous study that used the Surveillance, Epidemiology, and End Results (SEER) cancer registry data from 1973 to 2000 concluded that most secondary solid cancers in patients with breast cancer were not associated with radiotherapy [11]. However, the esophagus, pleura, bone, soft tissue, and contralateral breast

Table 2. Results of secondary cancer risk by postoperative radiotherapy after breast cancer surgery

Database	Subjects (n)	Outcomes (n)		IR (per 1,000 PY)		HR (95% CI)	p-value	Calibrated p-value
		Target	Comparator	Target	Comparator			
SNUBH	1,020	34	28	10.52	8.37	1.31 (0.79-2.19)	0.300	0.730
SNUH	3,406	162	230	12.55	17.14	0.74 (0.60-0.90)	0.003	< 0.001
Meta-analysis[a)]	4,426	196	258	12.15	15.39	0.93 (0.54-1.63)	0.811	

CI, confidence interval; HR, hazard ratio; IR, incidence rate; PY, person-years; SNUBH, Seoul National University Bundang Hospital; SNUH, Seoul National University Hospital. [a)]Meta-analysis results are based on a random-effects model determined by heterogeneity tests (I^2).

Table 3. Results by sites of secondary cancer: Seoul National University Hospital

Site group	Outcomes (n)		HR (95% CI)
	Target	Comparator	
Cardiovascular system	1	3	0.35 (0.02-2.73)
Digestive system	34	46	0.79 (0.51-1.23)
Endocrine	21	24	0.93 (0.51-1.67)
Genitourinary system	19	19	1.08 (0.57-2.06)
Lip, oral cavity, and pharynx	2	4	0.58 (0.08-2.98)
Musculoskeletal system/Connective tissue	21	45	0.49 (0.29-0.81)
Nervous system	6	23	0.27 (0.10-0.63)
Peritoneum	2	1	2.19 (0.21-47.27)
Respiratory system	25	29	0.91 (0.53-1.55)
Skin	4	10	0.42 (0.12-1.22)
Thorax	0	6	-
Lymphoid, hematopoietic	17	7	0.43 (0.17-1.00)
Unknown	2	3	-

Cardiovascular system (pericardium and vascular tissue), digestive system (ampulla of Vater, biliary tract, colon, gallbladder, large intestine, liver, pancreas, stomach, rectum), endocrine system (adrenal gland, thymoma, thyroid, thymus), genitourinary system (bladder, endometrium, kidney, ovary, pelvis, ureter, urethra, uterine cervix, uterus, uterine cervix), lip, oral cavity, pharynx (buccal mucosa, palate, tongue, tonsil, neck, parotid gland), musculoskeletal system/connective tissue (bone, soft tissues, vertebral column), nervous system (brain), peritoneum (peritoneal cavity, peritoneum), respiratory system (lung, pleura), skin (finger, skin), thorax (chest, mediastinum, sternum), lymphoid, hematopoietic system (leukemia, lymph nodes, lymphoma, myeloma). CI, confidence interval; HR, hazard ratio.

regions affected by high-dose radiation had a greater incidence of secondary cancers than those affected by surgery alone. A single cohort study [10] that evaluated the effect of radiotherapy on the risk of secondary cancer reported that radiotherapy increased the RR of all secondary cancers, but this was not statistically significant (RR, 1.22; 95% CI, 0.88 to 1.69). A study on a South Korean population showed no significant correlation between radiation exposure and the development of secondary cancers (by site) in patients with breast cancer [2]. However, a few studies have reported that radiotherapy increases the risk of secondary cancers in women with surgically resected breast cancer [3-5,12,15,16]. In particular, anatomic sites affected by high-dose radiation, age at diagnosis, and longer latency periods were considered prominent covariates for the development of secondary cancer in patients with breast cancer.

These studies estimated and compared the RR and standardized incidence ratio between a group of secondary cancers after radiotherapy and surgery and a group of secondary cancers with surgery alone without a strategy to adjust for selection bias [4,5,10-12]. Rather, they only performed multivariate analyses to control for the confounding factors they had and were aware of. Thus, it is difficult to conclude that the differences between the two groups were due to radiotherapy. In the current study, PS matching and negative controls were used to address selection bias. We used sex, age groups, previous diagnosis, and risk scores (CCI, DCSI, CHA_2DS_2VASc) as covariates for matching and tested pre-specified false hypotheses using 112 negative control outcomes.

Table 4. Results by sites of secondary cancer: Seoul National University Bundang Hospital

Site group	Outcomes (n)		HR (95% CI)
	Target	Comparator	
Digestive system	9	10	0.83 (0.32-2.12)
Endocrine	6	2	3.00 (0.69-20.50)
Genitourinary system	4	5	0.86 (0.21-3.28)
Lip, oral cavity, and pharynx	3	1	-
Musculoskeletal system/Connective tissue	1	5	0.20 (0.01-1.22)
Nervous system	0	1	-
Respiratory system	8	4	1.97 (0.62-7.38)
Lymphoid, hematopoietic	0	2	-
Unknown	1	1	-

Digestive system (ampulla of Vater, biliary tract, colon, gallbladder, liver, pancreas, stomach, rectum), endocrine system (thyroid, thymus), genitourinary system (bladder, endometrium, ovary, ureter, uterus, uterine cervix), lip, oral cavity, pharynx (tongue, neck, parotid gland), musculoskeletal system/connective tissue (bone, soft tissues, vertebral column), nervous system (brain), respiratory system (lung, pleura). CI, confidence interval; HR, hazard ratio.

Table 5. Sensitivity analyses of secondary cancer risk by alternative index date, latency, age subgroup, and radiation fraction number

Subgroup	Database	Outcomes (n)/Subjects (n)		IR (per 1,000 PY)		HR (95% CI)	p-value
		Target	Comparator	Target	Comparator		
Index date[a] revise	SNUBH	48/1,004	28/1,004	13.7	8.82	1.58 (0.99-2.57)	0.058
	SNUH	170/3,301	218/3,301	12.3	17.02	0.72 (0.59-0.88)	0.001
	Meta[a]	218/4,305	246/4,305	17.33	15.98	1.04 (0.48-2.23)	0.930
Latency (yr)							
< 10	SNUBH	18/686	9/676	16.27	8.39	1.93 (0.89-4.51)	0.112
	SNUH	46/1,906	72/1,828	13.19	21.63	0.61 (0.42-0.88)	0.009
	Meta[a]	64/2,592	81/2,504	13.93	18.4	1.02 (0.33-3.15)	0.967
≥ 10	SNUBH	13/248	16/248	7.43	8.68	0.93 (0.43-1.96)	0.846
	SNUH	94/1,107	136/1,248	12.22	15.75	0.77 (0.59-1.01)	0.058
	Meta[a]	107/1,355	152/1,496	11.33	14.51	0.79 (0.62-1.01)	0.063
Age (yr)							
< 50	SNUBH	15/562	11/548	8.18	5.88	1.36 (0.63-3.04)	0.445
	SNUH	90/1,811	141/1,800	11.89	18.56	0.64 (0.49-0.83)	0.001
	Meta[a]	105/2,373	152/2,348	11.17	16.05	0.69 (0.54-0.89)	0.004
≥ 50	SNUBH	19/458	17/472	13.59	11.53	1.35 (0.69-2.69)	0.388
	SNUH	72/1,595	89/1,606	13.50	15.29	0.89 (0.65-1.22)	0.482
	Meta[a]	91/2,053	106/2,078	13.52	14.53	0.96 (0.72-1.28)	0.779
Radiotherapy fraction number							
	SNUBH	10/280	25/736	10.89	10.75	0.99 (0.46-2.03)	0.992
	SNUH	20/285	136/2,967	17.36	12.20	1.38 (0.84-2.16)	0.181
	Meta[a]	30/565	161/3,703	14.49	11.95	1.26 (0.84-1.88)	0.258

CI, confidence interval; HR, hazard ratio; IR, incidence rate; PY, person-years; SNUBH, Seoul National University Bundang Hospital; SNUH, Seoul National University Hospital. [a]Date of first breast cancer diagnosis.

Among previous studies that analyzed the effects of radiotherapy on people with breast cancer using SEER cancer registries, only one study used age, marital status, race, T category, N category, tumor grade, estrogen receptor status, progesterone receptor status, and treatment options as covariates for PS matching [17].

As the treatment guidelines for patients with breast cancer have been revised or updated, the amount of radiation

applied to the patient has decreased, and their effect has improved [4]. Whereas previous studies collected data on people with breast cancer from the 1970s or 1980s to recent years [4,5,10,11], the current study analyzed data of people with breast cancer admitted after 2003 considering the revision of clinical practice guidelines. Therefore, changes in the regimen may have influenced the risk of developing secondary cancer in patients with breast cancer.

In this study, there was no significant relationship between the secondary cancer site and radiotherapy, which is consistent with the findings of a previous study conducted at three tertiary medical centers in South Korea [2]. However, the prevalence of secondary cancers of the digestive, respiratory, genitourinary system, and endocrine systems was higher than that of other secondary cancers in both databases. Previous studies also found that thyroid cancer (an endocrine system cancer) and gastric cancer (a digestive system cancer) were frequently observed after a diagnosis of primary breast cancer [2,18]. This finding may reflect the high prevalence of thyroid, lung, and gastric cancer in South Korea. As of 2019, these were the most common cancers in South Korea. In particular, it has been reported that thyroid cancer tends to develop secondary to radiotherapy [4,19].

We defined the end date in the target cohort as 180 days after the first radiotherapy date. According to the 8th revision of the breast cancer clinical practice guidelines published in 2019 by the Korean Breast Cancer Society, 45-50 Gy of radiotherapy is recommended for 5-6 weeks after breast-conserving surgery. In the database, 75% of the population in the target cohort completed radiotherapy within 35 days of the first radiotherapy. Since all last radiotherapy dates in the target cohort were included within 180 days from the first radiotherapy after the removal of outliers, the cohort end date for the target cohort was set to be 180 days from the date the patient received their first radiotherapy.

There are several limitations in the use of OMOP CDM data to evaluate radiotherapy-related secondary cancer risks in patients with breast cancer. First, despite applying strategies such as PS matching and negative controls to adjust for selection bias, genetic factors affecting breast cancer susceptibility (e.g., *BRCA*, p53, *PTEN*), history of smoking, lifestyle factors, and family history of cancer could not be included as matching confounders due to a lack of CDM data. Other treatments such as hormone therapy were also not considered in the main analysis, as the study focused on the presence or absence of radiotherapy. Second, we used "fraction number" instead of "cumulative dose" of radiotherapy to analyze the relationship between the radiation dose and the occurrence of secondary cancer due to the absence of converted CDM data. Third, cancer stage information was stored in an unstructured text format and could not be utilized, as it was not standardized within the CDM structure. This is a common limitation of using real-world hospital data. Fourth, there were still a limited number of secondary cancer cases obtained by collecting the study populations from two tertiary university hospitals. Fifth, although the follow-up period exceeded 5 years (as shown in Table 1), more than 50% of patients in the initial cohort had no follow-up after 3 years, which is the TAR period, and therefore the risk of secondary cancer may be underestimated or overestimated. In addition, the number of outcome events may still be insufficient to determine clinical significance. Further studies with a larger sample size are needed to validate these findings. However, this finding was partially supplemented by a meta-analysis.

This study analyzed the risk of secondary cancer in patients who underwent breast cancer surgery depending on whether they received radiation therapy. Although the incidence of certain secondary cancers was slightly higher in the radiotherapy group, there was no significant difference in the overall risk of secondary cancer. In clinical practice, radiotherapy, when administered according to treatment guidelines, does not appear to increase the risk of secondary cancer. This information may help reduce concerns among patients and their families regarding the safety of radiotherapy. In the future, analyses including data specific to patients undergoing cancer treatment are needed.

Electronic Supplementary Material

Supplementary materials are available at Cancer Research and Treatment website (https://www.e-crt.org).

Ethical Statement

The study protocol was approved by the institutional review boards (SNUBH No. X-1910-570-901, SNUH No. E-1910-043-1067), and informed consent was exempted from institutional review boards. All aspects of the study were conducted in accordance with the Declaration of Helsinki.

Author Contributions

Conceived and designed the analysis: Yoo S, Kim B, Kim K (Kyubo Kim), Kim K (Kwangsoo Kim), Song E, Kim J, Ryoo HG, Paeng JC, Choi IY, Ko S, Yoo IR, Lee HY, Park RW.
Collected the data: Kim S, Yoo S.
Contributed data or analysis tools: Kim S, Boo D, Yoo S, Kim B.
Performed the analysis: Kim S, Boo D.
Wrote the paper: Kim S, Boo D.
Writing review: Kim K (Kyubo Kim), Kim K (Kwangsoo Kim), Song E, Kim J, Ryoo HG, Paeng JC, Choi IY, Ko S, Yoo IR.
Inspected the manuscript critically and took part in the revision of manuscript: Lee HY, Park RW.

ORCID iDs
Seok Kim : https://orcid.org/0000-0003-4996-8613
Dachung Boo : https://orcid.org/0000-0002-6855-3723
Rae Woong Park : https://orcid.org/0000-0003-4989-3287
Ho-Young Lee : https://orcid.org/0000-0001-6518-0602

Conflicts of Interest
Conflict of interest relevant to this article was not reported.

Funding
This research was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health and Welfare, Republic of Korea (grant numbers: HI19C0378, RS-2022-KH125200). This work was also partly supported by the Technology Innovation Program (grant number: 20004927), funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Author Details
[1]Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam, [2]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, [3]Department of Biomedical Systems Informatics, Yonsei University College of Medicine Institute for Innovation in Digital Healthcare, Seoul, [4]Department of Radiation Oncology, Seoul National University Bundang Hospital, Seongnam, [5]Department of Transdisciplinary Medicine, Seoul National University Hospital, Seoul, [6]Center for Data Science, Healthcare AI Research Institute, Seoul National University Hospital, Seoul, [7]Department of Medicine, Seoul National University College of Medicine, Seoul, [8]Department of Data Science Research, Innovative Medical Technology Research Institute, Seoul National University Hospital, Seoul, [9]Interdisciplinary Program in Bioengineering, Seoul National University, Seoul, [10]Department of Nuclear Medicine, Seoul National University Hospital, Seoul, [11]Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul, [12]Division of Nuclear Medicine, Department of Radiology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, [13]Department of Nuclear Medicine, Seoul National University Bundang Hospital, Seongnam, [14]Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul, Korea

References

1. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. CA Cancer J Clin. 2019;69:363-85.
2. Jung HK, Park S, Kim NW, Lee JE, Kim Z, Han SW, et al. Development of second primary cancer in Korean breast cancer survivors. Ann Surg Treat Res. 2017;93:287-92.
3. Grantzau T, Overgaard J. Risk of second non-breast cancer after radiotherapy for breast cancer: a systematic review and meta-analysis of 762,468 patients. Radiother Oncol. 2015;114:56-65.
4. Burt LM, Ying J, Poppe MM, Suneja G, Gaffney DK. Risk of secondary malignancies after radiation therapy for breast cancer: Comprehensive results. Breast. 2017;35:122-9.
5. Bazire L, De Rycke Y, Asselain B, Fourquet A, Kirova YM. Risks of second malignancies after breast cancer treatment: long-term results. Cancer Radiother. 2017;21:10-5.
6. Wang J. To use or not to use propensity score matching? Pharm Stat. 2021;20:15-24.
7. Ye Y, Kaskutas LA. Using propensity scores to adjust for selection bias when assessing the effectiveness of alcoholics anonymous in observational studies. Drug Alcohol Depend. 2009;104:56-64.
8. Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? JAMA. 2013;309:241-2.
9. Arnold BF, Ercumen A. Negative control outcomes: a tool to detect bias in randomized trials. JAMA. 2016;316:2597-8.
10. Zhang W, Becciolini A, Biggeri A, Pacini P, Muirhead CR. Second malignancies in breast cancer patients following radiotherapy: a study in Florence, Italy. Breast Cancer Res. 2011;13:R38.
11. Berrington de Gonzalez A, Curtis RE, Gilbert E, Berg CD, Smith SA, Stovall M, et al. Second solid cancers after radiotherapy for breast cancer in SEER cancer registries. Br J Cancer. 2010;102:220-6.
12. Grantzau T, Mellemkjaer L, Overgaard J. Second primary cancers after adjuvant radiotherapy in early breast cancer patients: a national population based study under the Danish Breast Cancer Cooperative Group (DBCG). Radiother Oncol. 2013;106:42-9.
13. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assoc. 2012;19:54-60.
14. Travis LB. The epidemiology of second primary cancers. Cancer Epidemiol Biomarkers Prev. 2006;15:2020-6.
15. Sun Q, Chen Y, Li T, Ni B, Zhu X, Xu B, et al. Risk and prognosis of secondary esophagus cancer after radiotherapy for breast cancer. Sci Rep. 2023;13:3968.
16. Okonogi N, Karasawa K, Nitta Y, Mori Y, Murata K, Wakatsuki M, et al. Risk of secondary malignancy after radiotherapy for breast cancer: long-term follow-up of Japanese patients with breast cancer. Breast Cancer Res Treat. 2022;194:561-7.

17. Li Y, Chen M, Pardini B, Dragomir MP, Lucci A, Calin GA. The role of radiotherapy in metaplastic breast cancer: a propensity score-matched analysis of the SEER database. J Transl Med. 2019;17:318.
18. Kim JY, Song HS. Metachronous double primary cancer after treatment of breast cancer. Cancer Res Treat. 2015;47:64-71.
19. Joseph KR, Edirimanne S, Eslick GD. The association between breast cancer and thyroid cancer: a meta-analysis. Breast Cancer Res Treat. 2015;152:173-81.