의학박사 학위논문

건강보험 심사평가원 청구 자료를 이용한 유방암 치료를 위한 국내 의료 여행에 관한 연구

Domestic medical travel from non-Seoul regions to Seoul for breast cancer treatment: a nationwide cohort study

울산대학교 대학원

의 학 과

정 재 호

Domestic medical travel from non-Seoul regions to Seoul for breast cancer treatment: a nationwide cohort study

지 도 교 수 정 경 해

이 논문을 의학박사 학위 논문으로 제출함

2022 년 2 월

울산대학교 대학원

의 학 과

정 재 호

정재호의 의학박사 학위논문을 인준함

심사위원 박 연 희 인

심사위원 정 경 해 인

심사위원 김 규 표 인

심사위원 정 일 용 인

심사위원 김 정 은 인

울 산 대 학 교 대 학 원

2022 년 2 월

Abstract

Purpose

To investigate the annual frequencies and patterns of breast cancer treatment by region, assess the trend of domestic medical travel from non-Seoul regions to Seoul for initial treatment, and identify factors associated with medical travel in breast cancer patients.

Materials and Methods

A nationwide retrospective cohort study was performed using the Health Insurance Review and Assessment data of Republic of Korea. Patients were classified according to the regions in which they underwent pathologic examination (Seoul vs. metropolitan cities vs. other regions). Frequencies of pathologic examinations, diagnosis, and treatment were analyzed according to regions. Domestic medical travel was analyzed according to cities and provinces, and factors associated with medical travel were investigated.

Results

A total of 150,709 breast cancer survivors who were diagnosed between January 2010 and December 2017 were included. The annual frequencies of pathologic examinations, diagnosis, and treatment increased in all regions, and the difference in the frequencies of surgery between Seoul and non-Seoul regions increased over time. The rate of medical travel from non-Seoul regions to Seoul increased from 14.2% (1,161/8,150) in 2010 to 19.8% (2,762/13,964) in 2017. Approximately a quarter of patients from other regions traveled to Seoul, and over 40% of patients from Chungbuk, Gyeongbuk, and Jeju regions traveled to Seoul for treatment in 2017. Young age and regions other than metropolitan cities were significantly associated with the likelihood of domestic medical travel. Patients covered by medical aid or those with other diseases were significantly less likely to travel to Seoul for breast cancer treatment.

Conclusion

The number of patients traveling to Seoul for breast cancer treatment increased over time. Young age, comorbidities, and type of insurance were significant factors associated with medical travel.

Table of Contents

Abstract i

Table of contents iii

List of Tables and Figures iv

Introduction 1

Materials and Methods 3

Data source and extraction 3

Study population 4

Variables and operational definitions 6

Statistical analysis 7

Results 8

Baseline characteristics 8

Annual frequencies of pathologic examinations, diagnosis, and treatment by regions 8

Domestic medical travel for breast cancer treatment 8

Factors associated with domestic medical travel for breast cancer treatment 18

Discussion 28

References 32

Appendices 36

1. IRB documents 36

2. R codes for analysis 40

국문 초록 64

List of Tables and Figures

Table 1. Characteristics of study population 9

Table 2. Frequencies of breast cancer management by region using quasi-Poisson regression analysis 11

Table 3. Domestic medical travel after diagnosis 19

Table 4. Domestic medical travel from non-Seoul regions to Seoul for initial treatment (2010 – 2014, continued) 21

Table 5. Factors associated with medical travel from non-Seoul to Seoul using multivariable logistic regression analysis 27

Figure 1. Study population 5

Figure 2. Annual number of patients who got pathologic examination by regions 12

Figure 3. Annual number of patients who diagnosed with breast cancer by regions 13

Figure 4. Annual number of patients who got initial treatment of breast cancer by regions .. 14

Figure 5. Annual number of patients who underwent breast cancer surgery by regions 15

Figure 6. Annual number of patients who treated with chemotherapy of breast cancer by regions 16

Figure 7. Annual number of patients who got radiotherapy of breast cancer by regions 17

Figure 8. Sankey diagram of domestic medical travel 20

Figure 9. Regional distribution of medical travel to Seoul 25

Figure 10. Annual proportion of medical travel to Seoul by region 26

Introduction

The incidence of breast cancer has rapidly increased and breast cancer has become the major type of cancer among women worldwide (1). In 2012, approximately 1.7 million women were diagnosed with breast cancer, but with a sharp rise worldwide, the number of newly diagnosed female breast cancer cases were approximately 2.3 million, according to GLOBOCAN 2020, accounting for nearly one in four female cancer cases. (1, 2). This increasing trend is observed in many countries, also in Asian countries such as Republic of Korea. In Republic of Korea, more than 26,000 patients were newly diagnosed with breast cancer in 2017. Breast cancer is the most common type of cancer in women (3), of which about 70% of patients are diagnosed with stage I/II (4). This increase in breast cancer eventually led to increased demands for breast cancer surgery and other adjuvant treatments (5).

In order to provide optimal treatment to the patients, breast cancer treatment should be provided by a multidisciplinary team, consisting of medical oncologists, breast surgeons, radiation oncologists, breast radiologists, and breast pathologists (6, 7). A timely access to definitive care, along with multidisciplinary approach, is important for improving patient outcomes (8-11).

Richard et al. conducted systematic review for the influence of delay on survival in patients with breast cancer, and concluded that there is substantial evidence that a treatment delay of 3 months or more in breast cancer patients is associated with a lower survival rate (8). One of the factors that can affect delay in treatment is domestic medical travel, which is moving from one region to other regions for treatment after diagnosis.

Stitzenberg et al. reported that there has been extensive centralization of complex cancer surgery over the past decade, and centralization is increasing patient travel (12). Therefore, the effect of centralization on treatment delay and subsequent lower survival in cancer patients should be considered for establishing national health policy.

Studies on domestic medical travel have been conducted primarily in cancers other than breast cancer (12-14). In Republic of Korea, centralization with domestic medical travel from non-capital regions to Seoul for cancer treatments has been increasing (15-17). However, little is known about the trend of medical travel for patients with breast cancer in Republic of Korea.

Therefore, I conducted a nationwide retrospective cohort study using the national claims data to investigate the patterns and frequencies of breast cancer treatments by region, assess domestic medical travel over time, and identify the factors associated with medical travel in patients with breast cancer.

Materials and Methods

Data source and extraction

The Health Insurance Review and Assessment (HIRA) assesses the medical services provided in Republic of Korea and decides their reimbursement. Republic of Korea's national health insurance is a government-managed public health insurance, with about 96% of the total population and the 4% (mostly low-income households) receiving medical aid (18). Accordingly, the HIRA database archives the nationwide claims data from all healthcare providers.

The patient samples of HIRA provide 5 tables: 1) 200 table (general information: beneficiary identification (ID), age, gender, type of insurance, indicators for inpatient/outpatient care, medical institution ID, type of medical institution, etc.); 2) 300 table (healthcare services: procedures, inpatient prescriptions, operations, diagnostic tests, etc.); 3) 400 table (diagnostic information: indicator for major diagnosis, department, etc.); 4) 530 table (outpatient prescriptions: classification type, service codes, drug codes, etc.); and 5) table of providers (location, zip code, name of providers, types of provider, etc.) (19, 20). The diagnosis data is archived using the International Classification of Diseases, 10th revision (ICD-10) (21), and newly diagnosed cancers are coded with the reimbursement claim code (V193).

Study population

I extracted all available data of 254,796 patients who were coded with the ICD-10 C50 (malignant neoplasm of breast) (21) and V193 codes between January 2010 and December 2017. After washing out prevalent cases between January 2008 and December 2009 (n=87,646), 167,150 newly diagnosed breast cancer patients were included in the analysis. Of them, I excluded cases with male breast cancer (n=1,214), history of ductal carcinoma in situ (n=3,937), recurrent breast cancer (n=2,642), absence of previous diagnostic biopsy (n=8,495), absence of follow-up record (n=152), and absence of the record of the medical facility where the biopsy was performed (n=1) (Figure 1).

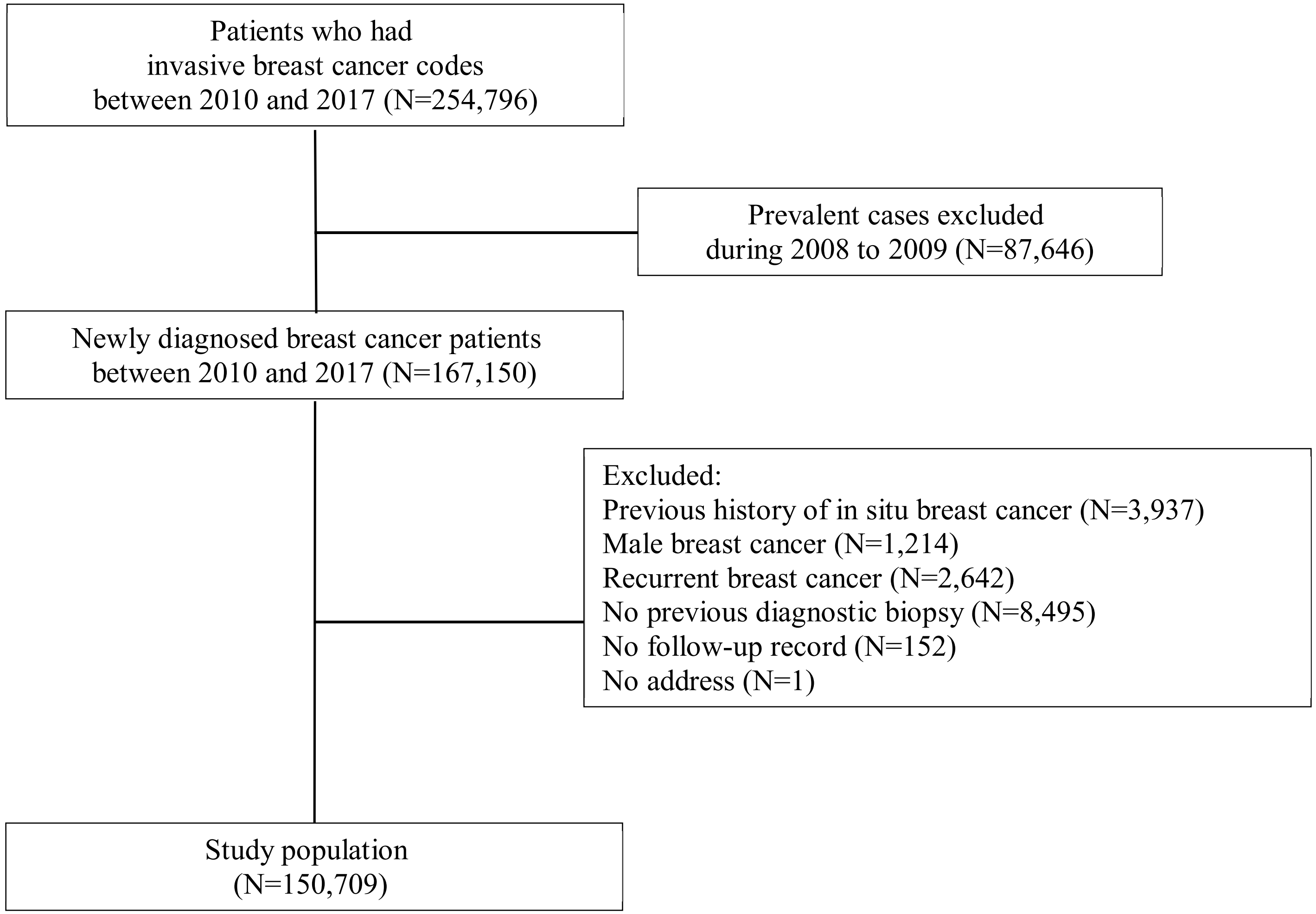

Figure 1. Study population

Variables and operational definitions

The basic demographics of the patients and their Charlson Comorbidity Index (CCI) using ICD-10 codes were analyzed (22, 23). Medical history such as hypertension (I10–13, 15), diabetes mellitus (E10–14), and dyslipidemia (E78) were defined using the relevant ICD-10 codes combined with medications.

Pathologic examination for breast cancer diagnosis was defined with the pathologic examination claims codes (C5509, C5911-C5917, C5500, C5504, C5508). The date of the pathologic examination was defined as the latest date among the dates when pathologic examination claims codes were recorded prior to breast cancer diagnosis. The date of breast cancer diagnosis was defined as the first date when C50 plus V193 codes were recorded. The V193 code is a unique code which is given after the definite diagnosis of cancer following biopsy, and it could more accurately identify cancer-diagnosed patients. Treatment of breast cancer such as surgery, chemotherapy, radiotherapy, endocrine therapy, and trastuzumab were defined based on the claims data recorded within 1 year after breast cancer diagnosis. The date of the surgery or the initiation of chemotherapy was collected.

Geographic regions were categorized into Seoul, metropolitan cities (Daegu, Daejeon, Gwangju, Incheon, Pusan, and Ulsan), and other regions (Chungbuk, Chungnam, Gangwon, Gyeongbuk, Gyeonggi, Gyeonggam, Jeju, Jeonbuk, Jeonnam, and Sejong). The patients with breast cancer were divided into three groups according to the regions in which they had received pathologic examinations. Regions where pathologic examination, diagnosis, and initiation of treatment were performed were also analyzed.

Statistical analysis

Baseline characteristics were presented according to the regions where the pathologic examinations were performed. To visualize the domestic medical travels of breast cancer survivors in Republic of Korea during the total study period, a Sankey diagram was used (24). The annual frequencies of pathologic examinations, diagnosis, initial treatment (surgery, chemotherapy, and radiotherapy) for breast cancer by geographic region were analyzed.

Generalize linear models with quasi-Poisson distribution adjusted for regions and years with regions other than Seoul and 2010 as reference categories were performed, and adjusted odds ratio (OR) and 95% confidence interval (CI) were estimated to investigate the differences in the frequencies of pathologic examination, diagnosis, initial treatment, surgery, chemotherapy, and radiotherapy. The numbers and proportions of patients who underwent pathologic examinations in non-Seoul areas traveling to Seoul for initial breast cancer treatment were analyzed. Factors associated with domestic medical travel from non-Seoul regions to Seoul were analyzed using a multivariable logistic regression model adjusted by age at diagnosis, insurance (health insurance vs. medical aid), CCI, region where the pathologic examination was performed (metropolitan cities vs. other regions), and year (from 2010 to 2017).

Statistical analyses were conducted using SAS statistical software (version 9.4.2, SAS Institute Inc., Cary, NC, USA) and R software version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org). This study was approved by the Institutional Review Board of Asan Medical Center (IRB no. 2020-1769).

Results

Baseline characteristics

A total of 150,709 patients diagnosed with breast cancer between January 2010 and December 2017 were included in this study. The mean age of the total population was 52.0 years (Table 1). Of the patients, 135,886 (90.2%) underwent surgery. The number of patients received radiotherapy and endocrine therapy were 100,064 (66.4%) and 100,102 (66.4%), respectively. The proportions of breast cancer patients who received chemotherapy and trastuzumab were 58.4% and 15.6%, respectively.

Annual frequencies of pathologic examinations, diagnosis, and treatment by regions

From 2010 to 2017, the annual frequencies of pathologic examinations, diagnosis, and treatment for breast cancer had increased in all regions (Figure 2–7). A greater number of patients underwent pathologic examinations, diagnosis, and initial treatment in Seoul than in non-Seoul regions (Table 2). The difference in the annual frequencies of pathologic examination and diagnosis between Seoul and other regions decreased over time (Figure 2 and 3). The difference in the annual frequencies of chemotherapy and radiotherapy between Seoul and non-Seoul regions did not increase over time (Figure 6 and 7). However, the difference in the annual frequencies of initial treatment and surgery between Seoul and non-Seoul regions increased over time (Figure 4 and 5).

Domestic medical travel for breast cancer treatment

During the total study period, 65,852 (43.7%) patients were diagnosed with breast cancer in Seoul. After the exclusion of 8,045 (5.3%) patients who did not receive any treatment,

Table 1. Characteristics of study population

Characteristics	Seoul (n = 60,542, %)		Metropolitan cities (n = 40,893, %)		Other regions (n = 49,274, %)		Total (n = 150,709, %)	
Age at diagnosis (years, mean ± SD)	51.6 ± 11.3		52.4 ± 11.2		52.3 ± 11.7		52.0 ± 11.4	
Insurance								
Health insurance	59,228	97.8	39,105	95.6	47,391	96.2	145,724	96.7
Medical aid	1,314	2.2	1,788	4.37	1,883	3.8	4,985	3.3
CCI (mean ± SD)	2.2 ± 2.1		2.3 ± 2.1		2.3 ± 2.1		2.3 ± 2.1	
previous diabetes mellitus	5,343	8.8	4,070	10.0	5,218	10.6	14,631	9.7
previous hypertension	16,042	26.5	11,734	28.7	14,728	29.9	42,504	28.2
previous dyslipidemia	15,931	26.3	11,593	28.4	13,771	28.0	41,295	27.4
Surgery*	54,083	89.3	37,407	91.5	44,396	90.1	135,886	90.2
Chemotherapy*	33,984	56.1	24,491	59.9	29,466	59.8	87,941	58.4
Radiotherapy*	39,799	65.7	26,667	65.2	33,598	68.2	100,064	66.4
Endocrine therapy*	39,735	65.6	27,016	66.1	33,351	67.7	100,102	66.4
Trastuzumab*	8,756	14.5	7,000	17.1	7,786	15.8	23,542	15.6

Duration after cohort entry (months, median, IQR)	51 (29-77)	49 (28-74)	46 (26-71)	49 (28-74)

CCI, Charlson Comorbidity Index; SD, standard deviation; IQR, Interquartile range

* Treatment of breast cancer such as surgery, chemotherapy, radiotherapy, endocrine therapy, and trastuzumab were defined based on the claims data recorded within 1 year after breast cancer diagnosis.

Table 2. Frequencies of breast cancer management by region using quasi-Poisson regression analysis

	OR[1]	95% CI		P value
		lower	upper	
Pathologic examination				
Seoul	1.230	1.143	1.324	<0.001
Metropolitan city	0.830	0.766	0.901	<0.001
Diagnosis				
Seoul	1.419	1.350	1.492	<0.001
Metropolitan city	0.829	0.783	0.878	<0.001
Initial treatment*				
Seoul	1.870	1.791	1.952	<0.001
Metropolitan city	0.844	0.802	0.889	<0.001
Surgery				
Seoul	1.902	1.823	1.985	<0.001
Metropolitan city	0.861	0.818	0.906	<0.001
Chemotherapy				
Seoul	1.781	1.702	1.865	<0.001
Metropolitan city	0.847	0.802	0.894	<0.001
Radiotherapy				
Seoul	1.162	1.081	1.249	0.001
Metropolitan city	0.687	0.633	0.747	<0.001

[1]adjusted by years (from 2010 to 2017) with 2010 as a reference and regions (Seoul, metropolitan cities, and others) with others as a reference

*Any kinds of treatment (surgery, chemotherapy, and radiotherapy)

OR, odds ratio; CI, confidence interval

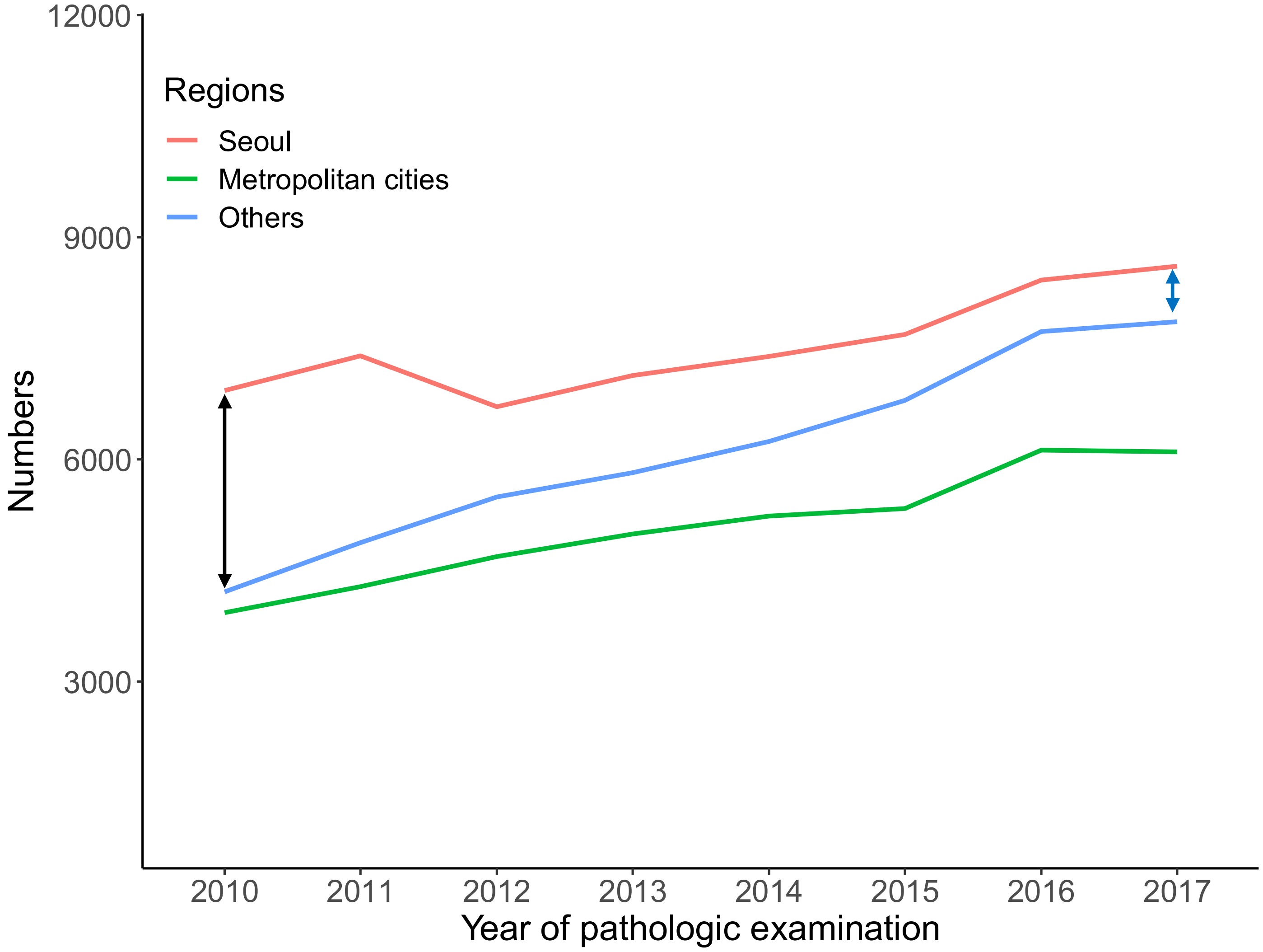

Figure 2. Annual number of patients who got pathologic examination by regions

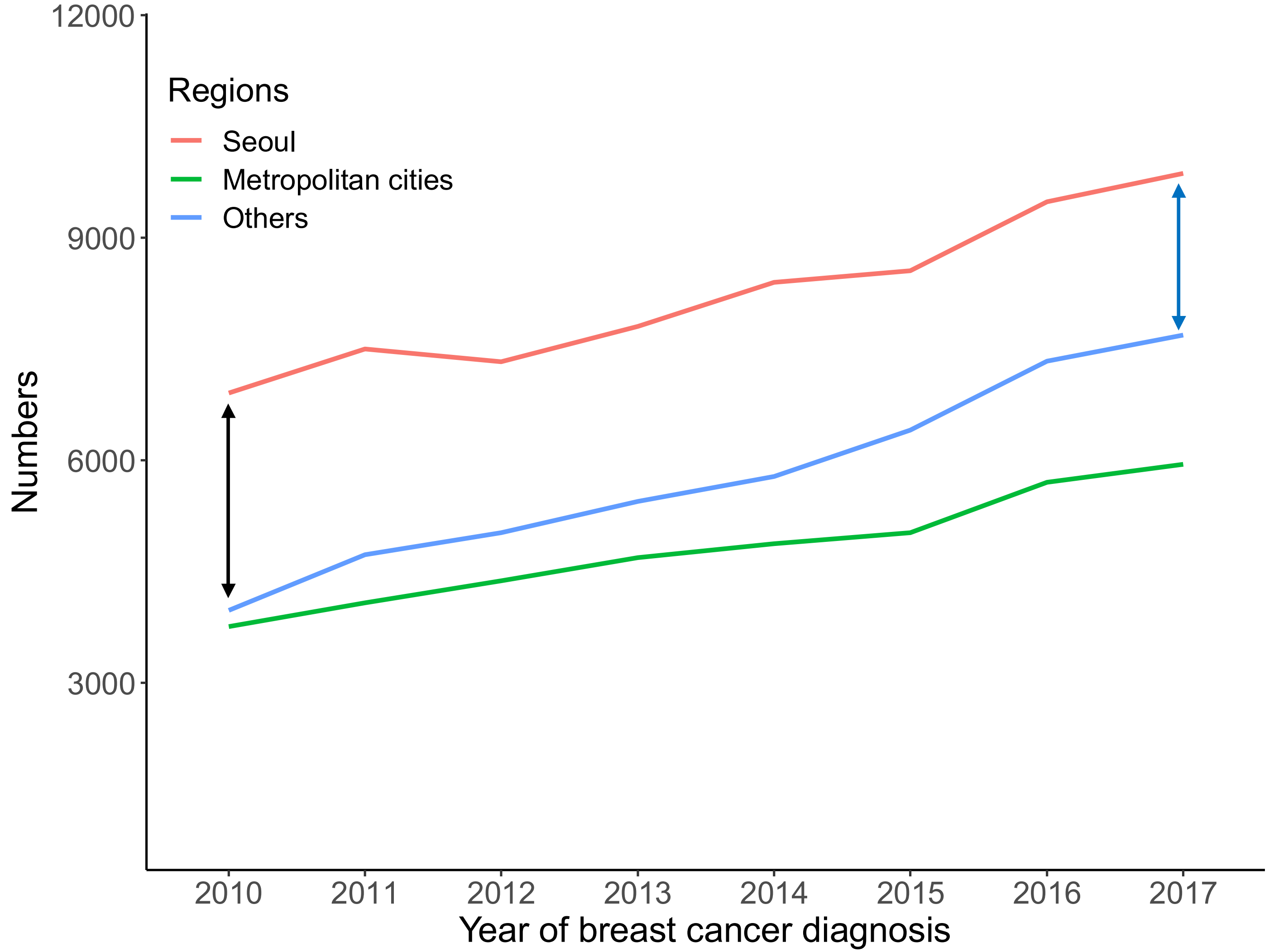

Figure 3. Annual number of patients who diagnosed with breast cancer by regions

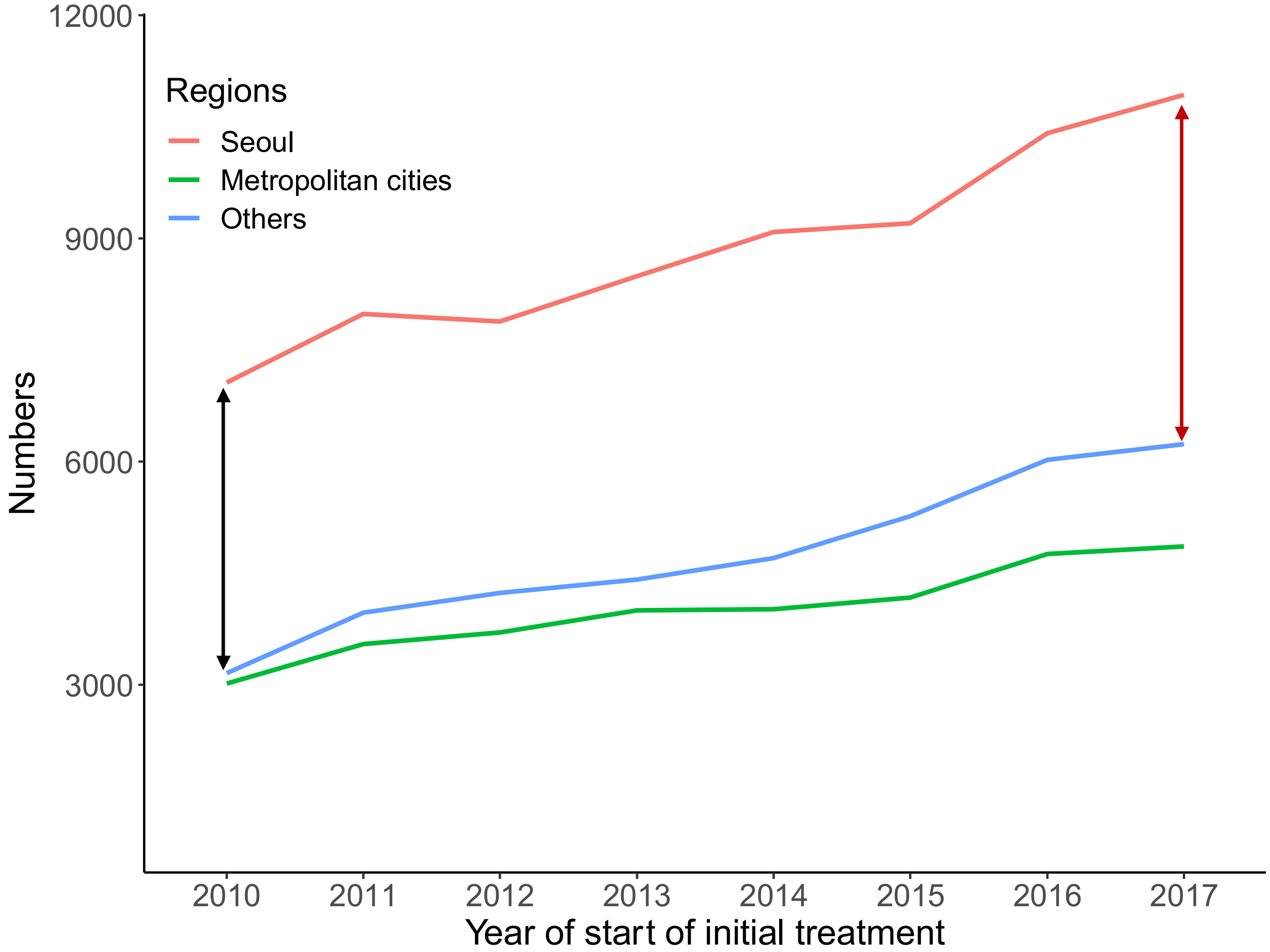

Figure 4. Annual number of patients who got initial treatment of breast cancer by regions

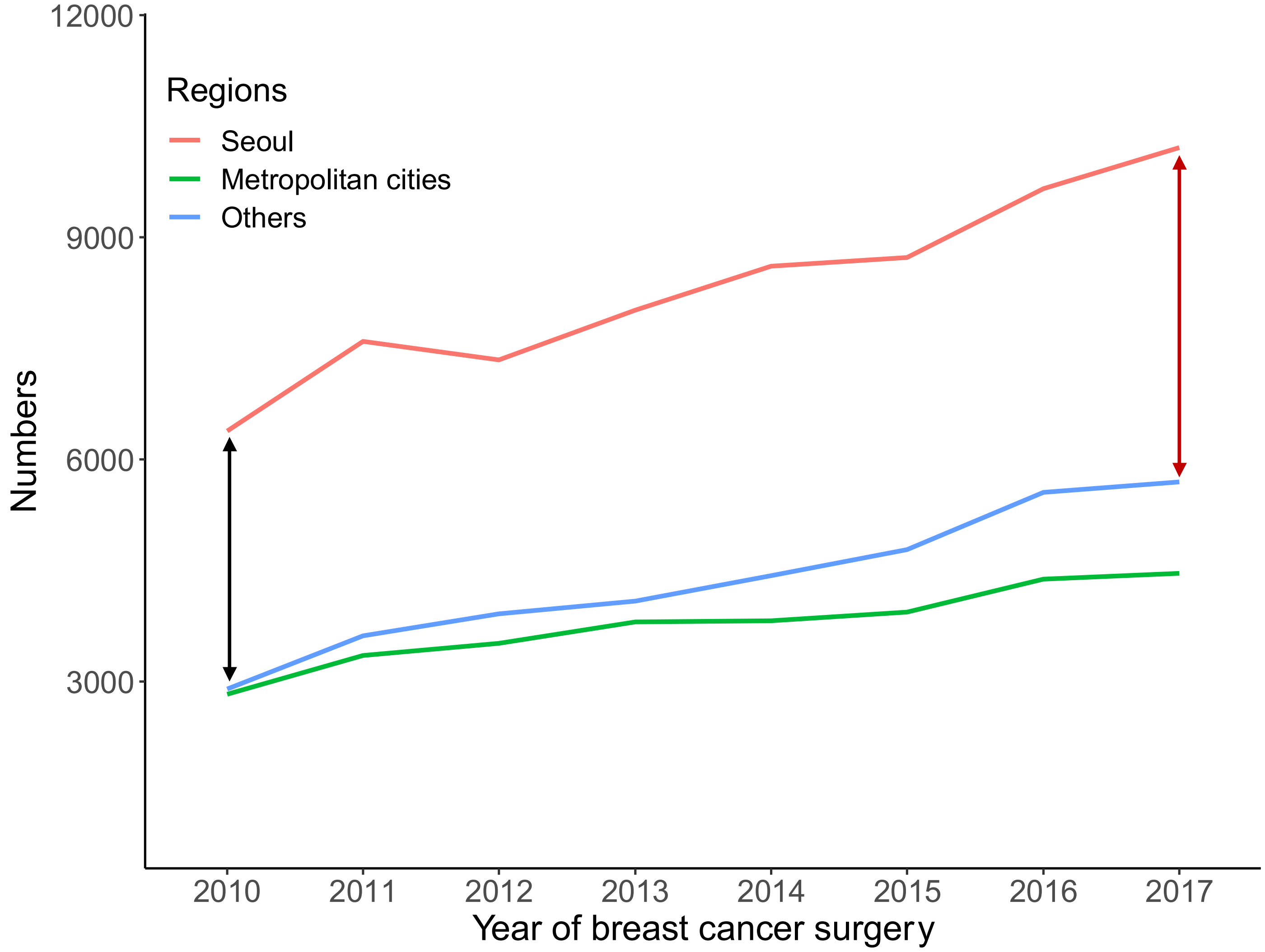

Figure 5. Annual number of patients who underwent breast cancer surgery by regions

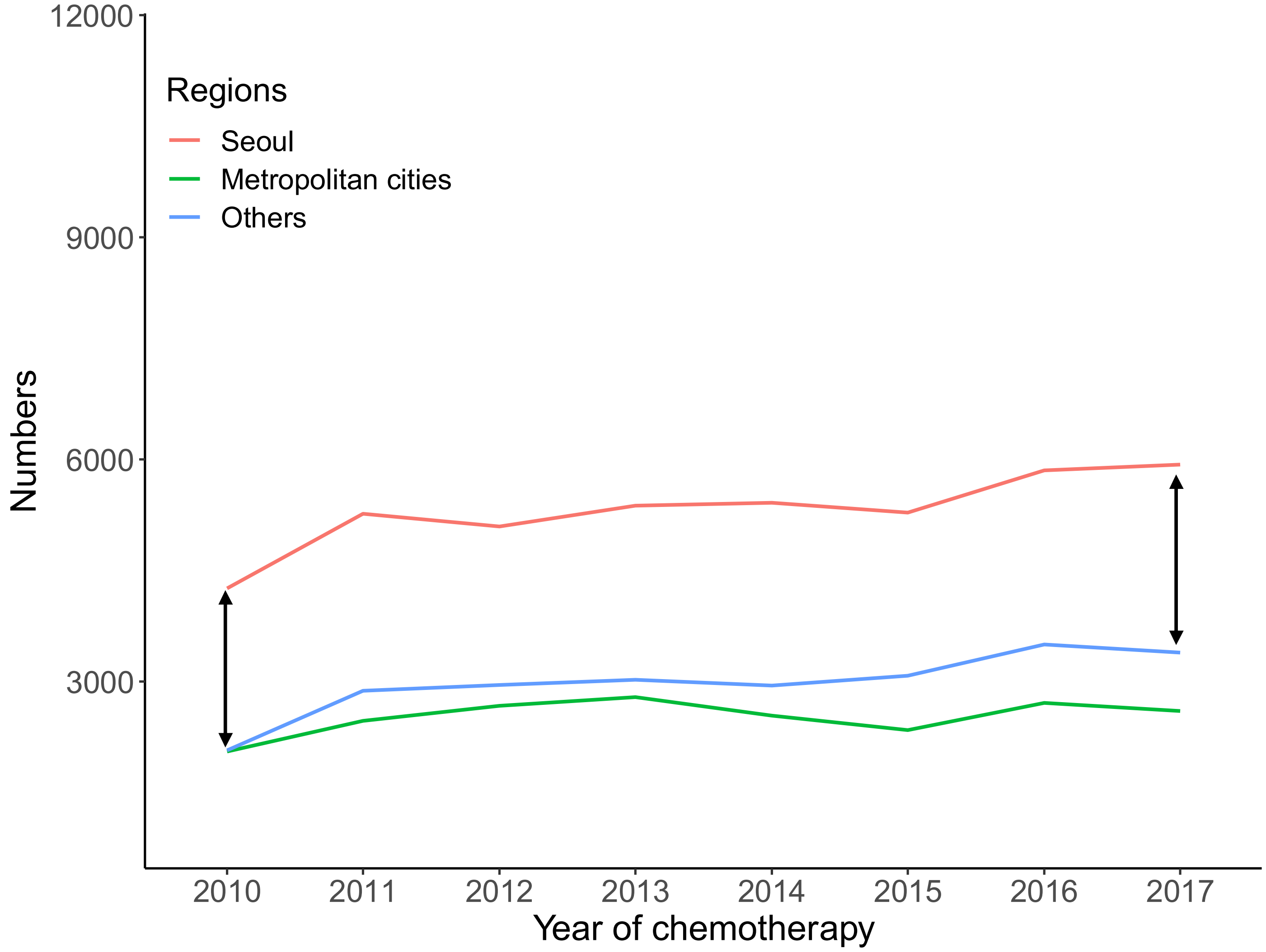

Figure 6. Annual number of patients who treated with chemotherapy of breast cancer by regions

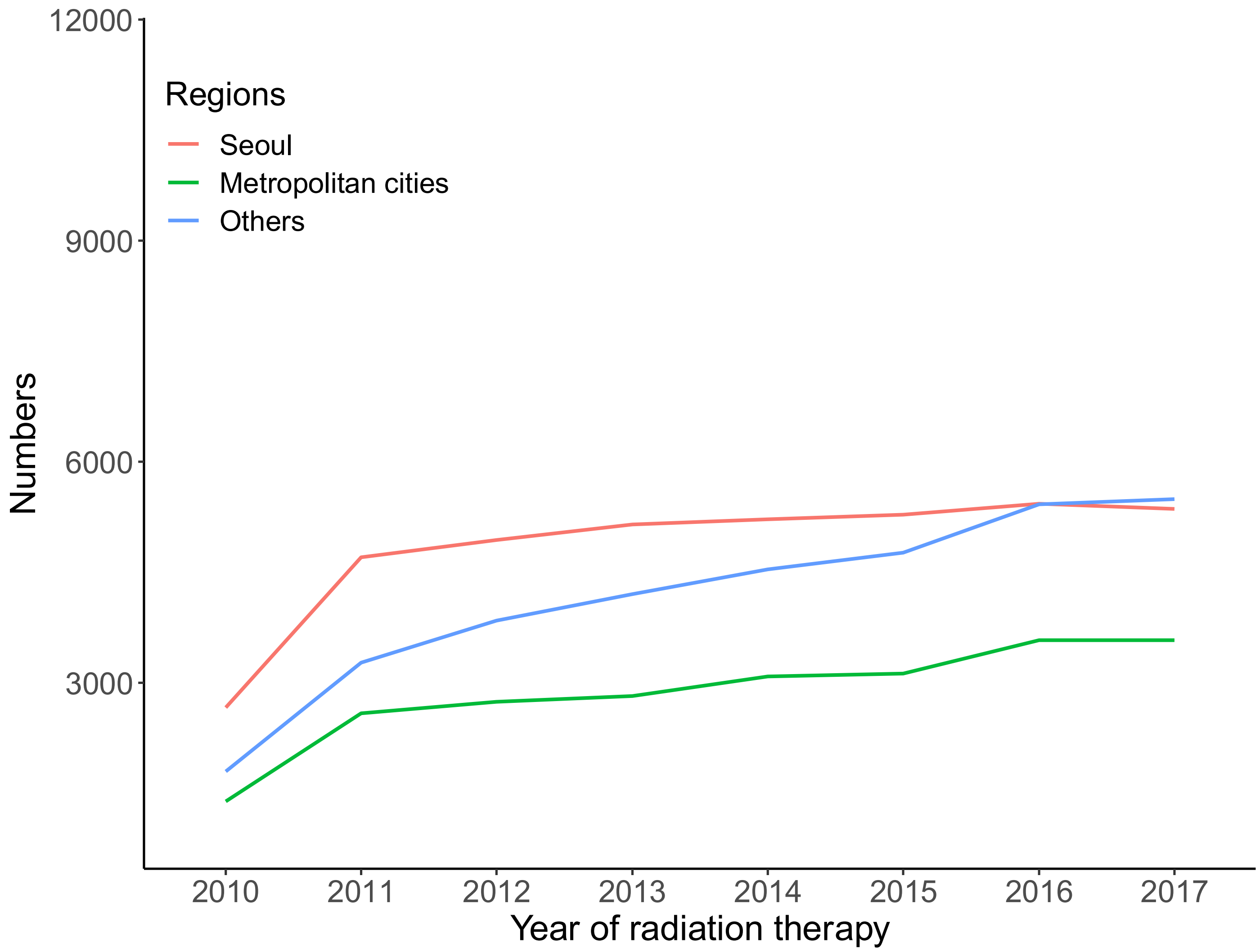

Figure 7. Annual number of patients who got radiotherapy of breast cancer by regions

50.4% (71,925/142,664) of patients started breast cancer treatment in Seoul (Table 3 and Figure 8).

Among the patients who underwent pathologic examination in non-Seoul regions, after excluding 434 patients without a record on the date of pathologic examinations, the rate of domestic medical travel from non-Seoul regions to Seoul for initial treatment increased from 14.2% (1,161/8,150) in 2010 to 19.8% (2,762/13,964) in 2017 (Table 4). Approximately a quarter of patients (23.6%) from other regions traveled to Seoul in 2017. Especially, more than 40% of breast cancer patients who underwent pathologic examination in Chungbuk, Gyeongbuk, and Jeju regions had traveled to Seoul in 2017. (Figure 9 and Figure 10).

Factors associated with domestic medical travel for breast cancer treatment

After excluding 4,171 patients who did not receive any treatment, multivariable logistic regression analyses showed age younger than 40 years (OR, 2.13; 95% CI, 2.00–2.28; $p < 0.001$) was significantly associated with a higher likelihood for domestic medical travel to Seoul for breast cancer treatment (Table 5). Patients registered with medical aid (OR, 0.26; 95% CI, 0.23–0.31; $p < 0.001$) and those with other medical diseases (OR, 0.98; 95% CI, 0.97–0.99; $p < 0.001$) were significantly less likely to travel to Seoul. Patients who underwent pathologic examination in other regions were significantly more likely to travel to Seoul than those who underwent an examination in metropolitan cities (OR, 1.77; 95% CI, 1.71–1.84; $p < 0.001$).

Table 3. Domestic medical travel after diagnosis

Region	No. of patients diagnosed (%)	No. of patients treated[§] (%)	From diagnosis to treatment	No. of patients (%)
Seoul	65,852 (43.7)	71,925 (50.4)	Seoul → Seoul	61,471 (93.3)
			Seoul → Other regions	887 (1.3)
			Seoul → Metropolitan cities	136 (0.2)
			Seoul → no treatment	3,358 (5.1)
Metropolitan cities	38,464 (25.5)	32,376 (22.7)	Metropolitan cities → Metropolitan cities	31,012 (80.6)
			Metropolitan cities → Seoul	3,642 (9.5)
			Metropolitan cities → Other regions	1,891 (4.9)
			Metropolitan cities → no treatment	1,919 (5.0)
Other regions	46,393 (30.8)	38,363 (26.9)	Other regions → Other regions	35,585 (76.7)
			Other regions → Seoul	6,812 (14.7)
			Other regions → Metropolitan cities	1,228 (2.6)
			Other regions → no treatment	2,768 (6.0)

§ Calculated only for patients who received any kinds of initial treatment (surgery, chemotherapy, and radiotherapy)

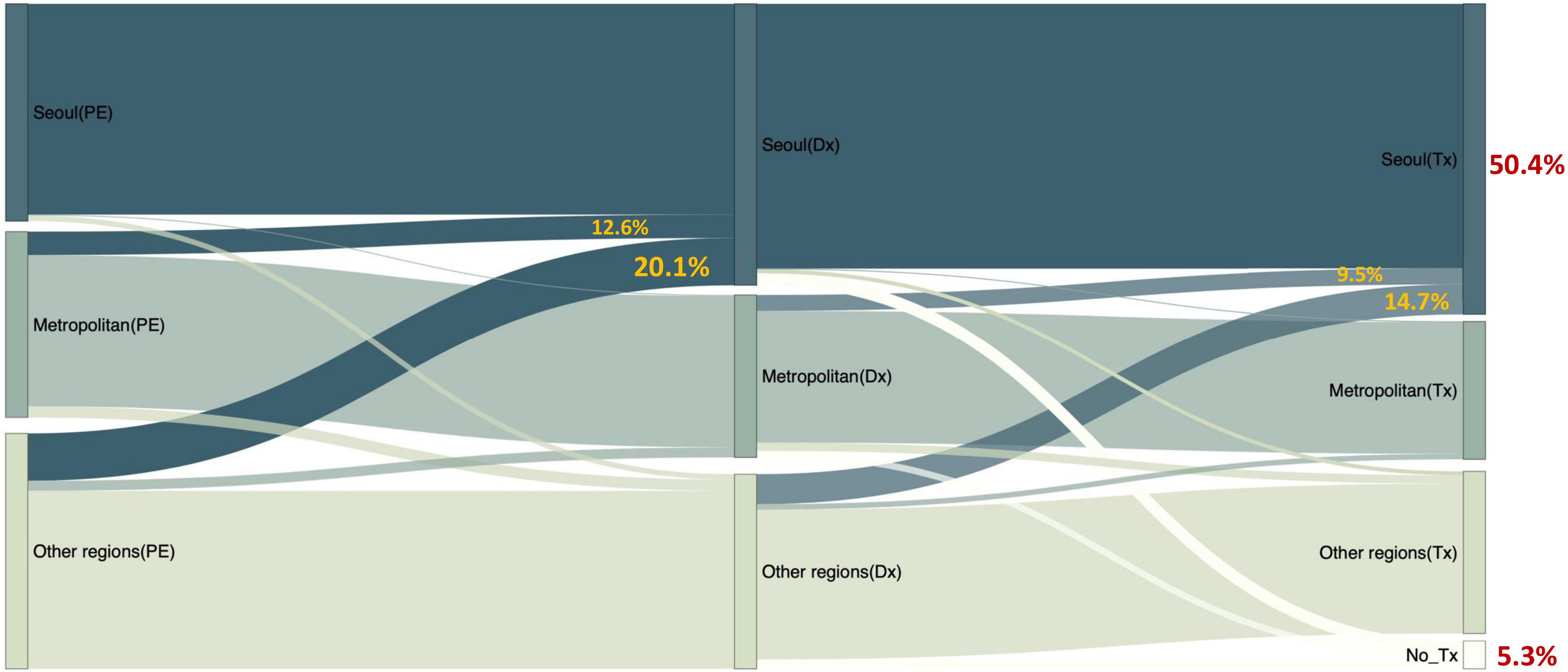

Figure 8. Sankey diagram of domestic medical travel

PE, pathologic examination, Dx, diagnosis, Tx, treatment

Table 4. Domestic medical travel from non-Seoul regions to Seoul for initial treatment (2010 – 2014, continued)

	2010			2011			2012			2013			2014		
	n[a]	n[b]	%	n[a]	n[b]	%	n[a]	n[b]	%	n[a]	n[b]	%	n[a]	n[b]	%
Metropolitan city	3,935	443	11.3	4,286	432	10.1	4,692	624	13.3	4992	696	13.9	5,232	803	15.3
Other regions	4,215	718	17.0	4,873	796	16.3	5,495	1,190	21.7	5818	1,326	22.8	6,240	1,529	24.5
total	8,150	1,161	14.2	9,159	1,228	13.4	10,187	1,814	17.8	10810	2,022	18.7	11,472	2,332	20.3
Chungbuk	183	57	31.1	197	61	31.0	269	99	36.8	247	113	45.7	289	126	43.6
Chungnam	215	53	24.7	248	44	17.7	281	62	22.1	298	71	23.8	297	87	29.3
Daegu	902	63	7.0	970	70	7.2	1,027	98	9.5	1,139	116	10.2	1,220	123	10.1
Daejeon	441	80	18.1	469	74	15.8	568	150	26.4	520	133	25.6	617	181	29.3
Gangwon	219	60	27.4	236	57	24.2	292	100	34.2	325	102	31.4	341	117	34.3
Gwangju	323	48	14.9	380	39	10.3	440	75	17.0	522	93	17.8	485	91	18.8
Gyeongbuk	166	49	29.5	218	72	33.0	284	97	34.2	288	99	34.4	307	132	43.0
Gyeonggi	2,358	326	13.8	2,793	381	13.6	3,105	568	18.3	3,302	613	18.6	3,614	716	19.8
Gyeongnam	307	48	15.6	367	55	15.0	377	82	21.8	451	116	25.7	470	123	26.2
Incheon	645	71	11.0	684	72	10.5	783	106	13.5	837	131	15.7	849	132	15.5

Jeju	68	24	35.3	89	27	30.3	117	34	29.1	127	44	34.6	132	45	34.1
Jeonbuk	367	55	15.0	415	57	13.7	436	91	20.9	442	96	21.7	447	115	25.7
Jeonnam	332	46	13.9	310	42	13.5	334	57	17.1	336	71	21.1	343	68	19.8
Pusan	1,442	160	11.1	1,572	147	9.4	1,642	157	9.6	1,713	196	11.4	1,712	224	13.1
Sejong	NA	NA		NA	NA		NA	NA		2	1	50.0	NA	NA	
Ulsan	182	21	11.5	211	30	14.2	232	38	16.4	261	27	10.3	349	52	14.9

n^a: number of patients who underwent pathologic examination, n^b: number of patients who got initial treatment in Seoul

NA, not available

Table 4. The domestic medical travel from non-Seoul regions to Seoul for initial treatment (2015 – 2017)

	2015			2016			2017			Total		
	n[a]	n[b]	%	n[a]	n[b]	%	n[a]	n[b]	%	n[a]	n[b]	%
Metropolitan city	5,339	735	13.8	6,128	874	14.3	6,105	906	14.8	40,709	5,513	13.5
Other regions	6,797	1,468	21.6	7,727	1,758	22.8	7,859	1,856	23.6	49,024	10,641	21.7
total	12,136	2,203	18.2	13,855	2,632	19.0	13,964	2,762	19.8	89,733	16,154	18.0
Chungbuk	350	146	41.7	352	141	40.1	297	122	41.1	2,184	865	39.6
Chungnam	381	96	25.2	409	107	26.2	386	98	25.4	2,515	618	24.6
Daegu	1257	118	9.4	1,339	136	10.2	1,362	159	11.7	9,216	883	9.6
Daejeon	615	163	26.5	741	187	25.2	673	167	24.8	4,644	1,135	24.4
Gangwon	369	108	29.3	435	134	30.8	444	130	29.3	2,661	808	30.4
Gwangju	497	72	14.5	604	110	18.2	655	127	19.4	3,906	655	16.8
Gyeongbuk	340	148	43.5	395	168	42.5	406	178	43.8	2,404	943	39.2
Gyeonggi	3,758	612	16.3	4,347	778	17.9	4,506	889	19.7	27,783	4,883	17.6
Gyeongnam	570	133	23.3	634	158	24.9	687	167	24.3	3,863	882	22.8
Incheon	952	141	14.8	1148	150	13.1	1,113	138	12.4	7,011	941	13.4

Jeju	158	53	33.5	156	56	35.9	190	80	42.1	1,037	363	35.0
Jeonbuk	502	105	20.9	596	134	22.5	561	110	19.6	3,766	763	20.3
Jeonnam	368	66	17.9	401	82	20.4	381	82	21.5	2,805	514	18.3
Pusan	1,724	201	11.7	1,904	237	12.4	1,884	242	12.8	13,593	1,564	11.5
Sejong	1	1	100.0	2	0	0.0	1	0	0.0	6	2	33.3
Ulsan	294	40	13.6	392	54	13.8	418	73	17.5	2,339	335	14.3

n^a: number of patients who underwent pathologic examination, n^b: number of patients who got initial treatment in Seoul

NA, not available

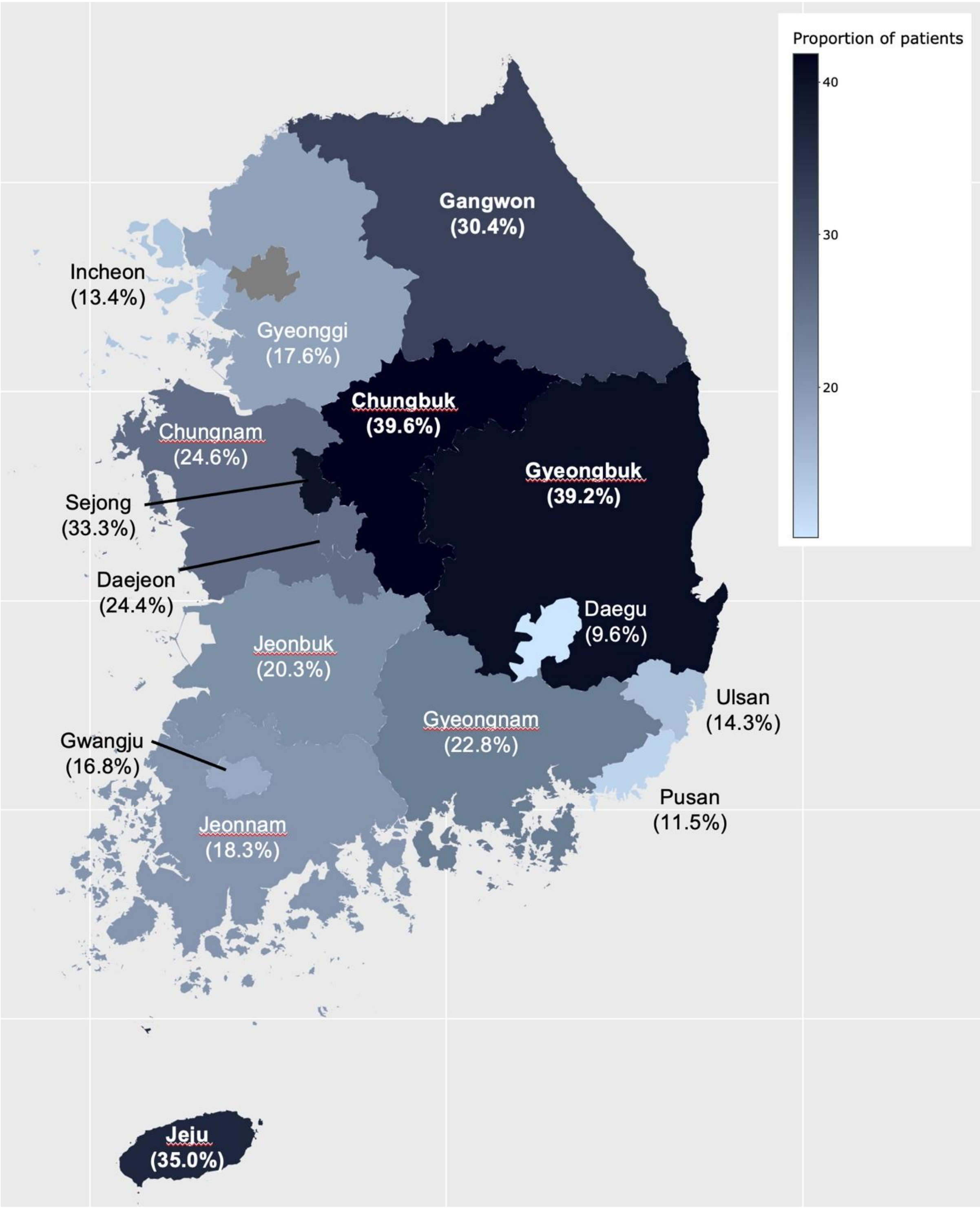

Figure 9. Regional distribution of medical travel to Seoul

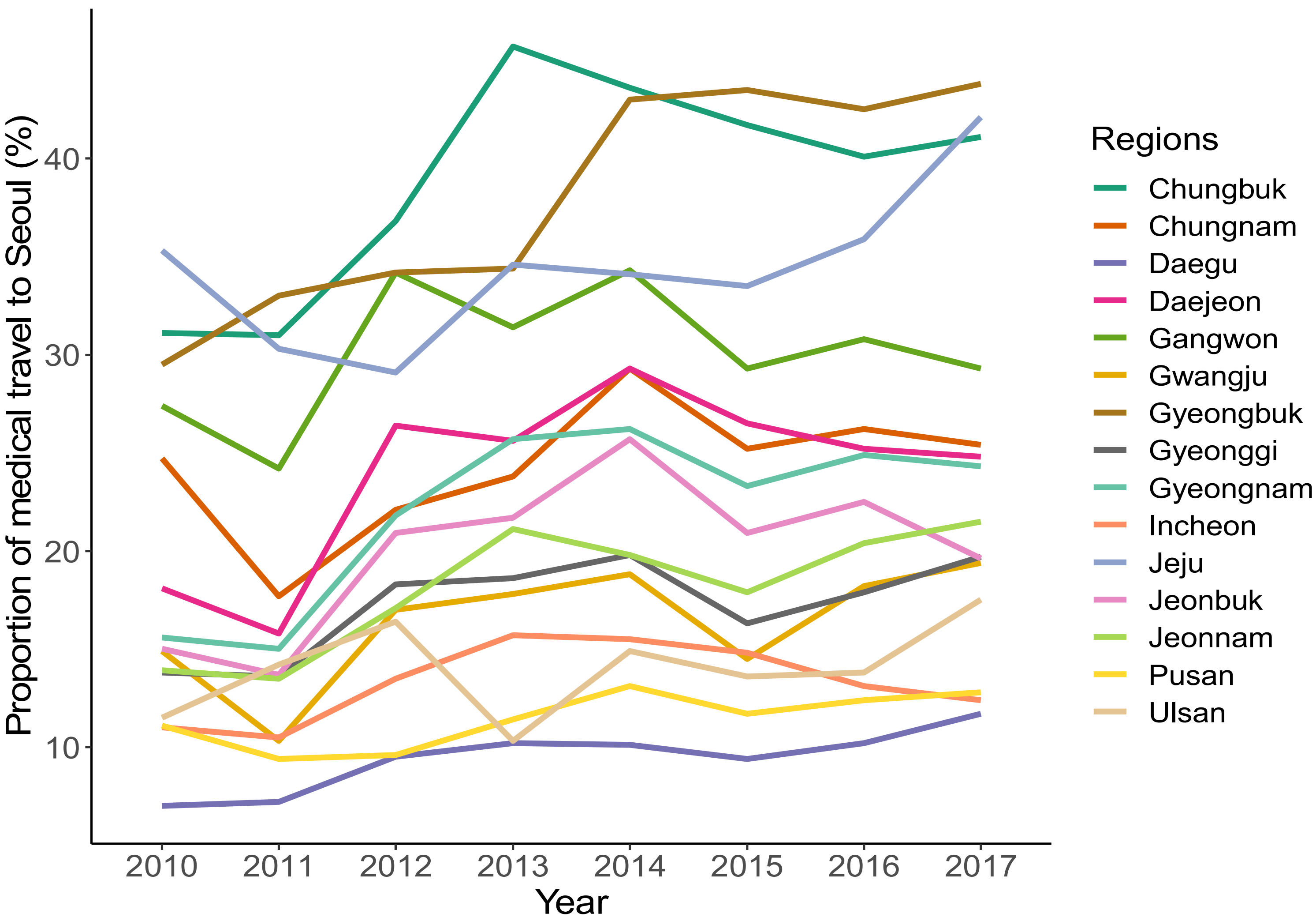

Figure 10. Annual proportion of medical travel to Seoul by region

Table 5. Factors associated with medical travel from non-Seoul to Seoul using multivariable logistic regression analysis

	Cases No.	Events No.	Proportion %	OR[1]	95% CI lower	95% CI upper	P value
Age at diagnosis							
60≤	20,021	2,548	12.7	reference			
50-59	25,980	4,872	18.8	1.51	1.43	1.59	<0.001
40-49	30,310	6,411	21.2	1.72	1.63	1.81	<0.001
<40	9,251	2,323	25.1	2.13	2.00	2.28	<0.001
Insurance type							
Health insurance	82,243	15,986	19.4	reference			
Medical aid	3,319	168	5.1	0.27	0.23	0.31	<0.001
CCI (continuous)	NA	NA	NA	0.98	0.97	0.99	<0.001
Regions							
Metropolitan cities	38,866	5,513	14.2	reference			
Others	46,696	10,641	22.8	1.77	1.71	1.84	<0.001
Years							
2010	7,659	1,161	15.2	reference			
2011	8,753	1,228	14.0	0.91	0.83	0.99	<0.001
2012	9,746	1,814	18.6	1.29	1.19	1.40	<0.001
2013	10,335	2,022	19.6	1.38	1.27	1.49	<0.001
2014	10,927	2,332	21.3	1.55	1.44	1.68	<0.001
2015	11,539	2,203	19.1	1.34	1.24	1.45	<0.001
2016	13,222	2,632	19.9	1.43	1.32	1.54	<0.001
2017	13,381	2,762	20.6	1.50	1.39	1.62	<0.001

OR, odds ratio; CI, confidence interval, CCI, Charlson Comorbidity Index

Discussion

This study showed that domestic medical travel for breast cancer treatment from non-Seoul regions to Seoul increased over time in Republic of Korea. The numbers and proportions of domestic medical travel from non-Seoul regions to Seoul differed by region, and in some regions, more than 40% of the patients traveled to Seoul for breast cancer treatment in 2017. Young patients and patients in other regions were more likely to travel to Seoul after pathologic examination. Patients who were registered with medical aid or those with other medical diseases were less likely to travel to Seoul. To my knowledge, this is the first nationwide study to investigate the trend in domestic medical travel of breast cancer patients in Republic of Korea.

A previous study showed the centralization of lung cancer surgery in Republic of Korea (15), as the proportion of lung cancer surgeries significantly increased over time in Seoul while the proportion in other areas decreased. For prostate cancer, one study showed that the frequencies of medical travel from non-Seoul regions to Seoul for treatment decreased from 2005 to 2014, although a large proportion of medical travel remained unchanged (17). Another study in the US (New York State) was conducted to investigate the impact of centralizing rectal cancer surgeries to high-volume centers (14), which showed that there was a statistically significant increase in the average distance traveled by patients during a 10-year period from 2004 to 2014, indicating that centralization might add to the travel burden of patients.

Centralization of cancer treatment is an important issue for patients and healthcare providers alike because there is a risk of prolonged waiting time for treatment. Delay in the start of treatment often causes anxiety, reduces the quality of life, adversely affects mental health, and may lead to detrimental effects on survival (25). In breast cancer, prolonged time from diagnosis to surgery significantly decreases cancer-specific survival as well as overall survival (9, 10). Delayed chemotherapy was also associated with poor prognosis, both in the adjuvant setting (26, 27) and in the metastatic setting (28).

The present study describes the patterns of medical travel of more than 150,000 patients diagnosed with breast cancer in different regions of Republic of Korea during a 7-year period. I found that young age, insurance type, and previous medical disease were the most important factors associated with medical travel from outside Seoul to Seoul. Medical travel to Seoul for treatment was more common in younger age groups, and the number of patients in their 30s was more than double that of patients aged 60 or older. A similar trend was also observed in other types of cancers, such as prostate cancer or head and neck cancer (17, 29). In this study, the OR for medical travel in patients on medical aid was one-quarter of that in patients with health insurance. Similar to the results of this study, one study conducted on parathyroid surgery in the US reported that the proportion of domestic medical travelers (defined as patients who underwent parathyroidectomy at a hospital in a different US region from the one in which they resided and traveled more than 150 miles to that hospital) were higher in patients with insurance (62.1%) than in those with Medicare (32.4%) or Medicaid/uninsured (0.3%) (30).

An interesting finding of the present study is that the proportion of patients receiving initial treatment in Seoul increased over time, but this differed depending on the treatment type (i.e., surgery, chemotherapy, and radiotherapy). The proportion of patients receiving treatment in Seoul increased sharply over time in the case of surgery, while only a modest increase was noted for chemotherapy; in contrast, the proportion of patients receiving radiotherapy increased rapidly in regions other than Seoul. A possible explanation for this discrepancy according to treatment type is the frequency of treatment; while surgery is a one-time treatment, chemotherapy requires visits every one-to-three weeks for more than three months; moreover, radiotherapy requires daily visits, which makes it less likely for patients to travel outside their residential region for treatment.

The limitations of this study should be noted. First, the actual address of the patients' residence could not be identified because the HIRA data is based on claims data. Thus, I categorized patients based on the regions where the pathologic examinations were performed instead of the patients' residences. Second, there are several potential factors associated with medical travel that were not included in the HIRA data, such as socioeconomic status, transportations, and route of information. For a more detailed investigation for the reason of medical travel, a subsequent questionnaire study will be conducted. Third, although the present study showed the centralization of breast cancer treatment in Republic of Korea, the waiting time for breast cancer treatment was not investigated. Further studies are needed to assess whether centralization leads to delays in treatment and causes harmful effects on breast cancer survival. In addition, the HIRA data did not include the stage of cancer, survival analysis could not be performed accordingly. Lastly, it was not clear in this study whether the reason patients moved to Seoul for treatment was because there were therapeutic benefits such as standard level of treatment. In countries like Korea, where all insurance companies are government, and countries where transportation costs are more problematic the cost of medical care, the national policies are inevitably important, and consensus is required as a policy.

In conclusion, the numbers and proportion of patients diagnosed with breast cancer who traveled to Seoul for treatment increased over time depending on the treatment method and regions. Young age, comorbidities, and type of insurance were significant factors associated with medical travel. Further investigation about the effect of centralization on waiting time for breast cancer treatment and survival should be performed.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65(2):87-108.
3. National Cancer Information Center. https://www.cancer.go.kr.
4. Kang SY, Kim YS, Kim Z, Kim HY, Kim HJ, Park S, et al. Breast Cancer Statistics in Korea in 2017: Data from a Breast Cancer Registry. J Breast Cancer. 2020;23(2):115-28.
5. Chung IY, Lee J, Park S, Lee JW, Youn HJ, Hong JH, et al. Nationwide Analysis of Treatment Patterns for Korean Breast Cancer Survivors Using National Health Insurance Service Data. J Korean Med Sci. 2018;33(44):e276.
6. Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger. Ann Oncol. 2019;30(8):1194-220.
7. NCCN clinical practice guidelines in oncology: breast cancer, version 8.2021. https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf.
8. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. Lancet. 1999;353(9159):1119-26.
9. Yun YH, Kim YA, Min YH, Park S, Won YJ, Kim DY, et al. The influence of hospital volume and surgical treatment delay on long-term survival after cancer surgery. Ann Oncol. 2012;23(10):2731-7.
10. Bleicher RJ, Ruth K, Sigurdson ER, Beck JR, Ross E, Wong YN, et al. Time to Surgery and Breast Cancer Survival in the United States. JAMA Oncol. 2016;2(3):330-9.

11. Maghous A, Rais F, Ahid S, Benhmidou N, Bellahamou K, Loughlimi H, et al. Factors influencing diagnosis delay of advanced breast cancer in Moroccan women. BMC Cancer. 2016;16:356.

12. Stitzenberg KB, Sigurdson ER, Egleston BL, Starkey RB, Meropol NJ. Centralization of cancer surgery: implications for patient access to optimal care. J Clin Oncol. 2009;27(28):4671-8.

13. Shalowitz DI, Nivasch E, Burger RA, Schapira MM. Are patients willing to travel for better ovarian cancer care? Gynecol Oncol. 2018;148(1):42-8.

14. Xu Z, Aquina CT, Justiniano CF, Becerra AZ, Boscoe FP, Schymura MJ, et al. Centralizing Rectal Cancer Surgery: What Is the Impact of Travel on Patients? Dis Colon Rectum. 2020;63(3):319-25.

15. Park S, Park IK, Kim ER, Hwang Y, Lee HJ, Kang CH, et al. Current Trends of Lung Cancer Surgery and Demographic and Social Factors Related to Changes in the Trends of Lung Cancer Surgery: An Analysis of the National Database from 2010 to 2014. Cancer Res Treat. 2017;49(2):330-7.

16. Hong NS, Lee KS, Kam S, Choi GS, Kwon OK, Ryu DH, et al. A Survival Analysis of Gastric or Colorectal Cancer Patients Treated With Surgery: Comparison of Capital and a Non-capital City. J Prev Med Public Health. 2017;50(5):283-93.

17. Kim JH, Kim SY, Yun SJ, Chung JI, Choi H, Yu HS, et al. Medical Travel among Non-Seoul Residents to Seek Prostate Cancer Treatment in Medical Facilities of Seoul. Cancer Res Treat. 2019;51(1):53-64.

18. 2021 National Health Insurance Act. https://www.nhis.or.kr/.

19. Kim JA, Yoon S, Kim LY, Kim DS. Towards Actualizing the Value Potential of Korea Health Insurance Review and Assessment (HIRA) Data as a Resource for Health Research: Strengths, Limitations, Applications, and Strategies for Optimal Use of HIRA Data. J Korean

Med Sci. 2017;32(5):718-28.

20. Kim L, Kim JA, Kim S. A guide for the utilization of Health Insurance Review and Assessment Service National Patient Samples. Epidemiol Health. 2014;36:e2014008.

21. World Health Organization. ICD-10 version: 2016 https://icd.who.int/browse10/2016/en.

22. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43(11):1130-9.

23. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol. 2004;57(12):1288-94.

24. Schmidt M. The Sankey diagram in energy and material flow management: Part II: Methodology and current applications. Journal of industrial ecology. 2008;12(2):173-85.

25. Song H, Fang F, Valdimarsdottir U, Lu D, Andersson TM, Hultman C, et al. Waiting time for cancer treatment and mental health among patients with newly diagnosed esophageal or gastric cancer: a nationwide cohort study. BMC Cancer. 2017;17(1):2.

26. Lohrisch C, Paltiel C, Gelmon K, Speers C, Taylor S, Barnett J, et al. Impact on survival of time from definitive surgery to initiation of adjuvant chemotherapy for early-stage breast cancer. J Clin Oncol. 2006;24(30):4888-94.

27. Chavez-MacGregor M, Clarke CA, Lichtensztajn DY, Giordano SH. Delayed Initiation of Adjuvant Chemotherapy Among Patients With Breast Cancer. JAMA Oncol. 2016;2(3):322-9.

28. Ho PJ, Cook AR, Binte Mohamed Ri NK, Liu J, Li J, Hartman M. Impact of delayed treatment in women diagnosed with breast cancer: A population-based study. Cancer Med. 2020;9(7):2435-44.

29. Lamont EB, Hayreh D, Pickett KE, Dignam JJ, List MA, Stenson KM, et al. Is patient travel distance associated with survival on phase II clinical trials in oncology? J Natl Cancer Inst. 2003;95(18):1370-5.

30. Hinson AM, Hohmann SF, Stack BC, Jr. Domestic Travel and Regional Migration for Parathyroid Surgery Among Patients Receiving Care at Academic Medical Centers in the United States, 2012-2014. JAMA Otolaryngol Head Neck Surg. 2016;142(7):641-7.

Appendices

1. IRB documents

Date: 2021/04/28

주소 : 05505 서울특별시 송파구 올림픽로 43길 88 서울아산병원 TEL : 02-3010-7166, FAX : 02-3010-4163

심의면제 확인서

과제명	건강보험심사평가원 자료를 활용한 유방암 진단 및 치료의 지역 및 의료종별 이용 및 의료접근성 연구 (면제확인번호:2020-1769)					
연구책임자	소속	종양내과	직위	임상조교수	성명	정재호

1. 생명윤리 및 안전에 관한 법률에 의거한 인간대상연구 (기본사항이 모두 '예'이고, 아래 항목 중 하나 이상에 해당하는 경우 심의면제가 가능함)	예	아니오
기본사항		
1) 연구대상자 및 공공에 미치는 위험이 미미한 경우임	☑	☐
2) 일반 대중에게 공개된 정보를 이용하는 연구 또는 개인식별정보를 수집·기록하지 않은 연구	☑	☐
3) 취약한 환경에 있는 연구대상자를 대상으로 하는 연구가 아님	☑	☐
아래		
1) 연구대상자를 직접 조작하거나 그 환경을 조작하는 연구 중 다음 어느 하나에 해당하는 연구 (1) 약물투여, 혈액채취 등 침습적 행위를 하지 않는 연구 (2) 신체적 변화가 따르지 않는 단순 접촉 측정장비 또는 관찰장비만을 사용하는 연구 (3) 「식품위생법 시행규칙」 제3조에 따라 판매 등이 허용되는 식품 또는 식품첨가물을 이용하여 맛이나 질을 평가하는 연구 (4) 「화장품법」 제8조에 따른 안전기준에 맞는 화장품을 이용하여 사용감 또는 만족도 등을 조사하는 연구	☐	☑
2) 연구대상자와 배아·난자·정자 또는 인체유래물의 기증자를 직접 대면하더라도 연구대상자와 배아·난자·정자 또는 인체유래물의 기증자가 특정되지 않고 「개인정보보호법」 제23조에 따른 민감정보를 수집하거나 기록하지 않는 연구(사상, 신념, 노동조합, 정당의 가입, 탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 유전자검사 등의 결과로 얻어진 유전정보, 형의 실효 등에 관한 법률 제2조제5호에 따른 범죄경력자료에 해당하는 정보)	☐	☑
3) 연구대상자와 배아·난자·정자 또는 인체유래물의 기증자에 대한 기존의 자료나 문서를 이용하는 연구. 다만, 서울아산병원 의무기록에 직접 접근하는 연구는 제외	☑	☐
2. 생명윤리 및 안전에 관한 법률에 의거한 인체유래물연구 (기본사항이 모두 '예'이고, 아래 항목 중 하나 이상에 해당하는 경우 심의면제가 가능함)	**예**	**아니오**
기본사항		
1) 인체유래물기증자 및 공공에 미치는 위험이 미미한 경우임	☐	☑
아래		

주소 : 05505 서울특별시 송파구 올림픽로 43길 88 서울아산병원 TEL : 02-3010-7166, FAX : 02-3010-4163

1) 연구자가 개인정보를 수집·기록하지 않은 연구 중 다음 어느 하나에 해당하는 연구 (1) 인체유래물은행이 수집·보관하고 있는 인체유래물과 그로부터 얻은 유전정보(이하 "인체유래물 등"이라 한다)를 제공받아 사용하는 연구로서 인체유래물 등을 제공한 인체유래물은행을 통하지 않으면 개인정보를 확인할 수 없는 연구 (2) 의료기관에서 치료 및 진단을 목적으로 사용하고 남은 인체유래물 등을 이용하여 정확도 검사 등 검사실 정도관리 및 검사법 평가 등을 수행하는 연구 (3) 인체유래물을 직접 채취하지 않는 경우로서 일반 대중이 이용할 수 있도록 인체유래물로부터 분리리가공된 연구재료(병원체, 세포주 등을 포함한다)를 사용하는 연구 (4) 연구자가 인체유래물 기증자의 개인식별정보를 알 수 없으며, 연구를 통해 얻어진 결과가 기증자 개인의 유전적 특징과 관계가 없는 연구. 다만, 배아줄기세포주를 이용한 연구는 제외		☐	☑
2) 「초·중등교육법」 제2조 및 「고등교육법」 제2조에 따른 학교와 보건복지부장관이 정하는 교육기관에서 통상적인 교육과정의 범위에서 실무와 관련하여 수행하는 연구		☐	☑
3) 공중보건상 긴급한 조치가 필요한 상황에서 국가 또는 지방자치단체가 직접 수행하거나 위탁한 연구		☐	☑
3. US DHHS 또는 US FDA의 규제를 받는 연구인 경우 **(아래 6가지 범주 중 하나 이상에 해당되어야 하고, 해당되는 범주의 세부 내용이 모두 '예'이어야 함)**		**예**	**아니오**
범주 1	1) 일반적으로 받아들여지는 교육환경에서 수행되는 연구임	☐	☑
	2) 일반적인 교육과정을 포함함(예, 정규 및 특수 교수법 전략에 대한 연구 교수법, 교과과정 또는 수업관리 방법들 간의 비교 혹은 효과성에 대한 연구)	☐	☑
	3) 연구대상자에 수감자를 포함하지 않은 연구	☐	☑
	4) U.S. FDA 규제를 받지 않은 연구	☐	☑
범주 2	1) 교육용 시험(인지, 진단, 적성, 성취도), 설문조사, 면접조사, 대중 행동 관찰 등이 한가지 이상 포함된 연구	☐	☑
	2) 연구대상자의 정보가 적절하게 보호되지 않아서 연구대상자가 민형사상의 책임을 지게 되거나 재정적인 상황, 고용, 보험 또는 명성을 악화시키거나 오명을 씌우게 하는 경우가 없음	☐	☑
	3) 연구대상자를 식별하거나 직·간접적인 식별자를 통해 연구대상자를 식별할 수 있는 상태로 정보가 기록되지 않음	☐	☑
범주 3	1) 교육용 시험(인지, 진단, 적성, 성취도), 설문조사, 면접조사, 대중 행동 관찰 등이 한가지 이상 포함된 연구	☐	☑
	2) 다음 조건 중 어느 하나에 해당될 때 · 공직에 근무 중이거나 근무 예정인 자를 대상으로 하는 연구 또는 만 12세 미만의 자를 포함하지 않는 연구 · 연구대상자의 사적 정보에 대한 비밀이 법률에 따라 연구기간 동안 또는 영구적으로 보장되는 경우	☐	☑
	3) 연구대상자에 수감자를 포함하지 않은 연구	☐	☑
	4) U.S. FDA 규제를 받지 않은 연구	☐	☑
범주 4	1) 문서, 기록, 검사결과, 문헌 그리고 병리적 표본 또는 진단표본 등과 같이 기존에 확보된 자료를 대상으로 하는 연구	☐	☑
	2) 다음 조건 중 어느 하나에 해당될 때 · 공개적으로 이용 가능한 자료 · 직접적으로 또는 연구대상자와 연결된 식별자를 통해 연구대상자를 확인할 수 없는 연구	☐	☑
범주 5	1) 프로젝트가 연구이거나 시범사업임	☐	☑

주소 : 05505 서울특별시 송파구 올림픽로 43길 88 서울아산병원 TEL : 02-3010-7166, FAX : 02-3010-4163

범주 5	2) 연방정부나 기관장의 승인 하에 수행되는 연구	☐	☑
	3) 다음과 같은 항목을 평가하기 위한 연구 · 공익 또는 서비스 프로그램 · 이러한 프로그램 하에 얻어지는 이익 또는 서비스의 절차 · 이러한 프로그램 또는 절차의 변경 또는 대체 가능성 · 이러한 프로그램 하에 이익 또는 서비스를 위한 지불수준 또는 방법의 변경 가능성	☐	☑
	4) 연구가 공공의 이익이나 서비스로 이어질 때	☐	☑
	5) 연구가 특수한 연방정보의 권한으로 실시될 때	☐	☑
	6) IRB 심의 요구조건이 없을 때	☐	☑
	7) 비침습적 또는 연구대상자의 사생활 침해가 포함되지 않은 연구	☐	☑
	8) 연구대상자에 수감자를 포함하지 않은 연구	☐	☑
	9) U.S. FDA 규제를 받지 않은 연구	☐	☑
범주 6	1) 식품의 질과 맛 평가 또는 소비자 선호도 확인을 위한 연구	☐	☑
	2) 다음 조건 중 어느 하나에 해당될 때 · 식품첨가물이 없거나 · 정부기관에서 안전성을 승인한 식품첨가물만을 사용한 식품 대상	☐	☑
	3) 연구대상자에 수감자를 포함하지 않은 연구	☐	☑

최종결과	심의면제 가능여부 판정일	2020년 11월 25일
	본 신규과제는 IRB에서 규정한 심의면제 조건에 부합하여 심의를 면제합니다. 판정일 이후부터 연구개시가 가능하며 본 확인서를 보관하시기 바랍니다.	

기타 의견

본 위원회에서는 연구자가 제출하신 심의면제를 검토한 결과 승인하기로 결정하였습니다.

심의면제 연구라도 AMC IRB SOP와 HRPP 규정을 준수하여야 함을 유의하시고 연구를 진행하여 주시기 바랍니다.

제출자료 목록 및 버전번호

연구계획서(version 1.0)

Date: 2021/04/28

주소 : 05505 서울특별시 송파구 올림픽로 43길 88 서울아산병원 TEL : 02-3010-7166, FAX : 02-3010-4163

임상연구심의위원회/기관생명윤리위원회	위원장 이무송 (인)

본 임상연구심의위원회는 국제표준화추진회의(ICH), 의약품임상시험관리기준/의료기기임상시험실시기준(KGCP) 및 생명윤리및안전에관한법률 등 관련 법규를 준수합니다. 본 연구와 이해상충관계가 있는 위원이 있을 경우 해당 위원은 연구의 심의에서 배제하였습니다.

2. R code for analysis

```
############## I. Preparation ##########


getwd()
setwd("/vol/userdata3")

BR_INT <- read.csv("/vol/userdata3/sta_room083/BR_INT_FINAL.csv", fileEncoding = "euc-kr")

attach(BR_INT)


########### 2. packages installation ###########

library(reshape2)
library(purrr)
library(dplyr)
library(ggplot2)
library(MASS)
library(survival)
library(magrittr)
library(ggpubr)
library(GGally)
library(survminer)
library(cmprsk)
library(janitor)
library(xtable)
library(prodlim)
library(flexsurv)
library(lubridate)
library(tidyverse)
library(car)


########### 3.Data tansformation ###########

BR_INT$JID <- as.factor(BR_INT$JID)
BR_INT$sex<- as.factor(BR_INT$sex)
BR_INT$insurance <- as.factor(BR_INT$insurance)

BR_INT$DM_Hx_cat <- as.factor(BR_INT$DM_Hx_cat)
BR_INT$HT_Hx_cat <- as.factor(BR_INT$HT_Hx_cat)
BR_INT$DYP_Hx_cat <- as.factor(BR_INT$DYP_Hx_cat)
BR_INT$Other_CA_cat <- as.factor(BR_INT$Other_CA_cat)

BR_INT$Surgery_cat <- as.factor(BR_INT$Surgery_cat)
BR_INT$ChemoTx_cat <- as.factor(BR_INT$ChemoTx_cat)
BR_INT$ETx_cat <- as.factor(BR_INT$ETx_cat)
BR_INT$RT_cat <- as.factor(BR_INT$RT_cat)
BR_INT$TZM_cat <- as.factor(BR_INT$TZM_cat)

BR_INT$BC_DATE <- as.Date(BR_INT$BC_DATE)

BR_INT$In_hospital_death_STATUS <- as.factor(BR_INT$In_hospital_death_STATUS)

BR_INT$PE_DATE <- as.Date(BR_INT$PE_DATE)
BR_INT$PE_Region <- as.factor(BR_INT$PE_Region)
BR_INT$PE_Hospital <- as.factor(BR_INT$PE_Hospital)
BR_INT$PE_Depart <- as.factor(BR_INT$PE_Depart)

BR_INT$SG_DATE <- as.Date(BR_INT$SG_DATE)
BR_INT$SG_Region <- as.factor(BR_INT$SG_Region)
BR_INT$SG_Hospital <- as.factor(BR_INT$SG_Hospital)
BR_INT$SG_Depart <- as.factor(BR_INT$SG_Depart)

BR_INT$CTX_DATE <- as.Date(   as.character(BR_INT$CTX_DATE), format="%Y-%m-%d")
BR_INT$CTX_Region <- as.factor(BR_INT$CTX_Region)
BR_INT$CTX_Hospital <- as.factor(BR_INT$CTX_Hospital)
BR_INT$CTX_Depart <- as.factor(BR_INT$CTX_Depart)

BR_INT$ET_DATE <- as.Date(BR_INT$ET_DATE)
BR_INT$ET_Region <- as.factor(BR_INT$ET_Region)
BR_INT$ET_Hospital <- as.factor(BR_INT$ET_Hospital)
BR_INT$ET_Depart <- as.factor(BR_INT$ET_Depart)

BR_INT$STX_DATE <- as.Date(BR_INT$STX_DATE)
BR_INT$STX_Region <- as.factor(BR_INT$STX_Region)
BR_INT$STX_Hospital <- as.factor(BR_INT$STX_Hospital)
BR_INT$STX_Depart <- as.factor(BR_INT$STX_Depart)

BR_INT$BC_DATE <- as.Date(BR_INT$BC_DATE)
BR_INT$BC_Region <- as.factor(BR_INT$BC_Region)
BR_INT$BC_Hospital <- as.factor(BR_INT$BC_Hospital)
BR_INT$BC_Depart <- as.factor(BR_INT$BC_Depart)

BR_INT$RT_DATE <- as.Date(BR_INT$RT_DATE)
BR_INT$RT_Region <- as.factor(BR_INT$RT_Region)
BR_INT$RT_Hospital <- as.factor(BR_INT$RT_Hospital)
BR_INT$RT_Depart <- as.factor(BR_INT$RT_Depart)

BR_INT$PE_Region_cat1 <- as.factor(BR_INT$PE_Region_cat1)
BR_INT$PE_Region_cat2 <- as.factor(BR_INT$PE_Region_cat2)
BR_INT$PE_Region_cat3 <- as.factor(BR_INT$PE_Region_cat3)
BR_INT$PE_Hospital_cat1 <- as.factor(BR_INT$PE_Hospital_cat1)
BR_INT$PE_Hospital_cat2 <- as.factor(BR_INT$PE_Hospital_cat2)
BR_INT$PE_Hospital_cat3 <- as.factor(BR_INT$PE_Hospital_cat3)
```

```
BR_INT$PE_Hospital_cat4 <- as.factor(BR_INT$PE_Hospital_cat4)
BR_INT$PE_Depart_cat1 <- as.factor(BR_INT$PE_Depart_cat1)

BR_INT$BC_Region_cat1 <- as.factor(BR_INT$BC_Region_cat1)
BR_INT$BC_Region_cat2 <- as.factor(BR_INT$BC_Region_cat2)
BR_INT$BC_Region_cat3 <- as.factor(BR_INT$BC_Region_cat3)
BR_INT$BC_Hospital_cat1 <- as.factor(BR_INT$BC_Hospital_cat1)
BR_INT$BC_Hospital_cat2 <- as.factor(BR_INT$BC_Hospital_cat2)
BR_INT$BC_Hospital_cat3 <- as.factor(BR_INT$BC_Hospital_cat3)
BR_INT$BC_Hospital_cat4 <- as.factor(BR_INT$BC_Hospital_cat4)
BR_INT$BC_Depart_cat1 <- as.factor(BR_INT$BC_Depart_cat1)

BR_INT$SG_Region_cat1 <- as.factor(BR_INT$SG_Region_cat1)
BR_INT$SG_Region_cat2 <- as.factor(BR_INT$SG_Region_cat2)
BR_INT$SG_Region_cat3 <- as.factor(BR_INT$SG_Region_cat3)
BR_INT$SG_Hospital_cat1 <- as.factor(BR_INT$SG_Hospital_cat1)
BR_INT$SG_Hospital_cat2 <- as.factor(BR_INT$SG_Hospital_cat2)
BR_INT$SG_Hospital_cat3 <- as.factor(BR_INT$SG_Hospital_cat3)
BR_INT$SG_Hospital_cat4 <- as.factor(BR_INT$SG_Hospital_cat4)
BR_INT$SG_Depart_cat1 <- as.factor(BR_INT$SG_Depart_cat1)

BR_INT$CTX_Region_cat1 <- as.factor(BR_INT$CTX_Region_cat1)
BR_INT$CTX_Region_cat2 <- as.factor(BR_INT$CTX_Region_cat2)
BR_INT$CTX_Region_cat3 <- as.factor(BR_INT$CTX_Region_cat3)
BR_INT$CTX_Hospital_cat1 <- as.factor(BR_INT$CTX_Hospital_cat1)
BR_INT$CTX_Hospital_cat2 <- as.factor(BR_INT$CTX_Hospital_cat2)
BR_INT$CTX_Hospital_cat3 <- as.factor(BR_INT$CTX_Hospital_cat3)
BR_INT$CTX_Hospital_cat4 <- as.factor(BR_INT$CTX_Hospital_cat4)
BR_INT$CTX_Depart_cat1 <- as.factor(BR_INT$CTX_Depart_cat1)


BR_INT$RT_Region_cat1 <- as.factor(BR_INT$RT_Region_cat1)
BR_INT$RT_Region_cat2 <- as.factor(BR_INT$RT_Region_cat2)
BR_INT$RT_Region_cat3 <- as.factor(BR_INT$RT_Region_cat3)
BR_INT$RT_Hospital_cat1 <- as.factor(BR_INT$RT_Hospital_cat1)
BR_INT$RT_Hospital_cat2 <- as.factor(BR_INT$RT_Hospital_cat2)
BR_INT$RT_Hospital_cat3 <- as.factor(BR_INT$RT_Hospital_cat3)
BR_INT$RT_Hospital_cat4 <- as.factor(BR_INT$RT_Hospital_cat4)
BR_INT$RT_Depart_cat1 <- as.factor(BR_INT$RT_Depart_cat1)

BR_INT$STX_Region_cat1 <- as.factor(BR_INT$STX_Region_cat1)
BR_INT$STX_Region_cat2 <- as.factor(BR_INT$STX_Region_cat2)
BR_INT$STX_Region_cat3 <- as.factor(BR_INT$STX_Region_cat3)
BR_INT$STX_Hospital_cat1 <- as.factor(BR_INT$STX_Hospital_cat1)
BR_INT$STX_Hospital_cat2 <- as.factor(BR_INT$STX_Hospital_cat2)
BR_INT$STX_Hospital_cat3 <- as.factor(BR_INT$STX_Hospital_cat3)
BR_INT$STX_Hospital_cat4 <- as.factor(BR_INT$STX_Hospital_cat4)
BR_INT$STX_Depart_cat1 <- as.factor(BR_INT$STX_Depart_cat1)

BR_INT$PE_BC_DAY <- as.integer(BR_INT$PE_BC_DAY)
BR_INT$PE_SG_DAY <- as.integer(BR_INT$PE_SG_DAY)
BR_INT$PE_CTX_DAY <- as.integer(BR_INT$PE_CTX_DAY)
BR_INT$PE_ET_DAY <- as.integer(BR_INT$PE_ET_DAY)
BR_INT$PE_STX_DAY <- as.integer(BR_INT$PE_STX_DAY)

BR_INT$BC_SG_DAY <- as.integer(BR_INT$BC_SG_DAY)
BR_INT$BC_CTX_DAY <- as.integer(BR_INT$BC_CTX_DAY)
BR_INT$BC_ET_DAY <- as.integer(BR_INT$BC_ET_DAY)
BR_INT$BC_STX_DAY <- as.integer(BR_INT$BC_STX_DAY)

BR_INT$SG_RT_DAY <- as.integer(BR_INT$SG_RT_DAY)
BR_INT$CTX_RT_DAY <- as.integer(BR_INT$CTX_RT_DAY)
BR_INT$SG_CTX_DAY <- as.integer(BR_INT$SG_CTX_DAY)


########### 4.re-level ###########

BR_INT$age_cat1 <- relevel(BR_INT$age_cat1, ref="<50")
BR_INT$age_cat2 <- relevel(BR_INT$age_cat2, ref="<40")
BR_INT$age_cat3 <- relevel(BR_INT$age_cat3, ref="<50")

BR_INT$insurance  <- relevel(BR_INT$insurance, ref='4')

BR_INT$CCI_cat    <- relevel(BR_INT$CCI_cat, ref='0')
BR_INT$DM_Hx_cat  <- relevel(BR_INT$DM_Hx_cat, ref='0')
BR_INT$HT_Hx_cat  <- relevel(BR_INT$HT_Hx_cat, ref='0')
BR_INT$DYP_Hx_cat <- relevel(BR_INT$DYP_Hx_cat, ref='0')
BR_INT$Other_CA_cat <- relevel(BR_INT$Other_CA_cat, ref='0')

BR_INT$Surgery_cat <- relevel(BR_INT$Surgery_cat, ref='0')
BR_INT$ChemoTx_cat <- relevel(BR_INT$ChemoTx_cat, ref='0')
BR_INT$ETx_cat <- relevel(BR_INT$ETx_cat, ref='0')
BR_INT$RT_cat     <- relevel(BR_INT$RT_cat, ref='0')
BR_INT$TZM_cat    <- relevel(BR_INT$TZM_cat, ref='0')

BR_INT$PE_Region_cat1 <- relevel(BR_INT$PE_Region_cat1, ref='Seoul')
BR_INT$PE_Region_cat2 <- relevel(BR_INT$PE_Region_cat2, ref='Seoul')
BR_INT$PE_Region_cat3 <- relevel(BR_INT$PE_Region_cat3, ref='Seoul')
BR_INT$PE_Hospital_cat1 <- relevel(BR_INT$PE_Hospital_cat1, ref='Tertiary hospital')
BR_INT$PE_Hospital_cat2 <- relevel(BR_INT$PE_Hospital_cat2, ref='Tertiary hospital')
BR_INT$PE_Hospital_cat3 <- relevel(BR_INT$PE_Hospital_cat3, ref='Tertiary hospital')
BR_INT$PE_Hospital_cat4 <- relevel(BR_INT$PE_Hospital_cat4, ref='Tertiary hospital')
BR_INT$PE_Depart_cat1 <- relevel(BR_INT$PE_Depart_cat1, ref='GS')
BR_INT$SG_Region_cat1 <- relevel(BR_INT$SG_Region_cat1, ref='Seoul')
BR_INT$SG_Region_cat2 <- relevel(BR_INT$SG_Region_cat2, ref='Seoul')
BR_INT$SG_Region_cat3 <- relevel(BR_INT$SG_Region_cat3, ref='Seoul')
BR_INT$SG_Hospital_cat1 <- relevel(BR_INT$SG_Hospital_cat1, ref='Tertiary hospital')
BR_INT$SG_Hospital_cat2 <- relevel(BR_INT$SG_Hospital_cat2, ref='Tertiary hospital')
```

```
BR_INT$SG_Hospital_cat3 <- relevel(BR_INT$SG_Hospital_cat3, ref='Tertiary hospital')
BR_INT$SG_Hospital_cat4 <- relevel(BR_INT$SG_Hospital_cat4, ref='Tertiary hospital')
BR_INT$SG_Depart_cat1 <- relevel(BR_INT$SG_Depart_cat1, ref='GS')

BR_INT$CTX_Region_cat1 <- relevel(BR_INT$CTX_Region_cat1, ref='Seoul')
BR_INT$CTX_Region_cat2 <- relevel(BR_INT$CTX_Region_cat2, ref='Seoul')
BR_INT$CTX_Region_cat3 <- relevel(BR_INT$CTX_Region_cat3, ref='Seoul')
BR_INT$CTX_Hospital_cat1 <- relevel(BR_INT$CTX_Hospital_cat1, ref='Tertiary hospital')
BR_INT$CTX_Hospital_cat2 <- relevel(BR_INT$CTX_Hospital_cat2, ref='Tertiary hospital')
BR_INT$CTX_Hospital_cat3 <- relevel(BR_INT$CTX_Hospital_cat3, ref='Tertiary hospital')
BR_INT$CTX_Hospital_cat4 <- relevel(BR_INT$CTX_Hospital_cat4, ref='Tertiary hospital')
BR_INT$CTX_Depart_cat1 <- relevel(BR_INT$CTX_Depart_cat1, ref='GS')

BR_INT$STX_Region_cat1 <- relevel(BR_INT$STX_Region_cat1, ref='Seoul')
BR_INT$STX_Region_cat2 <- relevel(BR_INT$STX_Region_cat2, ref='Seoul')
BR_INT$STX_Region_cat3 <- relevel(BR_INT$STX_Region_cat3, ref='Seoul')
BR_INT$STX_Hospital_cat1 <- relevel(BR_INT$STX_Hospital_cat1, ref='Tertiary hospital')
BR_INT$STX_Hospital_cat2 <- relevel(BR_INT$STX_Hospital_cat2, ref='Tertiary hospital')
BR_INT$STX_Hospital_cat3 <- relevel(BR_INT$STX_Hospital_cat3, ref='Tertiary hospital')
BR_INT$STX_Hospital_cat4 <- relevel(BR_INT$STX_Hospital_cat4, ref='Tertiary hospital')
BR_INT$STX_Depart_cat1 <- relevel(BR_INT$STX_Depart_cat1, ref='GS')

BR_INT$RT_Region_cat1 <- relevel(BR_INT$RT_Region_cat1, ref='Seoul')
BR_INT$RT_Region_cat2 <- relevel(BR_INT$RT_Region_cat2, ref='Seoul')
BR_INT$RT_Region_cat3 <- relevel(BR_INT$RT_Region_cat3, ref='Seoul')
BR_INT$RT_Hospital_cat1 <- relevel(BR_INT$RT_Hospital_cat1, ref='Tertiary hospital')
BR_INT$RT_Hospital_cat2 <- relevel(BR_INT$RT_Hospital_cat2, ref='Tertiary hospital')
BR_INT$RT_Hospital_cat3 <- relevel(BR_INT$RT_Hospital_cat3, ref='Tertiary hospital')
BR_INT$RT_Hospital_cat4 <- relevel(BR_INT$RT_Hospital_cat4, ref='Tertiary hospital')
BR_INT$RT_Depart_cat1 <- relevel(BR_INT$RT_Depart_cat1, ref='GS')

BR_INT$BC_Region_cat1 <- relevel(BR_INT$BC_Region_cat1, ref='Seoul')
BR_INT$BC_Region_cat2 <- relevel(BR_INT$BC_Region_cat2, ref='Seoul')
BR_INT$BC_Region_cat3 <- relevel(BR_INT$BC_Region_cat3, ref='Seoul')
BR_INT$BC_Hospital_cat1 <- relevel(BR_INT$BC_Hospital_cat1, ref='Tertiary hospital')
BR_INT$BC_Hospital_cat2 <- relevel(BR_INT$BC_Hospital_cat2, ref='Tertiary hospital')
BR_INT$BC_Hospital_cat3 <- relevel(BR_INT$BC_Hospital_cat3, ref='Tertiary hospital')
BR_INT$BC_Hospital_cat4 <- relevel(BR_INT$BC_Hospital_cat4, ref='Tertiary hospital')
BR_INT$BC_Depart_cat1 <- relevel(BR_INT$BC_Depart_cat1, ref='GS')


########### 5.Tranform ###########

##### Pathologic examination #####

BR_INT <- transform (BR_INT, PE_YEAR = ifelse(PE_DATE >= "2010-01-01" & PE_DATE <=
"2010-12-31", "2010",
                                      ifelse(PE_DATE >= "2011-01-01" & PE_DATE <= "2011-12-31",
"2011",
                                        ifelse(PE_DATE >= "2012-01-01" & PE_DATE <= "2012-12-31",
"2012",
                                          ifelse(PE_DATE >= "2013-01-01" & PE_DATE <= "2013-12-
31", "2013",
                                            ifelse(PE_DATE >= "2014-01-01" & PE_DATE <=
"2014-12-31", "2014",
                                              ifelse(PE_DATE >= "2015-01-01" & PE_DATE <=
"2015-12-31", "2015",
                                                ifelse(PE_DATE >= "2016-01-01" & PE_DATE
<= "2016-12-31", "2016",
                                                  ifelse(PE_DATE >= "2017-01-01" &
PE_DATE <= "2017-12-31", "2017",
                                                    ifelse(PE_DATE >= "2018-01-01" &
PE_DATE <= "2018-12-31", "2018",

                                                      "No"))))))))))


BR_INT <- transform (BR_INT, PE_YEAR1 = ifelse(PE_DATE >= "2010-01-01" & PE_DATE <=
"2015-12-31", "2010-2015",

                                      ifelse(PE_DATE >= "2016-01-01" & PE_DATE <= "2017-12-31",
"2016-2017",

                                      "No")))

BR_INT$PE_YEAR <- as.factor(BR_INT$PE_YEAR)
BR_INT$PE_YEAR1 <- as.factor(BR_INT$PE_YEAR1)



##### Diagnosis #####

BR_INT <- transform (BR_INT, BC_YEAR = ifelse(BC_DATE >= "2010-01-01" & BC_DATE <=
"2010-12-31", "2010",

                                      ifelse(BC_DATE >= "2011-01-01" & BC_DATE <= "2011-12-31",
"2011",
```

```
                                ifelse(BC_DATE >= "2012-01-01" & BC_DATE <= "2012-12-31",
"2012",
                                ifelse(BC_DATE >= "2013-01-01" & BC_DATE <= "2013-12-31",
"2013",
                                ifelse(BC_DATE >= "2014-01-01" & BC_DATE <= "2014-12-31",
"2014",
                                ifelse(BC_DATE >= "2015-01-01" & BC_DATE <= "2015-12-31",
"2015",
                                ifelse(BC_DATE >= "2016-01-01" & BC_DATE <= "2016-12-31",
"2016",
                                ifelse(BC_DATE >= "2017-01-01" & BC_DATE <= "2017-12-31",
"2017",
                                ifelse(BC_DATE >= "2018-01-01" & BC_DATE <= "2018-12-31",
"2018",
                                "No")))))))))

BR_INT <- transform (BR_INT, BC_YEAR1 = ifelse(BC_DATE >= "2010-01-01" & BC_DATE
<= "2015-12-31", "2010-2015",
                                ifelse(BC_DATE >= "2016-01-01" & BC_DATE <= "2017-12-31",
"2016-2017",
                                "No")))

BR_INT$BC_YEAR <- as.factor(BR_INT$BC_YEAR)
BR_INT$BC_YEAR1 <- as.factor(BR_INT$BC_YEAR1)


##### Initial treatment_All #####


BR_INT <- transform (BR_INT, STX_YEAR = ifelse(STX_DATE >= "2010-01-01" & STX_DATE
<= "2010-12-31", "2010",
                                ifelse(STX_DATE >= "2011-01-01" & STX_DATE <= "2011-12-31",
"2011",
                                    ifelse(STX_DATE >= "2012-01-01" & STX_DATE <= "2012-12-
31", "2012",
                                        ifelse(STX_DATE >= "2013-01-01" & STX_DATE <=
"2013-12-31", "2013",
                                            ifelse(STX_DATE >= "2014-01-01" & STX_DATE <=
"2014-12-31", "2014",
                                                ifelse(STX_DATE >= "2015-01-01" & STX_DATE
<= "2015-12-31", "2015",
                                                    ifelse(STX_DATE >= "2016-01-01" &
STX_DATE <= "2016-12-31", "2016",
                                                        ifelse(STX_DATE >= "2017-01-01" &
STX_DATE <= "2017-12-31", "2017",
                                                            ifelse(STX_DATE >= "2018-01-01" &
STX_DATE <= "2018-12-31", "2018",
                                                                "No")))))))))

BR_INT <- transform (BR_INT, STX_YEAR1 = ifelse(STX_DATE >= "2010-01-01" &
STX_DATE <= "2015-12-31", "2010-2015",
                                ifelse(STX_DATE >= "2016-01-01" & STX_DATE <= "2017-12-31",
"2016-2017",
                                "No")))

BR_INT$STX_YEAR <- as.factor(BR_INT$STX_YEAR)
BR_INT$STX_YEAR1 <- as.factor(BR_INT$STX_YEAR1)


##### Initial treatment_surgery #####


BR_INT <- transform (BR_INT, SG_YEAR = ifelse(SG_DATE >= "2010-01-01" & SG_DATE <=
"2010-12-31", "2010",
                                ifelse(SG_DATE >= "2011-01-01" & SG_DATE <= "2011-12-31",
"2011",
                                    ifelse(SG_DATE >= "2012-01-01" & SG_DATE <= "2012-12-31",
"2012",
```

```
                                        ifelse(SG_DATE >= "2013-01-01" & SG_DATE <= "2013-12-
31", "2013",
                                          ifelse(SG_DATE >= "2014-01-01" & SG_DATE <=
"2014-12-31", "2014",
                                            ifelse(SG_DATE >= "2015-01-01" & SG_DATE <=
"2015-12-31", "2015",
                                              ifelse(SG_DATE >= "2016-01-01" & SG_DATE
<= "2016-12-31", "2016",
                                                ifelse(SG_DATE >= "2017-01-01" &
SG_DATE <= "2017-12-31", "2017",
                                                  ifelse(SG_DATE >= "2018-01-01" &
SG_DATE <= "2018-12-31", "2018",
                                                    "NULL"))))))))))

BR_INT <- transform (BR_INT, SG_YEAR1 = ifelse(SG_DATE >= "2010-01-01" & SG_DATE <=
"2015-12-31", "2010-2015",
                          ifelse(SG_DATE >= "2016-01-01" & SG_DATE <= "2018-12-31",
"2016-2018",
                            "NULL")))

BR_INT$SG_YEAR <- as.factor(BR_INT$SG_YEAR)
BR_INT$SG_YEAR1 <- as.factor(BR_INT$SG_YEAR1)
```

```
##### Initial treatment_chemotherapy #####
```

```
BR_INT <- transform (BR_INT, CTX_YEAR = ifelse(CTX_DATE >= "2010-01-01" &
CTX_DATE <= "2010-12-31", "2010",
                          ifelse(CTX_DATE >= "2011-01-01" & CTX_DATE <= "2011-12-31",
"2011",
                            ifelse(CTX_DATE >= "2012-01-01" & CTX_DATE <= "2012-12-
31", "2012",
                              ifelse(CTX_DATE >= "2013-01-01" & CTX_DATE <=
"2013-12-31", "2013",
                                ifelse(CTX_DATE >= "2014-01-01" & CTX_DATE <=
"2014-12-31", "2014",
                                  ifelse(CTX_DATE >= "2015-01-01" & CTX_DATE
<= "2015-12-31", "2015",
                                    ifelse(CTX_DATE >= "2016-01-01" &
CTX_DATE <= "2016-12-31", "2016",
                                      ifelse(CTX_DATE >= "2017-01-01" &
CTX_DATE <= "2017-12-31", "2017",
                                        ifelse(CTX_DATE >= "2018-01-01" &
CTX_DATE <= "2018-12-31", "2018",
                                          "NULL"))))))))))

BR_INT <- transform (BR_INT, CTX_YEAR1 = ifelse(CTX_DATE >= "2010-01-01" &
CTX_DATE <= "2015-12-31", "2010-2015",
                          ifelse(CTX_DATE >= "2016-01-01" & CTX_DATE <= "2018-12-31",
"2016-2018",
                            "NULL")))

BR_INT$CTX_YEAR <- as.factor(BR_INT$CTX_YEAR)
BR_INT$CTX_YEAR1 <- as.factor(BR_INT$CTX_YEAR1)
```

```
##### Initial treatment_radiotherapy #####
```

```
BR_INT <- transform (BR_INT, RT_YEAR = ifelse(RT_DATE >= "2010-01-01" & RT_DATE <=
"2010-12-31", "2010",
                          ifelse(RT_DATE >= "2011-01-01" & RT_DATE <= "2011-12-31",
"2011",
                            ifelse(RT_DATE >= "2012-01-01" & RT_DATE <= "2012-12-31",
"2012",
                              ifelse(RT_DATE >= "2013-01-01" & RT_DATE <= "2013-12-
31", "2013",
```

```
                                        ifelse(RT_DATE >= "2014-01-01" & RT_DATE <=
"2014-12-31", "2014",
                                           ifelse(RT_DATE >= "2015-01-01" & RT_DATE <=
"2015-12-31", "2015",
                                              ifelse(RT_DATE >= "2016-01-01" & RT_DATE
<= "2016-12-31", "2016",
                                                 ifelse(RT_DATE >= "2017-01-01" &
RT_DATE <= "2017-12-31", "2017",
                                                    ifelse(RT_DATE >= "2018-01-01" &
RT_DATE <= "2018-12-31", "2018",
                                                       "NULL"))))))))))

BR_INT <- transform (BR_INT, RT_YEAR1 = ifelse(RT_DATE >= "2010-01-01" & RT_DATE <=
"2015-12-31", "2010-2015",
                    ifelse(RT_DATE >= "2016-01-01" & RT_DATE <= "2018-12-31",
"2016-2018",
                       "NULL")))

BR_INT$RT_YEAR <- as.factor(BR_INT$RT_YEAR)
BR_INT$RT_YEAR1 <- as.factor(BR_INT$RT_YEAR1)

BR_INT <- transform (BR_INT, Surgery_cat1 = ifelse(as.factor(SG_DATE) !="", '1','0'))
BR_INT <- transform (BR_INT, ChemoTx_cat1 = ifelse(as.factor(CTX_DATE) =="", '0','1'))

BR_INT <- transform (BR_INT, RT_cat1 = ifelse(as.factor(RT_DATE) =="", '0','1'))


########### 6. Subgroup ###########

BR_INT <- BR_INT %>% filter(PE_Region_cat2 != 'NULL')

BR_INT_PE_SEOUL <- BR_INT %>% filter(PE_Region_cat2 == 'Seoul')
BR_INT_PE_METRO <- BR_INT %>% filter(PE_Region_cat2 == 'Metropolitan city')
BR_INT_PE_OTHER <- BR_INT %>% filter(PE_Region_cat2 == 'Others')

BR_INT_BC_SEOUL <- BR_INT %>% filter(BC_Region_cat2 == 'Seoul')
BR_INT_BC_METRO <- BR_INT %>% filter(BC_Region_cat2 == 'Metropolitan city')
BR_INT_BC_OTHER <- BR_INT %>% filter(BC_Region_cat2 == 'Others')

BR_INT_STX_SEOUL <- BR_INT %>% filter(STX_Region_cat2 == 'Seoul')
BR_INT_STX_METRO <- BR_INT %>% filter(STX_Region_cat2 == 'Metropolitan city')
BR_INT_STX_OTHER <- BR_INT %>% filter(STX_Region_cat2 == 'Others')

BR_INT_SG_SEOUL <- BR_INT %>% filter(SG_Region_cat2 == 'Seoul')
BR_INT_SG_METRO <- BR_INT %>% filter(SG_Region_cat2 == 'Metropolitan city')
BR_INT_SG_OTHER <- BR_INT %>% filter(SG_Region_cat2 == 'Others')

BR_INT_CTX_SEOUL <- BR_INT %>% filter(CTX_Region_cat2 == 'Seoul')
BR_INT_CTX_METRO <- BR_INT %>% filter(CTX_Region_cat2 == 'Metropolitan city')
BR_INT_CTX_OTHER <- BR_INT %>% filter(CTX_Region_cat2 == 'Others')

BR_INT_RT_SEOUL <- BR_INT %>% filter(RT_Region_cat2 == 'Seoul')
BR_INT_RT_METRO <- BR_INT %>% filter(RT_Region_cat2 == 'Metropolitan city')
BR_INT_RT_OTHER <- BR_INT %>% filter(RT_Region_cat2 == 'Others')


BR_INT_PE_TH <- BR_INT %>% filter(PE_Hospital_cat2 == 'Tertiary hospital')
BR_INT_PE_GH <- BR_INT %>% filter(PE_Hospital_cat2 == 'General hospital')
BR_INT_PE_OH <- BR_INT %>% filter(PE_Hospital_cat2 == 'Others')

BR_INT_BC_TH <- BR_INT %>% filter(BC_Hospital_cat2 == 'Tertiary hospital')
BR_INT_BC_GH <- BR_INT %>% filter(BC_Hospital_cat2 == 'General hospital')
BR_INT_BC_OH <- BR_INT %>% filter(BC_Hospital_cat2 == 'Others')

BR_INT_STX_TH <- BR_INT %>% filter(STX_Hospital_cat2 == 'Tertiary hospital')
BR_INT_STX_GH <- BR_INT %>% filter(STX_Hospital_cat2 == 'General hospital')
BR_INT_STX_OH <- BR_INT %>% filter(STX_Hospital_cat2 == 'Others')

BR_INT_SG_TH <- BR_INT %>% filter(SG_Hospital_cat2 == 'Tertiary hospital')
BR_INT_SG_GH <- BR_INT %>% filter(SG_Hospital_cat2 == 'General hospital')
BR_INT_SG_OH <- BR_INT %>% filter(SG_Hospital_cat2 == 'Others')

BR_INT_CTX_TH <- BR_INT %>% filter(CTX_Hospital_cat2 == 'Tertiary hospital')
BR_INT_CTX_GH <- BR_INT %>% filter(CTX_Hospital_cat2 == 'General hospital')
BR_INT_CTX_OH <- BR_INT %>% filter(CTX_Hospital_cat2 == 'Others')

BR_INT_RT_TH <- BR_INT %>% filter(RT_Hospital_cat2 == 'Tertiary hospital')
BR_INT_RT_GH <- BR_INT %>% filter(RT_Hospital_cat2 == 'General hospital')
BR_INT_RT_OH <- BR_INT %>% filter(RT_Hospital_cat2 == 'Others')


BR_INT_TAM <- BR_INT %>% filter(ET_regimen_cat == 'INT'|ET_regimen_cat =='INT+AI')
BR_INT_NONE <- BR_INT %>% filter(ET_regimen_cat == 'None'|ET_regimen_cat =='AI')

str(BR_INT_TAM)
head(BR_INT_TAM)
str(BR_INT_NONE)
```

```
head(BR_INT_NONE)

########## II. Analysis ##########

##### Table 1. Characteristics of subjects by regions of primary care #####

attach(BR_INT)
detach(BR_INT)

ftable(sex)
100*prop.table(table(sex))

######## Total ########

##### Age at diagnosis
summarise(count=n(), BR_INT,age_mean=mean(age,na.rm=TRUE), age_sd=sd(age, na.rm=TRUE),
age_median=median(age,na.rm=TRUE),
        age_quartile_low=quantile(age, 0.25, na.rm=TRUE), age_quartile_high=quantile(age, 0.75,
na.rm=TRUE))

##### Insurance (4: Health insurance, 5: Medicare)
ftable(insurance)
100*prop.table(table(insurance))

##### CCI
summarise(count=n(), BR_INT,CCI_mean=mean(CCI,na.rm=TRUE), CCI_sd=sd(CCI,
na.rm=TRUE), CCI_median=median(CCI,na.rm=TRUE),
        CCI_quartile_low=quantile(CCI, 0.25, na.rm=TRUE), CCI_quartile_high=quantile(CCI, 0.75,
na.rm=TRUE))

##### Previous DM, HT, DYS
ftable(DM_Hx_cat)
100*prop.table(table(DM_Hx_cat))

ftable(HT_Hx_cat)
100*prop.table(table(HT_Hx_cat))

ftable(DYP_Hx_cat)
100*prop.table(table(DYP_Hx_cat))


##### Surgery
ftable(Surgery_cat1)
100*prop.table(table(Surgery_cat1))

##### Chemotherapy
ftable(ChemoTx_cat1)
100*prop.table(table(ChemoTx_cat1))
#### RT
ftable(RT_cat1)
100*prop.table(table(RT_cat1))

##### ETx
ftable(ETx_cat)
100*prop.table(table(ETx_cat))

##### TZM
ftable(TZM_cat)
100*prop.table(table(TZM_cat))


##### In_hospital_death_STATUS
ftable(In_hospital_death_STATUS)
100*prop.table(table(In_hospital_death_STATUS))


##### Median F/U
summarise(count=n(), BR_INT,BR_LAST_TIME_mean=mean(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_sd=sd(BR_LAST_TIME, na.rm=TRUE),
        BR_LAST_TIME_median=median(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_quartile_low=quantile(BR_LAST_TIME, 0.25, na.rm=TRUE),
        BR_LAST_TIME_quartile_high=quantile(BR_LAST_TIME, 0.75, na.rm=TRUE))
```

```
######## Seoul ########

##### Age at diagnosis
summarise(count=n(), BR_INT_PE_SEOUL, age_mean=mean(age,na.rm=TRUE), age_sd=sd(age,
na.rm=TRUE), age_median=median(age,na.rm=TRUE),
        age_quartile_low=quantile(age, 0.25, na.rm=TRUE), age_quartile_high=quantile(age, 0.75,
na.rm=TRUE))

##### Insurance (4: Health insurance, 5: Medicare)
ftable(BR_INT_PE_SEOUL$insurance)
100*prop.table(table(BR_INT_PE_SEOUL$insurance))

##### CCI
summarise(count=n(), BR_INT_PE_SEOUL, CCI_mean=mean(CCI,na.rm=TRUE),
CCI_sd=sd(CCI, na.rm=TRUE), CCI_median=median(CCI,na.rm=TRUE),
        CCI_quartile_low=quantile(CCI, 0.25, na.rm=TRUE), CCI_quartile_high=quantile(CCI, 0.75,
na.rm=TRUE))

##### Previous DM, HT, DYS
ftable(BR_INT_PE_SEOUL$DM_Hx_cat)
100*prop.table(table(BR_INT_PE_SEOUL$DM_Hx_cat))

ftable(BR_INT_PE_SEOUL$HT_Hx_cat)
100*prop.table(table(BR_INT_PE_SEOUL$HT_Hx_cat))

ftable(BR_INT_PE_SEOUL$DYP_Hx_cat)
100*prop.table(table(BR_INT_PE_SEOUL$DYP_Hx_cat))


##### Surgery
ftable(BR_INT_PE_SEOUL$Surgery_cat1)
100*prop.table(table(BR_INT_PE_SEOUL$Surgery_cat1))

##### Chemotherapy
ftable(BR_INT_PE_SEOUL$ChemoTx_cat1)
100*prop.table(table(BR_INT_PE_SEOUL$ChemoTx_cat1))

#### RT
ftable(BR_INT_PE_SEOUL$RT_cat1)
100*prop.table(table(BR_INT_PE_SEOUL$RT_cat1))

##### ETx
ftable(BR_INT_PE_SEOUL$ETx_cat)
100*prop.table(table(BR_INT_PE_SEOUL$ETx_cat))

##### TZM
ftable(BR_INT_PE_SEOUL$TZM_cat)
100*prop.table(table(BR_INT_PE_SEOUL$TZM_cat))

##### In_hospital_death_STATUS
ftable(BR_INT_PE_SEOUL$In_hospital_death_STATUS)
100*prop.table(table(BR_INT_PE_SEOUL$In_hospital_death_STATUS))


##### Median F/U
summarise(count=n(), BR_INT_PE_SEOUL,
BR_LAST_TIME_mean=mean(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_sd=sd(BR_LAST_TIME, na.rm=TRUE),
        BR_LAST_TIME_median=median(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_quartile_low=quantile(BR_LAST_TIME, 0.25, na.rm=TRUE),
        BR_LAST_TIME_quartile_high=quantile(BR_LAST_TIME, 0.75, na.rm=TRUE))



######## Metropolitan city ########

##### Age at diagnosis
summarise(count=n(), BR_INT_PE_METRO, age_mean=mean(age,na.rm=TRUE), age_sd=sd(age,
na.rm=TRUE), age_median=median(age,na.rm=TRUE),
        age_quartile_low=quantile(age, 0.25, na.rm=TRUE), age_quartile_high=quantile(age, 0.75,
na.rm=TRUE))

##### Insurance (4: Health insurance, 5: Medicare)
ftable(BR_INT_PE_METRO$insurance)
100*prop.table(table(BR_INT_PE_METRO$insurance))

##### CCI
summarise(count=n(), BR_INT_PE_METRO, CCI_mean=mean(CCI,na.rm=TRUE),
CCI_sd=sd(CCI, na.rm=TRUE), CCI_median=median(CCI,na.rm=TRUE),
        CCI_quartile_low=quantile(CCI, 0.25, na.rm=TRUE), CCI_quartile_high=quantile(CCI, 0.75,
na.rm=TRUE))

##### Previous DM, HT, DYS
ftable(BR_INT_PE_METRO$DM_Hx_cat)
100*prop.table(table(BR_INT_PE_METRO$DM_Hx_cat))

ftable(BR_INT_PE_METRO$HT_Hx_cat)
100*prop.table(table(BR_INT_PE_METRO$HT_Hx_cat))

ftable(BR_INT_PE_METRO$DYP_Hx_cat)
100*prop.table(table(BR_INT_PE_METRO$DYP_Hx_cat))


##### Surgery
```

```
ftable(BR_INT_PE_METRO$Surgery_cat1)
100*prop.table(table(BR_INT_PE_METRO$Surgery_cat1))

##### Chemotherapy
ftable(BR_INT_PE_METRO$ChemoTx_cat1)
100*prop.table(table(BR_INT_PE_METRO$ChemoTx_cat1))

#### RT
ftable(BR_INT_PE_METRO$RT_cat1)
100*prop.table(table(BR_INT_PE_METRO$RT_cat1))

##### ETx
ftable(BR_INT_PE_METRO$ETx_cat)
100*prop.table(table(BR_INT_PE_METRO$ETx_cat))

##### TZM
ftable(BR_INT_PE_METRO$TZM_cat)
100*prop.table(table(BR_INT_PE_METRO$TZM_cat))


##### In_hospital_death_STATUS
ftable(BR_INT_PE_METRO$In_hospital_death_STATUS)
100*prop.table(table(BR_INT_PE_METRO$In_hospital_death_STATUS))


##### Median F/U
summarise(count=n(), BR_INT_PE_METRO,
BR_LAST_TIME_mean=mean(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_sd=sd(BR_LAST_TIME, na.rm=TRUE),
        BR_LAST_TIME_median=median(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_quartile_low=quantile(BR_LAST_TIME, 0.25, na.rm=TRUE),
        BR_LAST_TIME_quartile_high=quantile(BR_LAST_TIME, 0.75, na.rm=TRUE))



######## Others ########

##### Age at diagnosis
summarise(count=n(), BR_INT_PE_OTHER, age_mean=mean(age,na.rm=TRUE), age_sd=sd(age,
na.rm=TRUE), age_median=median(age,na.rm=TRUE),
        age_quartile_low=quantile(age, 0.25, na.rm=TRUE), age_quartile_high=quantile(age, 0.75,
na.rm=TRUE))

##### Insurance (4: Health insurance, 5: Medicare)
ftable(BR_INT_PE_OTHER$insurance)
100*prop.table(table(BR_INT_PE_OTHER$insurance))

##### CCI
summarise(count=n(), BR_INT_PE_OTHER, CCI_mean=mean(CCI,na.rm=TRUE),
CCI_sd=sd(CCI, na.rm=TRUE), CCI_median=median(CCI,na.rm=TRUE),
        CCI_quartile_low=quantile(CCI, 0.25, na.rm=TRUE), CCI_quartile_high=quantile(CCI, 0.75,
na.rm=TRUE))

##### Previous DM, HT, DYS
ftable(BR_INT_PE_OTHER$DM_Hx_cat)
100*prop.table(table(BR_INT_PE_OTHER$DM_Hx_cat))

ftable(BR_INT_PE_OTHER$HT_Hx_cat)
100*prop.table(table(BR_INT_PE_OTHER$HT_Hx_cat))

ftable(BR_INT_PE_OTHER$DYP_Hx_cat)
100*prop.table(table(BR_INT_PE_OTHER$DYP_Hx_cat))


##### Surgery
ftable(BR_INT_PE_OTHER$Surgery_cat1)
100*prop.table(table(BR_INT_PE_OTHER$Surgery_cat1))

##### Chemotherapy
ftable(BR_INT_PE_OTHER$ChemoTx_cat1)
100*prop.table(table(BR_INT_PE_OTHER$ChemoTx_cat1))

#### RT
ftable(BR_INT_PE_OTHER$RT_cat1)
100*prop.table(table(BR_INT_PE_OTHER$RT_cat1))

##### ETx
ftable(BR_INT_PE_OTHER$ETx_cat)
100*prop.table(table(BR_INT_PE_OTHER$ETx_cat))

##### TZM
ftable(BR_INT_PE_OTHER$TZM_cat)
100*prop.table(table(BR_INT_PE_OTHER$TZM_cat))


##### In_hospital_death_STATUS
ftable(BR_INT_PE_OTHER$In_hospital_death_STATUS)
100*prop.table(table(BR_INT_PE_OTHER$In_hospital_death_STATUS))


##### Median F/U
summarise(count=n(), BR_INT_PE_OTHER,
BR_LAST_TIME_mean=mean(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_sd=sd(BR_LAST_TIME, na.rm=TRUE),
        BR_LAST_TIME_median=median(BR_LAST_TIME,na.rm=TRUE),
BR_LAST_TIME_quartile_low=quantile(BR_LAST_TIME, 0.25, na.rm=TRUE),
        BR_LAST_TIME_quartile_high=quantile(BR_LAST_TIME, 0.75, na.rm=TRUE))
```

```
########### Figure 3. Annual number of breast cancer patients by regions

library(ggplot2)

### PE

Regions <- c ("Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul",
        "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan",
"Metropolitan", "Metropolitan",
        "Others", "Others", "Others", "Others", "Others", "Others", "Others")
Numbers   <- c(length(which(BR_INT_PE_SEOUL$PE_YEAR=="2010")),
        length(which(BR_INT_PE_SEOUL$PE_YEAR=="2011")),
        length(which(BR_INT_PE_SEOUL$PE_YEAR=="2012")),
        length(which(BR_INT_PE_SEOUL$PE_YEAR=="2013")),
        length(which(BR_INT_PE_SEOUL$PE_YEAR=="2014")),
        length(which(BR_INT_PE_SEOUL$PE_YEAR=="2015")),
        length(which(BR_INT_PE_SEOUL$PE_YEAR=="2016")),

        length(which(BR_INT_PE_METRO$PE_YEAR=="2010")),
        length(which(BR_INT_PE_METRO$PE_YEAR=="2011")),
        length(which(BR_INT_PE_METRO$PE_YEAR=="2012")),
        length(which(BR_INT_PE_METRO$PE_YEAR=="2013")),
        length(which(BR_INT_PE_METRO$PE_YEAR=="2014")),
        length(which(BR_INT_PE_METRO$PE_YEAR=="2015")),
        length(which(BR_INT_PE_METRO$PE_YEAR=="2016")),

        length(which(BR_INT_PE_OTHER$PE_YEAR=="2010")),
        length(which(BR_INT_PE_OTHER$PE_YEAR=="2011")),
        length(which(BR_INT_PE_OTHER$PE_YEAR=="2012")),
        length(which(BR_INT_PE_OTHER$PE_YEAR=="2013")),
        length(which(BR_INT_PE_OTHER$PE_YEAR=="2014")),
        length(which(BR_INT_PE_OTHER$PE_YEAR=="2015")),
        length(which(BR_INT_PE_OTHER$PE_YEAR=="2016")))


Year <- c('2010','2011','2012','2013','2014','2015','2016',
      '2010','2011','2012','2013','2014','2015','2016',
      '2010','2011','2012','2013','2014','2015','2016')

PE_YEAR <- data.frame (Regions, Numbers, Year)

PE_YEAR$Regions <- relevel(PE_YEAR$Regions, ref='Seoul')

str(PE_YEAR)
summary(PE_YEAR)

dev.off()

ggplot(data=PE_YEAR, aes(x=Year, y=Numbers, group=Regions, color=Regions))+
  geom_line(size=0.8)+
  xlab("Year of pathologic examination")+
  ylim(500,10000) +
  scale_color_discrete("Regions") +
  theme(legend.position = c(0.15, 0.835),
     legend.title = element_text(size=13),
     legend.text = element_text(size=12),
     axis.title.x = element_text(size=15),
     axis.title.y = element_text(size=15),
     axis.text.x = element_text(size=12),
     axis.text.y = element_text(size=12)
     )



### BC

Regions <- c ("Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul",
        "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan",
"Metropolitan", "Metropolitan",
        "Others", "Others", "Others", "Others", "Others", "Others", "Others")


Numbers   <- c(length(which(BR_INT_BC_SEOUL$BC_YEAR=="2011")),
        length(which(BR_INT_BC_SEOUL$BC_YEAR=="2012")),
        length(which(BR_INT_BC_SEOUL$BC_YEAR=="2013")),
        length(which(BR_INT_BC_SEOUL$BC_YEAR=="2014")),
        length(which(BR_INT_BC_SEOUL$BC_YEAR=="2015")),
        length(which(BR_INT_BC_SEOUL$BC_YEAR=="2016")),
        length(which(BR_INT_BC_SEOUL$BC_YEAR=="2017")),

        length(which(BR_INT_BC_METRO$BC_YEAR=="2011")),
        length(which(BR_INT_BC_METRO$BC_YEAR=="2012")),
        length(which(BR_INT_BC_METRO$BC_YEAR=="2013")),
        length(which(BR_INT_BC_METRO$BC_YEAR=="2014")),
        length(which(BR_INT_BC_METRO$BC_YEAR=="2015")),
        length(which(BR_INT_BC_METRO$BC_YEAR=="2016")),
        length(which(BR_INT_BC_METRO$BC_YEAR=="2017")),

        length(which(BR_INT_BC_OTHER$BC_YEAR=="2011")),
        length(which(BR_INT_BC_OTHER$BC_YEAR=="2012")),
        length(which(BR_INT_BC_OTHER$BC_YEAR=="2013")),
```

```
            length(which(BR_INT_BC_OTHER$BC_YEAR=="2014")),
            length(which(BR_INT_BC_OTHER$BC_YEAR=="2015")),
            length(which(BR_INT_BC_OTHER$BC_YEAR=="2016")),
            length(which(BR_INT_BC_OTHER$BC_YEAR=="2017")))


Year <- c('2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017')

BC_YEAR <- data.frame (Regions, Numbers, Year)

BC_YEAR$Regions <- relevel(BC_YEAR$Regions, ref='Seoul')


str(BC_YEAR)

dev.off()


ggplot(data=BC_YEAR, aes(x=Year, y=Numbers, group=Regions, color=Regions))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer diagnosis")+
  ylim(500,10000) +
  scale_color_discrete("Regions") +
  theme(legend.position = c(0.15, 0.835),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15),
      axis.text.x = element_text(size=12),
      axis.text.y = element_text(size=12)
  )




### STX

Regions <- c ("Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul",
           "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan",
"Metropolitan", "Metropolitan",
           "Others", "Others", "Others", "Others", "Others", "Others", "Others")


Numbers   <- c(length(which(BR_INT_STX_SEOUL$STX_YEAR=="2011")),
            length(which(BR_INT_STX_SEOUL$STX_YEAR=="2012")),
            length(which(BR_INT_STX_SEOUL$STX_YEAR=="2013")),
            length(which(BR_INT_STX_SEOUL$STX_YEAR=="2014")),
            length(which(BR_INT_STX_SEOUL$STX_YEAR=="2015")),
            length(which(BR_INT_STX_SEOUL$STX_YEAR=="2016")),
            length(which(BR_INT_STX_SEOUL$STX_YEAR=="2017")),

            length(which(BR_INT_STX_METRO$STX_YEAR=="2011")),
            length(which(BR_INT_STX_METRO$STX_YEAR=="2012")),
            length(which(BR_INT_STX_METRO$STX_YEAR=="2013")),
            length(which(BR_INT_STX_METRO$STX_YEAR=="2014")),
            length(which(BR_INT_STX_METRO$STX_YEAR=="2015")),
            length(which(BR_INT_STX_METRO$STX_YEAR=="2016")),
            length(which(BR_INT_STX_METRO$STX_YEAR=="2017")),

            length(which(BR_INT_STX_OTHER$STX_YEAR=="2011")),
            length(which(BR_INT_STX_OTHER$STX_YEAR=="2012")),
            length(which(BR_INT_STX_OTHER$STX_YEAR=="2013")),
            length(which(BR_INT_STX_OTHER$STX_YEAR=="2014")),
            length(which(BR_INT_STX_OTHER$STX_YEAR=="2015")),
            length(which(BR_INT_STX_OTHER$STX_YEAR=="2016")),
            length(which(BR_INT_STX_OTHER$STX_YEAR=="2017")))


Year <- c('2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017')

STX_YEAR <- data.frame (Regions, Numbers, Year)

STX_YEAR$Regions <- relevel(STX_YEAR$Regions, ref='Seoul')



str(STX_YEAR)

dev.off()

ggplot(data=STX_YEAR, aes(x=Year, y=Numbers, group=Regions, color=Regions))+
  geom_line(size=0.8)+
  xlab("Year of initial breast cancer treatment")+
  ylim(500,10000) +
  scale_color_discrete("Regions") +
  theme(legend.position = c(0.15, 0.835),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15),
      axis.text.x = element_text(size=12),
      axis.text.y = element_text(size=12)
  )
```

```
### SG

Regions <- c ("Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul",
        "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan",
"Metropolitan", "Metropolitan",
        "Others", "Others", "Others", "Others", "Others", "Others", "Others")


Numbers   <- c(length(which(BR_INT_SG_SEOUL$SG_YEAR=="2011")),
         length(which(BR_INT_SG_SEOUL$SG_YEAR=="2012")),
         length(which(BR_INT_SG_SEOUL$SG_YEAR=="2013")),
         length(which(BR_INT_SG_SEOUL$SG_YEAR=="2014")),
         length(which(BR_INT_SG_SEOUL$SG_YEAR=="2015")),
         length(which(BR_INT_SG_SEOUL$SG_YEAR=="2016")),
         length(which(BR_INT_SG_SEOUL$SG_YEAR=="2017")),

         length(which(BR_INT_SG_METRO$SG_YEAR=="2011")),
         length(which(BR_INT_SG_METRO$SG_YEAR=="2012")),
         length(which(BR_INT_SG_METRO$SG_YEAR=="2013")),
         length(which(BR_INT_SG_METRO$SG_YEAR=="2014")),
         length(which(BR_INT_SG_METRO$SG_YEAR=="2015")),
         length(which(BR_INT_SG_METRO$SG_YEAR=="2016")),
         length(which(BR_INT_SG_METRO$SG_YEAR=="2017")),

         length(which(BR_INT_SG_OTHER$SG_YEAR=="2011")),
         length(which(BR_INT_SG_OTHER$SG_YEAR=="2012")),
         length(which(BR_INT_SG_OTHER$SG_YEAR=="2013")),
         length(which(BR_INT_SG_OTHER$SG_YEAR=="2014")),
         length(which(BR_INT_SG_OTHER$SG_YEAR=="2015")),
         length(which(BR_INT_SG_OTHER$SG_YEAR=="2016")),
         length(which(BR_INT_SG_OTHER$SG_YEAR=="2017")))


Year <- c('2011','2012','2013','2014','2015','2016','2017',
      '2011','2012','2013','2014','2015','2016','2017',
      '2011','2012','2013','2014','2015','2016','2017')


SG_YEAR <- data.frame (Regions, Numbers, Year)

SG_YEAR$Regions <- relevel(SG_YEAR$Regions, ref='Seoul')



str(SG_YEAR)

dev.off()

ggplot(data=SG_YEAR, aes(x=Year, y=Numbers, group=Regions, color=Regions))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer surgery")+
  ylim(500,10000) +
  scale_color_discrete("Regions") +
  theme(legend.position = c(0.15, 0.835),
     legend.title = element_text(size=13),
     legend.text = element_text(size=12),
     axis.title.x = element_text(size=15),
     axis.title.y = element_text(size=15),
     axis.text.x = element_text(size=12),
     axis.text.y = element_text(size=12))



### CTX

Regions <- c ("Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul",
        "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan",
"Metropolitan", "Metropolitan",
        "Others", "Others", "Others", "Others", "Others", "Others", "Others")


Numbers   <- c(length(which(BR_INT_CTX_SEOUL$CTX_YEAR=="2011")),
         length(which(BR_INT_CTX_SEOUL$CTX_YEAR=="2012")),
         length(which(BR_INT_CTX_SEOUL$CTX_YEAR=="2013")),
         length(which(BR_INT_CTX_SEOUL$CTX_YEAR=="2014")),
         length(which(BR_INT_CTX_SEOUL$CTX_YEAR=="2015")),
         length(which(BR_INT_CTX_SEOUL$CTX_YEAR=="2016")),
         length(which(BR_INT_CTX_SEOUL$CTX_YEAR=="2017")),

         length(which(BR_INT_CTX_METRO$CTX_YEAR=="2011")),
         length(which(BR_INT_CTX_METRO$CTX_YEAR=="2012")),
         length(which(BR_INT_CTX_METRO$CTX_YEAR=="2013")),
         length(which(BR_INT_CTX_METRO$CTX_YEAR=="2014")),
         length(which(BR_INT_CTX_METRO$CTX_YEAR=="2015")),
         length(which(BR_INT_CTX_METRO$CTX_YEAR=="2016")),
         length(which(BR_INT_CTX_METRO$CTX_YEAR=="2017")),

         length(which(BR_INT_CTX_OTHER$CTX_YEAR=="2011")),
         length(which(BR_INT_CTX_OTHER$CTX_YEAR=="2012")),
         length(which(BR_INT_CTX_OTHER$CTX_YEAR=="2013")),
         length(which(BR_INT_CTX_OTHER$CTX_YEAR=="2014")),
         length(which(BR_INT_CTX_OTHER$CTX_YEAR=="2015")),
         length(which(BR_INT_CTX_OTHER$CTX_YEAR=="2016")),
         length(which(BR_INT_CTX_OTHER$CTX_YEAR=="2017")))
Year <- c('2011','2012','2013','2014','2015','2016','2017',
```

```
        '2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017')

CTX_YEAR <- data.frame (Regions, Numbers, Year)

CTX_YEAR$Regions <- relevel(CTX_YEAR$Regions, ref='Seoul')

str(CTX_YEAR)

dev.off()

ggplot(data=CTX_YEAR, aes(x=Year, y=Numbers, group=Regions, color=Regions))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer chemotherapy")+
  ylim(500,10000) +
  scale_color_discrete("Regions") +
  theme(legend.position = c(0.15, 0.835),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15),
      axis.text.x = element_text(size=12),
      axis.text.y = element_text(size=12))




### RT

Regions <- c ("Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul", "Seoul",
           "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan", "Metropolitan",
"Metropolitan", "Metropolitan",
           "Others", "Others", "Others", "Others", "Others", "Others", "Others")


Numbers   <- c(length(which(BR_INT_RT_SEOUL$RT_YEAR=="2011")),
           length(which(BR_INT_RT_SEOUL$RT_YEAR=="2012")),
           length(which(BR_INT_RT_SEOUL$RT_YEAR=="2013")),
           length(which(BR_INT_RT_SEOUL$RT_YEAR=="2014")),
           length(which(BR_INT_RT_SEOUL$RT_YEAR=="2015")),
           length(which(BR_INT_RT_SEOUL$RT_YEAR=="2016")),
           length(which(BR_INT_RT_SEOUL$RT_YEAR=="2017")),

           length(which(BR_INT_RT_METRO$RT_YEAR=="2011")),
           length(which(BR_INT_RT_METRO$RT_YEAR=="2012")),
           length(which(BR_INT_RT_METRO$RT_YEAR=="2013")),
           length(which(BR_INT_RT_METRO$RT_YEAR=="2014")),
           length(which(BR_INT_RT_METRO$RT_YEAR=="2015")),
           length(which(BR_INT_RT_METRO$RT_YEAR=="2016")),
           length(which(BR_INT_RT_METRO$RT_YEAR=="2017")),

           length(which(BR_INT_RT_OTHER$RT_YEAR=="2011")),
           length(which(BR_INT_RT_OTHER$RT_YEAR=="2012")),
           length(which(BR_INT_RT_OTHER$RT_YEAR=="2013")),
           length(which(BR_INT_RT_OTHER$RT_YEAR=="2014")),
           length(which(BR_INT_RT_OTHER$RT_YEAR=="2015")),
           length(which(BR_INT_RT_OTHER$RT_YEAR=="2016")),
           length(which(BR_INT_RT_OTHER$RT_YEAR=="2017")))


Year <- c('2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017')

RT_YEAR <- data.frame (Regions, Numbers, Year)

RT_YEAR$Regions <- relevel(RT_YEAR$Regions, ref='Seoul')



str(RT_YEAR)

dev.off()

ggplot(data=RT_YEAR, aes(x=Year, y=Numbers, group=Regions, color=Regions))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer radiation therapy")+
  ylim(500,10000) +
  scale_color_discrete("Regions") +
  theme(legend.position = c(0.15, 0.835),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15),
      axis.text.x = element_text(size=12),
      axis.text.y = element_text(size=12))




########### Table 2. Poisson regression analysis

library(moonBook)
library(qcc)


### PE
```

```
PE_YEAR$Regions <- relevel(PE_YEAR$Regions, ref='Others')

hist(PE_YEAR$Numbers)
mean(PE_YEAR$Numbers)
var(PE_YEAR$Numbers)

qcc.overdispersion.test(PE_YEAR$Numbers, type="poisson")

poisson_PE <- glm(Numbers ~ Regions + Year, family=quasipoisson, data=PE_YEAR)

summary(poisson_PE)

extractOR(poisson_PE, digits=3)


### BC

BC_YEAR$Regions <- relevel(BC_YEAR$Regions, ref='Others')

hist(BC_YEAR$Numbers)
mean(BC_YEAR$Numbers)
var(BC_YEAR$Numbers)

qcc.overdispersion.test(BC_YEAR$Numbers, type="poisson")

poisson_BC <- glm(Numbers ~ Regions + Year, family=quasipoisson, data=BC_YEAR)

summary(poisson_BC)

extractOR(poisson_BC, digits=3)


### STX

STX_YEAR$Regions <- relevel(STX_YEAR$Regions, ref='Others')

hist(STX_YEAR$Numbers)
mean(STX_YEAR$Numbers)
var(STX_YEAR$Numbers)

qcc.overdispersion.test(STX_YEAR$Numbers, type="poisson")

poisson_STX <- glm(Numbers ~ Regions + Year, family=quasipoisson, data=STX_YEAR)

summary(poisson_STX)

extractOR(poisson_STX, digits=3)

### SG

SG_YEAR$Regions <- relevel(SG_YEAR$Regions, ref='Others')

hist(SG_YEAR$Numbers)
mean(SG_YEAR$Numbers)
var(SG_YEAR$Numbers)

qcc.overdispersion.test(SG_YEAR$Numbers, type="poisson")

poisson_SG <- glm(Numbers ~ Regions + Year, family=quasipoisson, data=SG_YEAR)

summary(poisson_SG)

extractOR(poisson_SG, digits=3)



### CTX

CTX_YEAR$Regions <- relevel(CTX_YEAR$Regions, ref='Others')

hist(CTX_YEAR$Numbers)
mean(CTX_YEAR$Numbers)
var(CTX_YEAR$Numbers)

qcc.overdispersion.test(CTX_YEAR$Numbers, type="poisson")

poisson_CTX <- glm(Numbers ~ Regions + Year, family=quasipoisson, data=CTX_YEAR)

summary(poisson_CTX)

extractOR(poisson_CTX, digits=3)



### RT


RT_YEAR$Regions <- relevel(RT_YEAR$Regions, ref='Others')

hist(RT_YEAR$Numbers)
mean(RT_YEAR$Numbers)
var(RT_YEAR$Numbers)
```

```
qcc.overdispersion.test(RT_YEAR$Numbers, type="poisson")

poisson_RT <- glm(Numbers ~ Regions + Year, family=quasipoisson, data=RT_YEAR)

summary(poisson_RT)

extractOR(poisson_RT, digits=3)




########### Odds ratios of quasi-Poisson regression analysis


ORplot(poisson_PE, type=2, show.CI=TRUE, main="Pathologic exam")

ORplot(poisson_BC, typC=2, show.CI=TRUE, main="Breast cancer diagnosis")

ORplot(poisson_STX, type=2, show.CI=TRUE, main="Initial breast cancer treatment")

ORplot(poisson_SG, type=2, show.CI=TRUE, main="Breast cancer surgery")

ORplot(poisson_CTX, type=2, show.CI=TRUE, main="Breast cancer chemotherapy")

ORplot(poisson_RT, type=2, show.CI=TRUE, main="Breast cancer radiation therapy")




########### Annual number of breast cancer patients by breast cancer care and hospital.


### PE

Hospital <- c ("Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary",
            "General","General","General","General","General","General","General",
            "Others","Others","Others","Others","Others","Others","Others")


Numbers   <- c(length(which(BR_INT_PE_TH$PE_YEAR=="2010")),
             length(which(BR_INT_PE_TH$PE_YEAR=="2011")),
             length(which(BR_INT_PE_TH$PE_YEAR=="2012")),
             length(which(BR_INT_PE_TH$PE_YEAR=="2013")),
             length(which(BR_INT_PE_TH$PE_YEAR=="2014")),
             length(which(BR_INT_PE_TH$PE_YEAR=="2015")),
             length(which(BR_INT_PE_TH$PE_YEAR=="2016")),
             length(which(BR_INT_PE_GH$PE_YEAR=="2010")),
             length(which(BR_INT_PE_GH$PE_YEAR=="2011")),
             length(which(BR_INT_PE_GH$PE_YEAR=="2012")),
             length(which(BR_INT_PE_GH$PE_YEAR=="2013")),
             length(which(BR_INT_PE_GH$PE_YEAR=="2014")),
             length(which(BR_INT_PE_GH$PE_YEAR=="2015")),
             length(which(BR_INT_PE_GH$PE_YEAR=="2016")),

             length(which(BR_INT_PE_OH$PE_YEAR=="2010")),
             length(which(BR_INT_PE_OH$PE_YEAR=="2011")),
             length(which(BR_INT_PE_OH$PE_YEAR=="2012")),
             length(which(BR_INT_PE_OH$PE_YEAR=="2013")),
             length(which(BR_INT_PE_OH$PE_YEAR=="2014")),
             length(which(BR_INT_PE_OH$PE_YEAR=="2015")),
             length(which(BR_INT_PE_OH$PE_YEAR=="2016")))



Year <- c('2010','2011','2012','2013','2014','2015','2016',
         '2010','2011','2012','2013','2014','2015','2016',
         '2010','2011','2012','2013','2014','2015','2016')

PE_YEAR <- data.frame (Hospital, Numbers, Year)

PE_YEAR$Hospital <- relevel(PE_YEAR$Hospital, ref='Tertiary')



str(PE_YEAR)
summary(PE_YEAR)

dev.off()

ggplot(data=PE_YEAR, aes(x=Year, y=Numbers, group=Hospital, color=Hospital))+
  geom_line(size=0.8)+
  xlab("Year of pathologic examination")+
  ylim(0,15000) +
  scale_color_discrete("Hospital") +
  theme(legend.position = c(0.15, 0.835),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15),
      axis.text.x = element_text(size=12),
      axis.text.y = element_text(size=12)
  )
```

```
### BC

Hospital <- c ("Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary",
              "General","General","General","General","General","General","General",
              "Others","Others","Others","Others","Others","Others","Others")


Numbers   <- c(length(which(BR_INT_BC_TH$BC_YEAR=="2011")),
             length(which(BR_INT_BC_TH$BC_YEAR=="2012")),
             length(which(BR_INT_BC_TH$BC_YEAR=="2013")),
             length(which(BR_INT_BC_TH$BC_YEAR=="2014")),
             length(which(BR_INT_BC_TH$BC_YEAR=="2015")),
             length(which(BR_INT_BC_TH$BC_YEAR=="2016")),
             length(which(BR_INT_BC_TH$BC_YEAR=="2017")),

             length(which(BR_INT_BC_GH$BC_YEAR=="2011")),
             length(which(BR_INT_BC_GH$BC_YEAR=="2012")),
             length(which(BR_INT_BC_GH$BC_YEAR=="2013")),
             length(which(BR_INT_BC_GH$BC_YEAR=="2014")),
             length(which(BR_INT_BC_GH$BC_YEAR=="2015")),
             length(which(BR_INT_BC_GH$BC_YEAR=="2016")),
             length(which(BR_INT_BC_GH$BC_YEAR=="2017")),

             length(which(BR_INT_BC_OH$BC_YEAR=="2011")),
             length(which(BR_INT_BC_OH$BC_YEAR=="2012")),
             length(which(BR_INT_BC_OH$BC_YEAR=="2013")),
             length(which(BR_INT_BC_OH$BC_YEAR=="2014")),
             length(which(BR_INT_BC_OH$BC_YEAR=="2015")),
             length(which(BR_INT_BC_OH$BC_YEAR=="2016")),
             length(which(BR_INT_BC_OH$BC_YEAR=="2017")))




Year <- c('2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017')

BC_YEAR <- data.frame (Hospital, Numbers, Year)

BC_YEAR$Hospital <- relevel(BC_YEAR$Hospital, ref='Tertiary')


str(BC_YEAR)

dev.off()

ggplot(data=BC_YEAR, aes(x=Year, y=Numbers, group=Hospital, color=Hospital))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer diagnosis")+
  ylim(0,15000) +
  scale_color_discrete("Hospital") +
  theme(legend.position = c(0.15, 0.835),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15),
      axis.text.x = element_text(size=12),
      axis.text.y = element_text(size=12)
  )



### STX

Hospital <- c ("Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary",
              "General","General","General","General","General","General","General",
              "Others","Others","Others","Others","Others","Others","Others")


Numbers   <- c(length(which(BR_INT_STX_TH$STX_YEAR=="2011")),
             length(which(BR_INT_STX_TH$STX_YEAR=="2012")),
             length(which(BR_INT_STX_TH$STX_YEAR=="2013")),
             length(which(BR_INT_STX_TH$STX_YEAR=="2014")),
             length(which(BR_INT_STX_TH$STX_YEAR=="2015")),
             length(which(BR_INT_STX_TH$STX_YEAR=="2016")),
             length(which(BR_INT_STX_TH$STX_YEAR=="2017")),

             length(which(BR_INT_STX_GH$STX_YEAR=="2011")),
             length(which(BR_INT_STX_GH$STX_YEAR=="2012")),
             length(which(BR_INT_STX_GH$STX_YEAR=="2013")),
             length(which(BR_INT_STX_GH$STX_YEAR=="2014")),
             length(which(BR_INT_STX_GH$STX_YEAR=="2015")),
             length(which(BR_INT_STX_GH$STX_YEAR=="2016")),
             length(which(BR_INT_STX_GH$STX_YEAR=="2017")),

             length(which(BR_INT_STX_OH$STX_YEAR=="2011")),
             length(which(BR_INT_STX_OH$STX_YEAR=="2012")),
             length(which(BR_INT_STX_OH$STX_YEAR=="2013")),
             length(which(BR_INT_STX_OH$STX_YEAR=="2014")),
             length(which(BR_INT_STX_OH$STX_YEAR=="2015")),
             length(which(BR_INT_STX_OH$STX_YEAR=="2016")),
             length(which(BR_INT_STX_OH$STX_YEAR=="2017")))
```

```
Year <- c('2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017')

STX_YEAR <- data.frame (Hospital, Numbers, Year)

STX_YEAR$Hospital <- relevel(STX_YEAR$Hospital, ref='Tertiary')



str(STX_YEAR)

dev.off()

ggplot(data=STX_YEAR, aes(x=Year, y=Numbers, group=Hospital, color=Hospital))+
  geom_line(size=0.8)+
  xlab("Year of initial breast cancer treatment")+
  ylim(0,15000) +
  scale_color_discrete("Hospital") +
  theme(legend.position = c(0.15, 0.835),
        legend.title = element_text(size=13),
        legend.text = element_text(size=12),
        axis.title.x = element_text(size=15),
        axis.title.y = element_text(size=15),
        axis.text.x = element_text(size=12),
        axis.text.y = element_text(size=12)
  )



### SG

Hospital <- c ("Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary",
               "General","General","General","General","General","General","General",
               "Others","Others","Others","Others","Others","Others","Others")


Numbers   <- c(length(which(BR_INT_SG_TH$SG_YEAR=="2011")),
               length(which(BR_INT_SG_TH$SG_YEAR=="2012")),
               length(which(BR_INT_SG_TH$SG_YEAR=="2013")),
               length(which(BR_INT_SG_TH$SG_YEAR=="2014")),
               length(which(BR_INT_SG_TH$SG_YEAR=="2015")),
               length(which(BR_INT_SG_TH$SG_YEAR=="2016")),
               length(which(BR_INT_SG_TH$SG_YEAR=="2017")),
               length(which(BR_INT_SG_GH$SG_YEAR=="2011")),
               length(which(BR_INT_SG_GH$SG_YEAR=="2012")),
               length(which(BR_INT_SG_GH$SG_YEAR=="2013")),
               length(which(BR_INT_SG_GH$SG_YEAR=="2014")),
               length(which(BR_INT_SG_GH$SG_YEAR=="2015")),
               length(which(BR_INT_SG_GH$SG_YEAR=="2016")),
               length(which(BR_INT_SG_GH$SG_YEAR=="2017")),

               length(which(BR_INT_SG_OH$SG_YEAR=="2011")),
               length(which(BR_INT_SG_OH$SG_YEAR=="2012")),
               length(which(BR_INT_SG_OH$SG_YEAR=="2013")),
               length(which(BR_INT_SG_OH$SG_YEAR=="2014")),
               length(which(BR_INT_SG_OH$SG_YEAR=="2015")),
               length(which(BR_INT_SG_OH$SG_YEAR=="2016")),
               length(which(BR_INT_SG_OH$SG_YEAR=="2017")))




Year <- c('2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017')


SG_YEAR <- data.frame (Hospital, Numbers, Year)

SG_YEAR$Hospital <- relevel(SG_YEAR$Hospital, ref='Tertiary')



str(SG_YEAR)

dev.off()

ggplot(data=SG_YEAR, aes(x=Year, y=Numbers, group=Hospital, color=Hospital))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer surgery")+
  ylim(0,15000) +
  scale_color_discrete("Hospital") +
  theme(legend.position = c(0.15, 0.835),
        legend.title = element_text(size=13),
        legend.text = element_text(size=12),
        axis.title.x = element_text(size=15),
        axis.title.y = element_text(size=15),
        axis.text.x = element_text(size=12),
        axis.text.y = element_text(size=12))
```

```
### CTX

Hospital <- c ("Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary",
            "General","General","General","General","General","General","General",
            "Others","Others","Others","Others","Others","Others","Others")


Numbers   <- c(length(which(BR_INT_CTX_TH$CTX_YEAR=="2011")),
            length(which(BR_INT_CTX_TH$CTX_YEAR=="2012")),
            length(which(BR_INT_CTX_TH$CTX_YEAR=="2013")),
            length(which(BR_INT_CTX_TH$CTX_YEAR=="2014")),
            length(which(BR_INT_CTX_TH$CTX_YEAR=="2015")),
            length(which(BR_INT_CTX_TH$CTX_YEAR=="2016")),
            length(which(BR_INT_CTX_TH$CTX_YEAR=="2017")),

            length(which(BR_INT_CTX_GH$CTX_YEAR=="2011")),
            length(which(BR_INT_CTX_GH$CTX_YEAR=="2012")),
            length(which(BR_INT_CTX_GH$CTX_YEAR=="2013")),
            length(which(BR_INT_CTX_GH$CTX_YEAR=="2014")),
            length(which(BR_INT_CTX_GH$CTX_YEAR=="2015")),
            length(which(BR_INT_CTX_GH$CTX_YEAR=="2016")),
            length(which(BR_INT_CTX_GH$CTX_YEAR=="2017")),

            length(which(BR_INT_CTX_OH$CTX_YEAR=="2011")),
            length(which(BR_INT_CTX_OH$CTX_YEAR=="2012")),
            length(which(BR_INT_CTX_OH$CTX_YEAR=="2013")),
            length(which(BR_INT_CTX_OH$CTX_YEAR=="2014")),
            length(which(BR_INT_CTX_OH$CTX_YEAR=="2015")),
            length(which(BR_INT_CTX_OH$CTX_YEAR=="2016")),
            length(which(BR_INT_CTX_OH$CTX_YEAR=="2017")))



Year <- c('2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017',
        '2011','2012','2013','2014','2015','2016','2017')

CTX_YEAR <- data.frame (Hospital, Numbers, Year)

CTX_YEAR$Hospital <- relevel(CTX_YEAR$Hospital, ref='Tertiary')



str(CTX_YEAR)

dev.off()

ggplot(data=CTX_YEAR, aes(x=Year, y=Numbers, group=Hospital, color=Hospital))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer chemotherapy")+
  ylim(0,15000) +
  scale_color_discrete("Hospital") +
  theme(legend.position = c(0.15, 0.835),
      legend.title = element_text(size=13),
      legend.text = element_text(size=12),
      axis.title.x = element_text(size=15),
      axis.title.y = element_text(size=15),
      axis.text.x = element_text(size=12),
      axis.text.y = element_text(size=12))



### RT

Hospital <- c ("Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary","Tertiary",
            "General","General","General","General","General","General","General",
            "Others","Others","Others","Others","Others","Others","Others")


Numbers   <- c(length(which(BR_INT_RT_TH$RT_YEAR=="2011")),
            length(which(BR_INT_RT_TH$RT_YEAR=="2012")),
            length(which(BR_INT_RT_TH$RT_YEAR=="2013")),
            length(which(BR_INT_RT_TH$RT_YEAR=="2014")),
            length(which(BR_INT_RT_TH$RT_YEAR=="2015")),
            length(which(BR_INT_RT_TH$RT_YEAR=="2016")),
            length(which(BR_INT_RT_TH$RT_YEAR=="2017")),

            length(which(BR_INT_RT_GH$RT_YEAR=="2011")),
            length(which(BR_INT_RT_GH$RT_YEAR=="2012")),
            length(which(BR_INT_RT_GH$RT_YEAR=="2013")),
            length(which(BR_INT_RT_GH$RT_YEAR=="2014")),
            length(which(BR_INT_RT_GH$RT_YEAR=="2015")),
            length(which(BR_INT_RT_GH$RT_YEAR=="2016")),
            length(which(BR_INT_RT_GH$RT_YEAR=="2017")),

            length(which(BR_INT_RT_OH$RT_YEAR=="2011")),
            length(which(BR_INT_RT_OH$RT_YEAR=="2012")),
            length(which(BR_INT_RT_OH$RT_YEAR=="2013")),
            length(which(BR_INT_RT_OH$RT_YEAR=="2014")),
            length(which(BR_INT_RT_OH$RT_YEAR=="2015")),
            length(which(BR_INT_RT_OH$RT_YEAR=="2016")),
            length(which(BR_INT_RT_OH$RT_YEAR=="2017")))
```

```r
Year <- c('2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017',
          '2011','2012','2013','2014','2015','2016','2017')

RT_YEAR <- data.frame (Hospital, Numbers, Year)

RT_YEAR$Hospital <- relevel(RT_YEAR$Hospital, ref='Tertiary')



str(RT_YEAR)

dev.off()

ggplot(data=RT_YEAR, aes(x=Year, y=Numbers, group=Hospital, color=Hospital))+
  geom_line(size=0.8)+
  xlab("Year of breast cancer radiation therapy")+
  ylim(0,15000) +
  scale_color_discrete("Hospital") +
  theme(legend.position = c(0.15, 0.835),
        legend.title = element_text(size=13),
        legend.text = element_text(size=12),
        axis.title.x = element_text(size=15),
        axis.title.y = element_text(size=15),
        axis.text.x = element_text(size=12),
        axis.text.y = element_text(size=12))





########### Annual trends of breast cancer care by hospital using quasi-Poisson regression analysis

library(moonBook)
library(qcc)


### PE

PE_YEAR$Hospital <- relevel(PE_YEAR$Hospital, ref='Others')

hist(PE_YEAR$Numbers)
mean(PE_YEAR$Numbers)
var(PE_YEAR$Numbers)

qcc.overdispersion.test(PE_YEAR$Numbers, type="poisson")

poisson_PE <- glm(Numbers ~ Hospital + Year, family=quasipoisson, data=PE_YEAR)

summary(poisson_PE)

extractOR(poisson_PE, digits=3)


### BC

BC_YEAR$Hospital <- relevel(BC_YEAR$Hospital, ref='Others')

hist(BC_YEAR$Numbers)
mean(BC_YEAR$Numbers)
var(BC_YEAR$Numbers)

qcc.overdispersion.test(BC_YEAR$Numbers, type="poisson")

poisson_BC <- glm(Numbers ~ Hospital + Year, family=quasipoisson, data=BC_YEAR)

summary(poisson_BC)

extractOR(poisson_BC, digits=3)


### STX

STX_YEAR$Hospital <- relevel(STX_YEAR$Hospital, ref='Others')

hist(STX_YEAR$Numbers)
mean(STX_YEAR$Numbers)
var(STX_YEAR$Numbers)

qcc.overdispersion.test(STX_YEAR$Numbers, type="poisson")

poisson_STX <- glm(Numbers ~ Hospital + Year, family=quasipoisson, data=STX_YEAR)

summary(poisson_STX)

extractOR(poisson_STX, digits=3)
```

```
### SG

SG_YEAR$Hospital <- relevel(SG_YEAR$Hospital, ref='Others')

hist(SG_YEAR$Numbers)
mean(SG_YEAR$Numbers)
var(SG_YEAR$Numbers)

qcc.overdispersion.test(SG_YEAR$Numbers, type="poisson")

poisson_SG <- glm(Numbers ~ Hospital + Year, family=quasipoisson, data=SG_YEAR)

summary(poisson_SG)

extractOR(poisson_SG, digits=3)



### CTX

CTX_YEAR$Hospital <- relevel(CTX_YEAR$Hospital, ref='Others')

hist(CTX_YEAR$Numbers)
mean(CTX_YEAR$Numbers)
var(CTX_YEAR$Numbers)

qcc.overdispersion.test(CTX_YEAR$Numbers, type="poisson")

poisson_CTX <- glm(Numbers ~ Hospital + Year, family=quasipoisson, data=CTX_YEAR)

summary(poisson_CTX)

extractOR(poisson_CTX, digits=3)



### RT


RT_YEAR$Hospital <- relevel(RT_YEAR$Hospital, ref='Others')

hist(RT_YEAR$Numbers)
mean(RT_YEAR$Numbers)
var(RT_YEAR$Numbers)

qcc.overdispersion.test(RT_YEAR$Numbers, type="poisson")

poisson_RT <- glm(Numbers ~ Hospital + Year, family=quasipoisson, data=RT_YEAR)

summary(poisson_RT)

extractOR(poisson_RT, digits=3)



########### Odds ratios of quasi-Poisson regression analysis


ORplot(poisson_PE, type=2, show.CI=TRUE, main="Pathologic exam")

ORplot(poisson_BC, typC=2, show.CI=TRUE, main="Breast cancer diagnosis")

ORplot(poisson_STX, type=2, show.CI=TRUE, main="Initial breast cancer treatment")

ORplot(poisson_SG, type=2, show.CI=TRUE, main="Breast cancer surgery")

ORplot(poisson_CTX, type=2, show.CI=TRUE, main="Breast cancer chemotherapy")

ORplot(poisson_RT, type=2, show.CI=TRUE, main="Breast cancer radiation therapy")
```

```
##### PE region, hospital, depar


ftable(PE_Region_cat1)
100*prop.table(table(PE_Region_cat1))

ftable(PE_Region_cat2)
100*prop.table(table(PE_Region_cat2))

ftable(PE_Region_cat3)
100*prop.table(table(PE_Region_cat3))

ftable(PE_Hospital_cat1)
100*prop.table(table(PE_Hospital_cat1))

ftable(PE_Hospital_cat2)
100*prop.table(table(PE_Hospital_cat2))

ftable(PE_Hospital_cat3)
100*prop.table(table(PE_Hospital_cat3))

ftable(PE_Hospital_cat4)
100*prop.table(table(PE_Hospital_cat4))

ftable(PE_Depart_cat1)
100*prop.table(table(PE_Depart_cat1))



##### BC region, hospital, depar


ftable(BC_Region_cat1)
100*prop.table(table(BC_Region_cat1))

ftable(BC_Region_cat2)
100*prop.table(table(BC_Region_cat2))

ftable(BC_Region_cat3)
100*prop.table(table(BC_Region_cat3))

ftable(BC_Hospital_cat1)
100*prop.table(table(BC_Hospital_cat1))

ftable(BC_Hospital_cat2)
100*prop.table(table(BC_Hospital_cat2))

ftable(BC_Hospital_cat3)
100*prop.table(table(BC_Hospital_cat3))

ftable(BC_Hospital_cat4)
100*prop.table(table(BC_Hospital_cat4))

ftable(BC_Depart_cat1)
100*prop.table(table(BC_Depart_cat1))



##### SG region, hospital, depar


ftable(SG_Region_cat1)
100*prop.table(table(SG_Region_cat1))

ftable(SG_Region_cat2)
100*prop.table(table(SG_Region_cat2))

ftable(SG_Region_cat3)
100*prop.table(table(SG_Region_cat3))

ftable(SG_Hospital_cat1)
100*prop.table(table(SG_Hospital_cat1))

ftable(SG_Hospital_cat2)
100*prop.table(table(SG_Hospital_cat2))

ftable(SG_Hospital_cat3)
100*prop.table(table(SG_Hospital_cat3))

ftable(SG_Hospital_cat4)
100*prop.table(table(SG_Hospital_cat4))

ftable(SG_Depart_cat1)
100*prop.table(table(SG_Depart_cat1))



##### CTX region, hospital, depar


ftable(CTX_Region_cat1)
100*prop.table(table(CTX_Region_cat1))
```

```
ftable(CTX_Region_cat2)
100*prop.table(table(CTX_Region_cat2))

ftable(CTX_Region_cat3)
100*prop.table(table(CTX_Region_cat3))

ftable(CTX_Hospital_cat1)
100*prop.table(table(CTX_Hospital_cat1))

ftable(CTX_Hospital_cat2)
100*prop.table(table(CTX_Hospital_cat2))

ftable(CTX_Hospital_cat3)
100*prop.table(table(CTX_Hospital_cat3))

ftable(CTX_Hospital_cat4)
100*prop.table(table(CTX_Hospital_cat4))

ftable(CTX_Depart_cat1)
100*prop.table(table(CTX_Depart_cat1))



##### STX region, hospital, depar


ftable(STX_Region_cat1)
100*prop.table(table(STX_Region_cat1))

ftable(STX_Region_cat2)
100*prop.table(table(STX_Region_cat2))

ftable(STX_Region_cat3)
100*prop.table(table(STX_Region_cat3))

ftable(STX_Hospital_cat1)
100*prop.table(table(STX_Hospital_cat1))

ftable(STX_Hospital_cat2)
100*prop.table(table(STX_Hospital_cat2))

ftable(STX_Hospital_cat3)
100*prop.table(table(STX_Hospital_cat3))

ftable(STX_Hospital_cat4)
100*prop.table(table(STX_Hospital_cat4))

ftable(STX_Depart_cat1)
100*prop.table(table(STX_Depart_cat1))



##### RT region, hospital, depar


ftable(RT_Region_cat1)
100*prop.table(table(RT_Region_cat1))

ftable(RT_Region_cat2)
100*prop.table(table(RT_Region_cat2))

ftable(RT_Region_cat3)
100*prop.table(table(RT_Region_cat3))

ftable(RT_Hospital_cat1)
100*prop.table(table(RT_Hospital_cat1))

ftable(RT_Hospital_cat2)
100*prop.table(table(RT_Hospital_cat2))

ftable(RT_Hospital_cat3)
100*prop.table(table(RT_Hospital_cat3))

ftable(RT_Hospital_cat4)
100*prop.table(table(RT_Hospital_cat4))

ftable(RT_Depart_cat1)
100*prop.table(table(RT_Depart_cat1))




BR_INT_SEOUL <- BR_INT %>% filter(PE_Region_cat2 == 'Seoul' & BC_Region_cat2 == 'Seoul'
& STX_Region_cat2 =='Seoul')
BR_INT_Metro <- BR_INT %>% filter(PE_Region_cat2 == 'Metropolitan city' & BC_Region_cat2
== 'Metropolitan city' & STX_Region_cat2 == 'Metropolitan city')
BR_INT_Others <- BR_INT %>% filter(PE_Region_cat2 == 'Others' & BC_Region_cat2 ==
'Others' & STX_Region_cat2 == 'Others')


##### Median F/U
summarise(count=n(), BR_INT_SEOUL,
        PE_BC_DAY_mean=mean(PE_BC_DAY,na.rm=TRUE),
        PE_BC_DAY_sd=sd(PE_BC_DAY, na.rm=TRUE),
```

```
        PE_BC_DAY_median=median(PE_BC_DAY,na.rm=TRUE),
        PE_STX_DAY_mean=mean(PE_STX_DAY,na.rm=TRUE),
        PE_STX_DAY_sd=sd(PE_STX_DAY, na.rm=TRUE),
        PE_STX_DAY_median=median(PE_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_mean=mean(BC_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_sd=sd(BC_STX_DAY, na.rm=TRUE),
        BC_STX_DAY_median=median(BC_STX_DAY,na.rm=TRUE))


##### Median F/U
summarise(count=n(), BR_INT_Metro,
        PE_BC_DAY_mean=mean(PE_BC_DAY,na.rm=TRUE),
        PE_BC_DAY_sd=sd(PE_BC_DAY, na.rm=TRUE),
        PE_BC_DAY_median=median(PE_BC_DAY,na.rm=TRUE),
        PE_STX_DAY_mean=mean(PE_STX_DAY,na.rm=TRUE),
        PE_STX_DAY_sd=sd(PE_STX_DAY, na.rm=TRUE),
        PE_STX_DAY_median=median(PE_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_mean=mean(BC_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_sd=sd(BC_STX_DAY, na.rm=TRUE),
        BC_STX_DAY_median=median(BC_STX_DAY,na.rm=TRUE))


##### Median F/U
summarise(count=n(), BR_INT_Others,
        PE_BC_DAY_mean=mean(PE_BC_DAY,na.rm=TRUE),
        PE_BC_DAY_sd=sd(PE_BC_DAY, na.rm=TRUE),
        PE_BC_DAY_median=median(PE_BC_DAY,na.rm=TRUE),
        PE_STX_DAY_mean=mean(PE_STX_DAY,na.rm=TRUE),
        PE_STX_DAY_sd=sd(PE_STX_DAY, na.rm=TRUE),
        PE_STX_DAY_median=median(PE_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_mean=mean(BC_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_sd=sd(BC_STX_DAY, na.rm=TRUE),
        BC_STX_DAY_median=median(BC_STX_DAY,na.rm=TRUE))

BR_INT_SSS <- BR_INT %>% filter(PE_Region_cat2 == 'Seoul' & BC_Region_cat2 == 'Seoul' &
STX_Region_cat2 =='Seoul')
BR_INT_MMM <- BR_INT %>% filter(PE_Region_cat2 == 'Metropolitan city' & BC_Region_cat2
== 'Metropolitan city' & STX_Region_cat2 == 'Metropolitan city')
BR_INT_OOO <- BR_INT %>% filter(PE_Region_cat2 == 'Others' & BC_Region_cat2 == 'Others'
& STX_Region_cat2 == 'Others')


BR_INT_S <- BR_INT %>% filter(PE_Region_cat2 == 'Seoul')
BR_INT_SS <- BR_INT %>% filter(PE_Region_cat2 == 'Seoul' & BC_Region_cat2 == 'Seoul')
BR_INT_SM <- BR_INT %>% filter(PE_Region_cat2 == 'Seoul' & BC_Region_cat2 ==
'Metropolitan city')
BR_INT_SO <- BR_INT %>% filter(PE_Region_cat2 == 'Seoul' & BC_Region_cat2 == 'Others')

summary(BR_INT_S)
summary(BR_INT_SS)
summary(BR_INT_SM)
summary(BR_INT_SO)


BR_INT_M <- BR_INT %>% filter(PE_Region_cat2 == 'Metropolitan city')
BR_INT_MS <- BR_INT %>% filter(PE_Region_cat2 == 'Metropolitan city' & BC_Region_cat2 ==
'Seoul')
BR_INT_MM <- BR_INT %>% filter(PE_Region_cat2 == 'Metropolitan city' & BC_Region_cat2
== 'Metropolitan city')
BR_INT_MO <- BR_INT %>% filter(PE_Region_cat2 == 'Metropolitan city' & BC_Region_cat2
== 'Others')

summary(BR_INT_M)
summary(BR_INT_MS)
summary(BR_INT_MM)
summary(BR_INT_MO)


BR_INT_O <- BR_INT %>% filter(PE_Region_cat2 == 'Others')
BR_INT_OS <- BR_INT %>% filter(PE_Region_cat2 == 'Others' & BC_Region_cat2 == 'Seoul')
BR_INT_OM <- BR_INT %>% filter(PE_Region_cat2 == 'Others' & BC_Region_cat2 ==
'Metropolitan city')
BR_INT_OO <- BR_INT %>% filter(PE_Region_cat2 == 'Others' & BC_Region_cat2 == 'Others')

summary(BR_INT_O)
summary(BR_INT_OS)
summary(BR_INT_OM)
summary(BR_INT_OO)


BR_INT_PE_SEOUL <- BR_INT %>% filter(PE_Region_cat1 == 'Seoul')
BR_INT_PE_Other <- BR_INT %>% filter(PE_Region_cat1 == 'Others')


str(BR_INT_PE_SEOUL)


##### Median F/U
summarise(count=n(), BR_INT_PE_SEOUL,
        PE_BC_DAY_mean=mean(PE_BC_DAY,na.rm=TRUE),
        PE_BC_DAY_sd=sd(PE_BC_DAY, na.rm=TRUE),
        PE_BC_DAY_median=median(PE_BC_DAY,na.rm=TRUE),
```

```
        PE_STX_DAY_mean=mean(PE_STX_DAY,na.rm=TRUE),
        PE_STX_DAY_sd=sd(PE_STX_DAY, na.rm=TRUE),
        PE_STX_DAY_median=median(PE_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_mean=mean(BC_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_sd=sd(BC_STX_DAY, na.rm=TRUE),
        BC_STX_DAY_median=median(BC_STX_DAY,na.rm=TRUE))

summarise(count=n(), BR_INT_PE_Other,
        PE_BC_DAY_mean=mean(PE_BC_DAY,na.rm=TRUE),
        PE_BC_DAY_sd=sd(PE_BC_DAY, na.rm=TRUE),
        PE_BC_DAY_median=median(PE_BC_DAY,na.rm=TRUE),
        PE_STX_DAY_mean=mean(PE_STX_DAY,na.rm=TRUE),
        PE_STX_DAY_sd=sd(PE_STX_DAY, na.rm=TRUE),
        PE_STX_DAY_median=median(PE_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_mean=mean(BC_STX_DAY,na.rm=TRUE),
        BC_STX_DAY_sd=sd(BC_STX_DAY, na.rm=TRUE),
        BC_STX_DAY_median=median(BC_STX_DAY,na.rm=TRUE))
```

국문 초록

배경 및 목적

한국에서는 암에 대한 치료를 위해 서울 이외의 지역에서 서울로 국내 의료 여행이 집중화 되고 있는 양상을 보인다. 하지만 국내 유방암 환자의 이러한 의료 여행의 동향에 대해서는 알려진 바가 거의 없다. 그래서 지역별 유방암 치료의 연간 빈도와 패턴을 조사하고, 초기 치료를 위해 서울 이외의 지역에서 서울로 국내 의료 여행 동향을 평가하고, 그리고 이런 유방암 환자의 의료 여행과 관련된 요인을 확인하고자 연구를 계획하였다.

연구 방법

한국 건강보험 심사평가원 청구 자료를 이용하여 전국적인 후향적 코호트 연구를 수행하였다. 환자들은 병리학적 검사를 받은 지역에 따라 분류하였다(서울 vs. 광역시 vs. 그 이외의 지역). 병리학적 검사, 진단 및 치료의 빈도를 지역별로 분석하였다. 국내 의료 여행을 시·도별로 분석하고, 의료 여행과 관련된 요인을 조사하였다.

결과

2010 년 1 월부터 2017 년 12 월 사이에 유방암 진단을 받은 총 150,709 명의 환자가 연구 분석에 포함되었다. 병리학적 검사, 진단 및 치료의 연간 빈도는 서울을 포함한 모든 지역에서 증가했으며, 특히 서울과 비 서울 지역 간의 수술 빈도 차이는 시간이 지남에 따라 증가하였다. 서울 이외의 지역에서 서울로의 의료 여행 비율은 2010 년 14.2%(1,161/8,150)에서 2017 년 19.8%(2,762/13,964)로 증가 소견을 보였다. 서울 및 대도시 이외의 지역 환자 중 약 4 분의 1 이 치료를 위해서 서울로 이동하였으며, 특히 2017 년에는 충북, 경북, 제주의 환자들의 40% 이상이 치료를 위해 서울의 의료 여행을 하였다. 40 세 이하의 젊은 연령과 광역시 이외의 지역 거주가 서울로의 국내 의료 여행이 높은 것과 유의한 연관성을 보였다. 반대로 의료급여 환자 이거나 다른 동반 질환이 있는 경우에는 서울로의 의료 여행이 유의하게 낮은 경향을 보였다.

결론

유방암 치료를 위한 서울로의 국내 의료 여행을 한 유방암 환자의 수와 비율은 치료 방법과 지역에 따른 차이가 있었지만 시간이 지남에 따라 증가하였다. 젊은 연령, 동반 질환, 보험 유형이 의료 여행과 관련된 중요한 요인이었다.