

## 목차

◆ 1. 유방암 임상 라이브러리에 요구되는 Table과 Column 정보 .....	2p
◆ 2. TABLE COLUMN 항목 수집 상태 .....	3p
- 환자대체번호 생성 과정 .....	12p
◆ 3. 유방암 라이브러리 구축 과정	
- 1) 암환자 등록 리스트 CTE .....	13p
- 2) 유방암_환자_건강정보 (BRCN_PT_HLINF) .....	13p
- 3) 유방암_환자_가족력 (BRCN_PT_FMHT) .....	14p
- 4) 유방암_환자_산과정보 (BRCN_PT_OBINF) .....	15p
- 5) 유방암_진단_신체계측정보 (BRCN_DIAG_ANINF) .....	16p
- 6) 유방암_진단_정보 (BRCN_DIAG_INF) .....	17p
- 7) 유방암_검사_영상 (BRCN_EXAM_IMAG) .....	18p
- 8) 유방암_검사_생식세포변이검사 (BRCN_EXAM_GMVX) .....	21p
- 9) 유방암_검사_유전자발현검사 (BRCN_EXAM_GNRX) .....	24p
- 10) 유방암_치료_수술 (BRCN_TRTM_OPRT) .....	25p
- 11) 유방암_검사_외과병리 (BRCN_EXAM_SGPT) .....	29p
- 12) 유방암_치료_약제성분 (BRCN_TRTM_DRIN) .....	35p
- 13) 유방암_치료_항암요법 (BRCN_TRTM_ANTP) .....	36p

## 1. 유방암 임상 라이브러리에 요구되는 Table과 Column 정보.

TABLE 영문명	TABLE 한문명	총 Column 수	수집 가능(동의) 수
BRCN_PT_HLINF	환자_건강정보	57	52
BRCN_PT_FMHT	환자_가족력	26	26
BRCN_PT_OBINF	환자_산과정보	26	17
BRCN_DIAG_ANINF	진단_신체계측정보	12	12
BRCN_DIAG_INF	진단_정보	13	13
BRCN_DIAG_MTST	진단_전이	12	0
BRCN_DIAG_STAG	진단_병기	12	11
BRCN_EXAM_DIAG	검사_진단	18	16
BRCN_EXAM_IMAG	검사_영상	24	24
BRCN_EXAM_BYPS	검사_생체	19	19
BRCN_EXAM_IMPT	검사_면역병리	15	15
BRCN_EXAM_MLPT	검사_분자병리	17	17
BRCN_EXAM_GMVX	검사_생식세포변이	27	27
BRCN_EXAM_GNRX	검사_유전자발현	10	10
BRCN_TRTM_OPRT	치료_수술	42	38
BRCN_TRTM_BLTR	치료_수혈	12	12
BRCN_EXAM_SGPT	검사_외과병리	53	52
BRCN_DIAG_AFOC	진단_수술후합병증	9	0
BRCN_TRTM_DRIN	치료_약제성분	29	27
BRCN_TRTM_ANTP	치료_항암요법	22	13
BRCN_TRTM_RD	치료_방사선	29	23
BRCN_FUOR_RLPS	추적관찰_재발	14	0
	테이블 총 22개	498	424

• Table 데이터(Row) 적재 상태 확인 (마지막 수정 : 2023-12-21)

Table	용량(Kb)	적재 Row 수	Table	용량(Kb)	적재 Row 수
BRCN_PT_HLINF	8320	5705	BRCN_EXAM_MLPT	2944	342
BRCN_PT_FMHT	4352	5702	BRCN_EXAM_GMVX	4224	11
BRCN_PT_OBINF	4224	1030	BRCN_EXAM_GNRX	2048	2
BRCN_DIAG_ANINF	2432	4979	BRCN_TRTM_OPRT	6272	2069
BRCN_DIAG_INF	2560	2774	BRCN_TRTM_BLTR	2304	56
BRCN_DIAG_MTST	640	0	BRCN_EXAM_SGPT	7552	986
BRCN_DIAG_STAG	2304	38	BRCN_DIAG_AFOC	640	0
BRCN_EXAM_DIAG	114432	807768	BRCN_TRTM_DRIN	12288	32041
BRCN_EXAM_IMAG	23936	34541	BRCN_TRTM_ANTP	3584	578
BRCN_EXAM_BYPS	3456	2364	BRCN_TRTM_RD	4608	1068
BRCN_EXAM_IMPT	3072	9707	BRCN_FUOR_RLPS	640	0
Column 합계	169728	874608	Column 합계	47104	37153

- 용량 총합계 : 0.21Gb, Row(적재 데이터) 총합계 : 911,761

## 2. TABLE COLUMN 항목 수집 상태.

### 1) 유방암\_환자\_건강정보 Table (BRCN\_PT\_HLINF)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	30	병력당뇨여부무응답구분명	O
2	환자대체번호	O	31	병력결핵여부무응답구분코드	O
3	입원일자	O	32	병력결핵여부무응답구분명	O
4	건강정보순번	O	33	병력간질환여부무응답구분코드	O
5	퇴원일자	O	34	병력간질환여부무응답구분명	O
6	현재음주여부무응답구분코드	O	35	병력고지혈증여부무응답구분코드	O
7	현재음주여부무응답구분명	O	36	병력고지혈증여부무응답구분명	O
8	음주력여부무응답구분코드	O	37	병력암여부무응답구분코드	O
9	음주력여부무응답구분명	O	38	병력암여부무응답구분명	O
10	음주시작연령값	X	39	병력암종류코드	O
11	음주종류코드	O	40	병력암종류명	O
12	음주종류명	O	41	병력우울증여부무응답구분코드	O
13	음주량	O	42	병력우울증여부무응답구분명	O
14	음주횟수	X	43	병력불면증여부무응답구분코드	O
15	음주기간년수	O	44	병력불면증여부무응답구분명	O
16	금주시작년도	X	45	병력심장질환여부무응답구분코드	O
17	현재흡연여부무응답구분코드	O	46	병력심장질환여부무응답구분명	O
18	현재흡연여부무응답구분명	O	47	기타병력여부무응답구분코드	O
19	흡연력여부무응답구분코드	O	48	기타병력여부무응답구분명	O
20	흡연력여부무응답구분명	O	49	기타병력질환내용	O
21	흡연시작연령값	X	50	주증상여부무응답구분코드	O
22	흡연량	O	51	주증상여부무응답구분명	O
23	흡연기간년수	O	52	주증상내용	O
24	금연시작년도	X	53	타병원진단후전원여부무응답구분코드	O
25	병력여부무응답구분코드	O	54	타병원진단후전원여부무응답구분명	O
26	병력여부무응답구분명	O	55	퇴원상태코드	O
27	병력고혈압여부무응답구분코드	O	56	퇴원상태명	O
28	병력고혈압여부무응답구분명	O	57	생성일시	O
29	병력당뇨여부무응답구분코드	O			

- '23 대장암\_환자\_건강정보(CLCN\_PT\_HLINF)와의 공통 Column 반영.

2) 유방암\_환자\_가족력 Table (BRCN\_PT\_FMHT)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	14	가족력결핵여부무응답구분코드	O
2	환자대체번호	O	15	가족력결핵여부무응답구분명	O
3	가족력기록일자	O	16	가족병력간질환여부무응답구분코드	O
4	가족력기록순번	O	17	가족병력간질환여부무응답구분명	O
5	가족력여부무응답구분코드	O	18	가족병력암여부무응답구분코드	O
6	가족력여부무응답구분명	O	19	가족병력암여부무응답구분명	O
7	환자가족관계코드	O	20	가족력암종류코드	O
8	환자가족관계명	O	21	가족력암종류명	O
9	환자가족관계기타내용	O	22	가족력암종류기타내용	O
10	가족병력고혈압여부무응답구분코드	O	23	가족병력기타여부무응답구분코드	O
11	가족병력고혈압여부무응답구분명	O	24	가족병력기타여부무응답구분명	O
12	가족병력당뇨여부무응답구분코드	O	25	가족병력기타내용	O
13	가족병력당뇨여부무응답구분명	O	26	생성일시	O

- '23 대장암\_환자\_가족력(CLCN\_PT\_FMHT)와의 공통 Column 반영.

3) 유방암\_환자\_산과정보 Table (BRCN\_PT\_OBINF)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	14	출산자녀수	O
2	환자대체번호	O	15	모유수유경험여부무응답구분코드	X
3	산과기록일자	O	16	모유수유경험여부무응답구분명	X
4	산과기록순번	O	17	모유수유월수	X
5	결혼여부무응답구분코드	O	18	경구피임약사용여부무응답구분코드	X
6	결혼여부무응답구분명	O	19	경구피임약사용여부무응답구분명	X
7	결혼상세코드	O	20	경구피임약사용월수	X
8	결혼상세명	O	21	폐경여부무응답구분코드	O
9	HRT경험여부무응답구분코드	X	22	폐경여부무응답구분명	O
10	HRT경험여부무응답구분명	X	23	폐경연령값	O
11	HRT시행월수	X	24	자궁적출술여부무응답구분코드	O
12	초경연령값	O	25	자궁적출술여부무응답구분명	O
13	출산연령값	O	26	생성일시	O

4) 유방암\_진단\_신체계측정보 Table (BRCN\_DIAG\_ANINF)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	7	BMI값	O
2	환자대체번호	O	8	이완혈압값	O
3	신체계측기록일자	O	9	수축혈압값	O
4	신체계측기록순번	O	10	ECOG코드	O
5	신장측정값	O	11	ECOG명	O
6	체중측정값	O	12	생성일시	O

- '23 대장암\_진단\_신체계측정보(CLCN\_DIAG\_ANINF)와의 공통 Column 반영.

5) 유방암\_진단\_정보 Table (BRCN\_DIAG\_INF)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	8	진단명	O
2	환자대체번호	O	9	진단KCD코드	O
3	진단등록일자	O	10	진단KCD명	O
4	진단등록순번	O	11	진단SNOMEDCT코드	O
5	유방암진단종류코드	O	12	진단SNOMEDCT명	O
6	유방암진단종류명	O	13	생성일시	O
7	진단코드	O			

- '23 대장암\_진단\_정보(CLCN\_DIAG\_INF)와의 공통 Column 반영.

6) 유방암\_진단\_전이 Table (BRCN\_DIAG\_MTST) - 수집 미동의

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	X	8	전이양상명	X
2	환자대체번호	X	9	전이진단방법코드	X
3	전이진단일자	X	10	전이진단방법명	X
4	전이진단순번	X	11	전이치료내용	X
5	전이부위코드	X	12	생성일시	X
6	전이부위명	X			
7	전이양상코드	X			

7) 유방암\_진단\_병기 Table (BRCN\_DIAG\_STAG)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	8	임상TNM병기값	O
2	환자대체번호	O	9	임상T병기값	O
3	진단병기기록일자	O	10	임상N병기값	O
4	진단병기기록순번	O	11	임상M병기값	O
5	임상종양특성코드	O	12	생성일시	O
6	임상종양특성명	O			
7	AJCC년도	X			

- '23 대장암\_진단\_병기(CLCN\_DIAG\_STAG)와의 공통 Column 반영.

8) 유방암\_검사\_진단 Table (BRCN\_EXAM\_DIAG)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	10	진단검사EDI명	O
2	환자대체번호	O	11	진단검사SNOMEDCT코드	O
3	진단검사일자	O	12	진단검사SNOMEDCT명	O
4	진단검사순번	O	13	진단검사LOINC코드	X
5	진단검사종류코드	O	14	진단검사LOINC명	X
6	진단검사종류명	O	15	진단검사결과내용	O
7	진단검사코드	O	16	진단검사결과단위내용	O
8	진단검사명	O	17	진단검사정상범위참고값	O
9	진단검사EDI코드	O	18	생성일시	O

- '23 대장암\_검사\_진단(CLCN\_EXAM\_DIAG)와의 공통 Column 반영.

9) 유방암\_검사\_영상 Table (BRCN\_EXAM\_IMAG)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	13	영상검사결과내용	O
2	환자대체번호	O	14	유방밀도분류코드	O
3	영상검사일자	O	15	유방밀도분류명	O
4	영상검사순번	O	16	영상검사결과진단분류코드	O
5	영상검사종류코드	O	17	영상검사결과진단분류명	O
6	영상검사종류명	O	18	미세석회화여부확인불가구분코드	O
7	영상검사코드	O	19	미세석회화여부확인불가구분명	O
8	영상검사명	O	20	비대칭여부확인불가구분코드	O
9	영상검사EDI코드	O	21	비대칭여부확인불가구분명	O
10	영상검사EDI명	O	22	전신전이여부확인불가구분코드	O
11	영상검사SNOMEDCT코드	O	23	전신전이여부확인불가구분명	O
12	영상검사SNOMEDCT명	O	24	생성일시	O

10) 유방암\_검사\_생체 Table (BRCN\_EXAM\_BYPS)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	○	11	생체검사방법명	○
2	환자대체번호	○	12	생체검사방법종류코드	○
3	생체검사일자	○	13	생체검사방법종류명	○
4	생체검사순번	○	14	생체검사결과내용	○
5	생체검사판독일자	○	15	조직학적진단코드	○
6	생체검사부위코드	○	16	조직학적진단명	○
7	생체검사부위명	○	17	조직학적등급코드	○
8	생체검사부위편측성코드	○	18	조직학적등급명	○
9	생체검사부위편측성명	○	19	생성일시	○
10	생체검사방법코드	○			

- '23 대장암\_검사\_생체(CLCN\_EXAM\_BYPS)와의 공통 Column 반영.

11) 유방암\_검사\_면역병리 Table (BRCN\_EXAM\_IMPT)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	○	9	면역병리검사코드	○
2	환자대체번호	○	10	면역병리검사소견분류코드	○
3	면역병리검사일자	○	11	면역병리검사소견분류명	○
4	면역병리검사순번	○	12	면역병리검사결과값	○
5	면역병리판독일자	○	13	면역병리검사결과단위코드	○
6	면역병리검사종류코드	○	14	면역병리검사결과단위명	○
7	면역병리검사종류명	○	15	생성일시	○
8	면역병리검사명	○			

12) 유방암\_검사\_분자병리 Table (BRCN\_EXAM\_MLPT)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	○	10	분자병리검사방법코드	○
2	환자대체번호	○	11	분자병리검사방법명	○
3	분자병리검사일자	○	12	분자병리검사결과코드	○
4	분자병리검사순번	○	13	분자병리검사결과명	○
5	분자병리판독일자	○	14	분자병리검사결과종류코드	○
6	분자병리검사코드	○	15	분자병리검사결과종류명	○
7	분자병리검사명	○	16	분자병리검사결과값	○
8	분자병리검사종류코드	○	17	생성일시	○
9	분자병리검사종류명	○			

13) 유방암\_검사\_생식세포변이검사 Table (BRCN\_EXAM\_GMVX)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	15	미확인변형체감지여부구분명	O
2	환자대체번호	O	16	DNA변이정보A값	O
3	생식세포변이검사일자	O	17	DNA변이정보B코드	O
4	생식세포변이검사순번	O	18	DNA변이정보B명	O
5	생식세포변이검사코드	O	19	DNA변이정보C코드	O
6	생식세포변이검사명	O	20	DNA변이정보C명	O
7	생식세포변이검사방법코드	O	21	아미노산변이정보A코드	O
8	생식세포변이검사방법명	O	22	아미노산변이정보A명	O
9	생식세포변이검사유전자종류코드	O	23	아미노산변이정보B값	O
10	생식세포변이검사유전자종류명	O	24	아미노산변이정보C코드	O
11	생식세포전이검사결과내용	O	25	아미노산변이정보C명	O
12	병원변형체감지여부구분코드	O	26	참고순번	O
13	병원변형체감지여부구분명	O	27	생성일시	O
14	미확인변형체감지여부구분코드	O			

- 2023년 신규 수집 상태. 데이터 확인 필요.

14) 유방암\_치료\_수술 Table (BRCN\_TRTM\_OPRT)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	22	수술부위명	O
2	환자대체번호	O	23	유방암수술방법코드	O
3	수술일자	O	24	유방암수술방법명	O
4	수술순번	O	25	ASA점수코드	O
5	수술종류코드	O	26	ASA점수명	O
6	수술종류명	O	27	수술절개방법내용	O
7	수술코드	O	28	수술절개선유형코드	O
8	수술명	O	29	수술절개선유형명	O
9	수술EDI코드	O	30	수술종양위치코드	O
10	수술EDI명	O	31	수술종양위치명	O
11	수술SNOMEDCT코드	O	32	수술종양상세위치코드	O
12	수술SNOMEDCT명	O	33	수술종양상세위치명	O
13	수술ICD9CM코드	X	34	재건술분류코드	O
14	수술ICD9CM명	X	35	재건술분류명	O
15	수술마취부위코드	O	36	재건술방법코드	O
16	수술마취부위명	O	37	재건술방법명	O
17	추정실혈량	O	38	감시림프절검출방법코드	O
18	수술소요시간	O	39	감시림프절검출방법명	O
19	수술목적코드	O	40	수술후평가항목코드	X
20	수술목적명	O	41	수술후평가항목명	X
21	수술부위코드	O	42	생성일시	O



15) 유방암\_검사\_유전자발현검사 Table (BRCN\_EXAM\_GNRX)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	○	6	유전자발현검사종류명	○
2	환자대체번호	○	7	유전자발현검사결과종류코드	○
3	유전자발현검사일자	○	8	유전자발현검사결과종류명	○
4	유전자발현검사순번	○	9	유전자발현검사결과내용	○
5	유전자발현검사종류코드	○	10	생성일시	○

- 2023 신규 수집 상태. 데이터 확인 필요.

16) 유방암\_치료\_수혈 Table (BRCN\_TRTM\_BLTR)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	○	7	혈액제제코드	○
2	환자대체번호	○	8	혈액제제명	○
3	수혈처방일자	○	9	혈액제제용량	○
4	수혈처방순번	○	10	수혈팩수	○
5	혈액제제종류코드	○	11	수혈총량	○
6	혈액제제종류명	○	12	생성일시	○

- '23 대장암\_치료\_수혈(CLCN\_TRTM\_BLTR)와의 공통 Column 반영.

17) 유방암\_치료\_약제성분 Table (BRCN\_TRTM\_DRIN)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	○	16	약제명	○
2	환자대체번호	○	17	약제EDI코드	○
3	약제처방일자	○	18	약제EDI명	○
4	약제처방순번	○	19	약제ATC코드	○
5	약제분류코드	○	20	약제ATC명	○
6	약제분류명	○	21	약제RXNORM코드	X
7	항암약제구분코드	○	22	약제RXNORM명	X
8	항암약제구분명	○	23	약제처방용량	○
9	약제성분종류코드	○	24	약제처방용량단위코드	○
10	약제성분종류명	○	25	약제처방용량단위명	○
11	약제성분코드	○	26	약제처방일수	○
12	약제성분명	○	27	약제투여경로코드	○
13	약제성분SNOMEDCT코드	○	28	약제투여경로명	○
14	약제성분SNOMEDCT명	○	29	생성일시	○
15	약제코드	○			

18) 유방암\_진단\_수술후합병증 Table (BRCN\_DIAG\_AFOC) – 수집 미동의

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	X	6	수술후합병증내용	X
2	환자대체번호	X	7	합병증등급코드	X
3	합병증발생수술일자	X	8	합병증등급명	X
4	합병증발생수술순번	X	9	생성일시	X
5	합병증발생일자	X			

19) 유방암\_치료\_항암요법 Table (BRCN\_TRTM\_ANTP)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	12	항암요법종료이유명	X
2	환자대체번호	O	13	항암요법평가일자	X
3	항암요법시작일자	O	14	항암요법평가항목코드	X
4	항암요법순번	O	15	항암요법평가항목명	X
5	항암요법종료일자	X	16	항암요법부작용항목코드	O
6	항암요법명	O	17	항암요법부작용항목명	O
7	항암요법주기수	O	18	항암요법부작용등급코드	O
8	항암요법라인수	X	19	항암요법부작용등급명	O
9	항암요법임시중단여부확인불가구분코드	X	20	항암요법치료목적코드	O
10	항암요법임시중단여부확인불가구분명	X	21	항암요법치료목적명	O
11	항암요법종료이유코드	X	22	생성일시	O

20) 유방암\_치료\_방사선 Table (BRCN\_TRTM\_RD)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	16	방사선치료부위명	O
2	환자대체번호	O	17	방사선선량	O
3	방사선치료처방일자	O	18	방사선시행횟수	O
4	방사선치료처방순번	O	19	방사선총선량	O
5	방사선치료종류코드	O	20	방사선치료평가일자	O
6	방사선치료종류명	O	21	방사선치료부작용시기구분코드	X
7	방사선치료코드	O	22	방사선치료부작용시기구분명	X
8	방사선치료명	O	23	방사선치료부작용종류코드	X
9	방사선치료EDI코드	O	24	방사선치료부작용종류명	X
10	방사선치료EDI명	O	25	방사선치료부작용등급코드	X
11	방사선치료SNOMEDCT코드	O	26	방사선치료부작용등급명	X
12	방사선치료SNOMEDCT명	O	27	방사선치료평가항목코드	O
13	방사선치료목적코드	O	28	방사선치료평가항목명	O
14	방사선치료목적명	O	29	생성일시	O
15	방사선치료부위코드	O			

21) 유방암\_검사\_외과병리 Table (BRCN\_EXAM\_SGPT)

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	O	31	상피내암핵등급코드	O
2	환자대체번호	O	32	상피내암핵등급명	O
3	외과병리검사일자	O	33	수술절제면안전경계길이값	O
4	외과병리검사순번	O	34	수술절제면상세내용	O
5	외과병리판독일자	O	35	수술절제면잔존코드	O
6	외과병리채취부위코드	O	36	수술절제면잔존명	O
7	외과병리채취부위명	O	37	전이암크기값	O
8	외과병리검사부위편측성코드	O	38	동반병변코드	O
9	외과병리검사부위편측성명	O	39	동반병변명	O
10	외과병리검사결과종양수	O	40	잔여종양코드	O
11	종양가로 길이값	O	41	잔여종양명	O
12	종양세로 길이값	O	42	림프절총수	O
13	종양높이값	O	43	양성림프절수	O
14	종양최대직경값	O	44	감시림프절총수	O
15	조직학적진단코드	O	45	양성감시림프절수	O
16	조직학적진단명	O	46	비감시림프절총수	O
17	조직학적등급코드	O	47	양성비감시림프절수	O
18	조직학적등급명	O	48	AJCC년도	X
19	림프혈관침윤존재여부구분코드	O	49	수술후병리TNM병기값	O
20	림프혈관침윤존재여부구분명	O	50	수술후병리T병기값	O
21	림프절주위침윤존재여부구분코드	O	51	수술후병리N병기값	O
22	림프절주위침윤존재여부구분명	O	52	수술후병리M병기값	O
23	유두침윤존재여부구분코드	O	53	생성일시	O
24	유두침윤존재여부구분명	O			
25	외과병리미세석회화존재여부구분코드	O			
26	외과병리미세석회화존재여부구분명	O			
27	상피내암괴사존재여부구분코드	O			
28	상피내암괴사존재여부구분명	O			
29	상피내암동반존재여부구분코드	O			
30	상피내암동반존재여부구분명	O			

22) 유방암\_추적관찰\_재발 Table (BRCN\_FUOR\_RLPS) – 수집 미동의

컬럼 번호	컬럼명	작성	컬럼 번호	컬럼명	작성
1	병원코드	X	8	재발진단방법코드	X
2	환자대체번호	X	9	재발진단방법명	X
3	재발진단일자	X	10	재발종류코드	X
4	재발진단순번	X	11	재발종류명	X
5	재발부위코드	X	12	재발증상코드	X
6	재발부위명	X	13	재발증상명	X
7	재발치료내용	X	14	생성일시	X

★ Ref. 환자대체번호 생성 과정

- 환자번호는 PID로 존재하며 이는 ODS.PAM\_PIDREPID Table에서 확인할 수 있다.
- 환자'대체'번호 생성에 필요한 연구대체번호(R로 시작)
  - 'R' || TOCHAR(SYSDATE, 'HH24MISSUS') 과정으로 생성. 단, DB2에서는 **NOW**로 사용해야 US부분까지 나온다.
  - HH24 : 24시간 형태의 **시간**
  - MI: **분**, SS: **초**, US: **Micro초**
- 연구대체번호(EX. R135318947995)를 이용해 환자대체번호(C로 시작) 생성.
  - R1353 1 8 947995 --> 여기서 C100을 앞에 붙여 C1001947995 번호가 만들어진다.

```
SELECT 'R' || TO_CHAR(NOW, 'HH24MISSUS'), NOW, SYSDATE
FROM SYSIBM.SYSDUMMY1
```

Figure 1. 연구대체번호 생성 테스트 Query

연구번호 확인 테이블 : USR.TM\_950019\_20230721111635248

### 3. 유방암 라이브러리 구축 과정

- 대장암 라이브러리를 구축하며 적용한 공통방식은 생략하며 유방암 특화 데이터 수집의 경우를 중점으로 기술.

#### 1) 암환자 등록 리스트 CTE

- LCD.PT\_CNRG\_TRGT Table (암종별연구대상자)에서 유방암 대상 환자 리스트 추출.
  - 테이블 적재 조건 : 최초 진단일 기준 2010 ~ 2022년 대상의 환자만 추출.
- 해당 CTE는 원천 테이블에서 데이터(ROW)를 필터링할 때 JOIN 조건으로 활용한다 (처리 시간 감소).

```
WITH PTINF AS (  
    SELECT  
        PT_SBST_NO AS PTSBSTNO  
    FROM  
        LCD.PT_CNRG_TRGT  
    WHERE CKND_CLSF_CD = 'CD0002' /*유방암코드*/  
    AND FRDX_YMD BETWEEN '2010-01-01' AND '2022-12-31'  
    /*최초진단일시 2010~2022년*/  
)
```

Figure 2. 유방암 진단 환자 리스트 추출 CTE

#### 2) 유방암\_환자\_건강정보 (BRCN\_PT\_HLINF)

- 주요 원천 테이블 : CDW.MR\_NRIE (간호초기평가 Table)
- 과거 병력과 관련된 소견이 저장된 Column : PASTHIST

```
1  [발생일자] 20년전 [병력내역] 고혈압 [비고] 무  
2  [발생일자] 20년전 [병력내역] 고지혈증 [비고] 무  
3  [발생일자] 10년전 [병력내역] 심장질환 [비고] 부정맥  
4  [발생일자] 2021 [병력내역] 기타 [비고] 손가락 관절염  
5  [발생일자] 2017 [병력내역] 기타 [비고] knee pain으로 op(menisectomy)  
6  
7  
8  [발생일자] 2015 [병력내역] 기타 [비고] Lt.pontine infarction  
9  [발생일자] 2021.12.3 [병력내역] 종양 [비고] Lt. Breast Ca  
10 [발생일자] 2017.02 [병력내역] 기타 [비고] Brain aneurysm  
11 [발생일자] 모름 [병력내역] 간염 [비고] HBV  
12 [발생일자] [병력내역] [비고]  
13  
14  
15 # 과거 병력이 없는 데이터 예시  
16 [발생일자] [병력내역] [비고]  
17 [발생일자] [병력내역] [비고]  
18 [발생일자] [병력내역] [비고]  
19 [발생일자] [병력내역] [비고]  
20 [발생일자] [병력내역] [비고]
```

Figure 3. PASTHIST Column 데이터 샘플

- 과거 병력 소견 데이터에서 정규표현식을 통해 '과거 암', '기타 질환', '고혈압' 등의 여부 데이터를 추출한다.
  - 'K-CURE 대장암 임상 라이브러리'의 경우에도 같은 데이터를 추출하므로 코드가 유사해 정규표현식 SQL 생략.
- 추출 결과 예시
  - 암종 : "Lt. Breast Ca --> 유방".
  - 기타 질환 : "[병력내역] 기타 [비고]" 위치에 작성된 내용 --> 손가락 관절염, Brain aneurysm 등.
  - 고혈압 여부 : "고혈압", "htn (Hypertension)" 또는 "hbp (High Blood Pressure)" 내용이 작성된 경우.

### 3) 유방암\_환자\_가족력 (BRCN\_PT\_FMHT)

- 주요 원천 테이블 : **CDW.MR\_NRIE** (간호초기평가 Table)
- 과거 병력과 관련된 소견이 저장된 Column : **FAMILYHIST**

1	[관계]	부	[생존여부]	N	[병력내역]	종양	[비고]	전립선암
2	[관계]	모	[생존여부]	Y	[병력내역]	종양	[비고]	자궁암
3	[관계]	자매	[생존여부]	Y	[병력내역]	종양	[비고]	갑상선, 유방암
4	[관계]		[생존여부]		[병력내역]		[비고]	
5	[관계]		[생존여부]		[병력내역]		[비고]	
6								
7	[관계]	배우자	[생존여부]	Y	[병력내역]	고혈압	[비고]	무
8	[관계]	자매	[생존여부]	N	[병력내역]	종양	[비고]	유방암
9	[관계]	형제	[생존여부]	Y	[병력내역]	기타	[비고]	신부전
10	[관계]	형제	[생존여부]	Y	[병력내역]	당뇨	[비고]	무
11	[관계]	형제	[생존여부]	Y	[병력내역]	고혈압	[비고]	무
12								
13	#	소견	정보가 없는 경우	예시				
14	[관계]		[생존여부]		[병력내역]		[비고]	
15	[관계]		[생존여부]		[병력내역]		[비고]	
16	[관계]		[생존여부]		[병력내역]		[비고]	
17	[관계]		[생존여부]		[병력내역]		[비고]	
18	[관계]		[생존여부]		[병력내역]		[비고]	

Figure 4. FAMILYHIST Column 데이터 샘플

- 가족 병력(가족력) 소견 데이터에서 정규표현식을 통해 '가족 암 여부', '가족력 기타 질환', '고혈압' 등의 여부 데이터를 추출한다.
- 'K-CURE 대장암 임상 라이브러리'의 경우에도 같은 데이터를 추출하므로 코드가 유사해 정규표현식 SQL 생략.
- 추출 결과 예시
  - 암종 : "전립선암 --> 전립선", "유방암 --> 유방".
  - 기타 질환 : "[병력내역] 기타 [비고]" 위치에 작성된 내용 --> 신부전 등.
  - 고혈압 여부 : "고혈압", "htn (Hypertension)" 또는 "hbp (High Blood Pressure)" 내용이 작성된 경우.
- "[관계]" 항목에 작성된 내용을 바탕으로 가족관계 종류를 구분한다.
- "부" -> 아버지, "모" -> 어머니, "자녀, 아들, 딸" -> 자녀 등.

```

CASE WHEN TRIM(REGEXP_SUBSTR(TB.FAMILYHIST,'(?<=₩[관계₩]).*(?<=₩[생존여부₩]))' <> '' THEN
CASE WHEN REGEXP_LIKE(TB.FAMILYHIST, '₩s 부₩s') THEN '아버지'
      WHEN REGEXP_LIKE(TB.FAMILYHIST, '₩s 모₩s') THEN '어머니'
      WHEN REGEXP_LIKE(TB.FAMILYHIST, '₩s 자녀₩s|아들|딸') THEN '자녀'
      WHEN REGEXP_LIKE(TB.FAMILYHIST, '₩s 형제₩s|₩s 자매₩s|동생|형|누나|언니|오빠') THEN '형제자매'
      WHEN REGEXP_LIKE(TB.FAMILYHIST, '₩s 이모|삼촌|외조모|외조부|고모|조모|조부') THEN '친척'
      WHEN REGEXP_LIKE(TB.FAMILYHIST, '₩s 기타₩s|손자|배우자|조카') THEN '기타'
END
ELSE '무응답'
END AS PT_FM_REL_CONT /*가족관계 상세*/

```

Figure 5. 가족관계 데이터 추출 목적의 CASE-WHEN-THEN

#### 4) 유방암\_환자\_산과정보 (BRCN\_PT\_OBINF)

- 주요 원천 테이블 : CDW.MR\_NRIE (간호초기평가), ODS.EMR\_MNWDGIR (간호정보조사지산과정보).

##### A. '#5 #6, #7, #8 결혼여부응답, 결혼상세' 데이터 추출.

- CDW.MR\_NRIE Table의 'MARYSTATNM' (결혼상태) Column.

ABC MARYSTATNM ↑
기타
기혼
독신
동거
미혼
별거
불명
사별
이혼
[NULL]

```

CASE
  WHEN M1.MARYSTATNM IN ('기혼', '사별', '이혼', '동거', '별거') THEN '예'
  WHEN M1.MARYSTATNM IN ('미혼', '독신') THEN '아니오'
  ELSE '무응답'
END
AS MARG_YN_NOANS_SPNM /*#6 결혼여부응답구분명*/

/* NULL 은 '무응답'. 미혼과 독신은 '아니오' (결혼 無). 외는 '예' (결혼 有) */

```

Figure 6. #5, #6결혼여부 데이터 추출 Query

```

CASE
  WHEN M1.MARYSTATNM = '미혼' THEN '1'
  WHEN M1.MARYSTATNM = '기혼' THEN '2'
  WHEN M1.MARYSTATNM = '이혼' THEN '3'
  WHEN M1.MARYSTATNM = '사별' THEN '4'
  WHEN M1.MARYSTATNM IS NOT NULL THEN '9' /*기타*/
END
AS MARG_DETL_CD /*#7 결혼상세코드*/
, CASE
  WHEN M1.MARYSTATNM IN ('미혼', '기혼', '이혼', '사별') THEN M1.MARYSTATNM
  WHEN M1.MARYSTATNM IS NOT NULL THEN '기타'
END
AS MARG_DETL_NM /*#8 결혼상세명*/

```

Figure 7. #7, #8 결혼상세 데이터 추출 Query

##### B. '#24, #25 자궁적출술여부' 데이터 추출.

- Keyword : Hysterectomy
- CDW.PE\_OP\_LIST (수술내역) Table에서 자궁적출(절제)수술 이력이 있는지 확인한다.

ABC OPCD	ABC OPENM	ABC OPKNM
7499.004.03	Cesarean hysterectomy	제왕자궁절제술
686.057.00	Laparoscopic assisted modified radical hysterectomy	복강경적 변형 근치적 자궁 절제
686.056.00	Laparoscopic assisted radical hysterectomy	복강경적 근치적 자궁 절제
686.056.01	Laparoscopic radical hysterectomy (I)	복강경적 근치적 자궁절제술 (I형)
686.056.02	Laparoscopic radical hysterectomy (II)	복강경적 근치적 자궁절제술 (II형)
686.056.03	Laparoscopic radical hysterectomy (III)	복강경적 근치적 자궁절제술 (III형)
683.008.00	Laparoscopic subtotal abdominal hysterectomy	복강경하 아전적 자궁절제술

Figure 8. Hysterectomy 수술 종류 샘플.

```

LEFT JOIN ( /* 자궁적출술(Hysterectomy) 관련 수술내역이 있는지 확인 */
  SELECT
    PTSBSTNO
    , 'Y' AS UTECY /* ROW 가 매칭되면 수술 경험이 있다고 판단 */
  FROM CDW.PE_OP_LIST
  WHERE REGEXP_LIKE(OPENM, 'Hysterectomy', 'i')
) T
ON T.PTSBSTNO = P.PTSBSTNO

```

Figure 9. Hysterectomy 관련 수술 경험이 있는지 확인하는 JOIN

## 5) 유방암\_진단\_신체계측정보 (BRCN\_DIAG\_ANINF)

- 주요 원천 테이블 : CDW.MR\_CLNC\_REC (임상관찰기록 Table).

### A. 필수 데이터 수집 (신장, 체중, BMI, 이완기 혈압, 수축기 혈압).

- 원천 테이블에 적재된 데이터 확인.

ITEMCD  
 IT000015: BP(S) - 수축혈압  
 IT000016: BP(D) - 이완혈압  
 IT000030: 키(CM)  
 IT000031: 체중(KG)  
 IT000032: 체중(G)

ABC ITEMCD	ABC ITEMNM	ABC SUPCD	ABC SUPI	123 RECRSLT	ABC ORNGRECR	REC YMD
IT000016	BP(D)	IT000001	V/S	70	70	2014-02-28
IT000022	맥박	IT000001	V/S	78	78	2014-02-28
IT000026	호흡	IT000001	V/S	20	20	2014-02-28
IT000028	체온	IT000001	V/S	36.5	36.5	2014-02-28
IT000015	BP(S)	IT000001	V/S	130	130	2013-10-07
IT000016	BP(D)	IT000001	V/S	60	60	2013-10-07
IT000022	맥박	IT000001	V/S	86	86	2013-10-07
IT000026	호흡	IT000001	V/S	20	20	2013-10-07
IT000028	체온	IT000001	V/S	36.7	36.7	2013-10-07
IT000030	키(cm)	IT000002	신체계측	162	162	2013-10-07
IT000031	체중(Kg)	IT000002	신체계측	52	52	2013-10-07
IT000077	SpO <sub>2</sub> (%)	IT000006	호흡기계	99	99	2013-10-07

Figure 10. 관찰 기록 정보 원본 데이터

- Row는 각각 1개의 검사 항목에 대한 기록이기 때문에 1명의 환자에 대한 전체(신장, 체중, 혈압 등) 항목의 기록을 1개의 Row로 결합하기 위해서는 SubQuery와 함께 '그룹화'가 필요하다.

- 그룹화 후 필요 데이터 추출.
- 그룹화를 하면 1개의 Row에 전체 기록 항목들에 대한 결과가 결합된 것을 확인할 수 있다.
- 결과로 추출된 Temporary Table에서 신장, 체중, 혈압 데이터가 모두 기록된 Row를 최종 활용한다.

	신장 측정일	신장	체중 측정일	체중	수축 혈압	이완 혈압
ABC REC YMD	123 MX_HT_DD	123 MX_HT	123 MX_WT_DD	123 MX_WT	123 MX_BPS	123 MX_BPD
20211026	20211026	160	2021-10-26	59	120	80
20211027	[NULL]	[NULL]	[NULL]	[NULL]	140	80
20211029	[NULL]	[NULL]	[NULL]	[NULL]	129	78
20211117	20211117	161.8	2021-11-17	60.5	120	80
20211118	[NULL]	[NULL]	[NULL]	[NULL]	140	90
20211119	[NULL]	[NULL]	[NULL]	[NULL]	120	70
20211208	20211208	160	2021-12-08	62.4	132	71
20211209	[NULL]	[NULL]	[NULL]	[NULL]	126	76
20211210	[NULL]	[NULL]	[NULL]	[NULL]	116	68
20161230	20161230	155	2016-12-30	59	140	70

Figure 11. 그룹화 후 최종으로 사용할 데이터 Row 확인

### B. 품질진단 시 확인된 데이터 유효 범위.

- '대장암 라이브러리 품질진단'의 경우 '신장, 체중, BMI'에 대한 유효범위 검사 SQL을 포함했다.
- 신장: 250이상의 값 이상치. - 체중: 250이상의 값 이상치. - BMI: 45이상의 값 이상치.



## 6) 유방암\_진단\_정보 (BRCN\_DIAG\_INF)

- 주요 원천 테이블 : **CDW.RG\_DIAG\_LIST** (진단정보 Table).
- 유방암 대상 환자가 2010-01-01 ~ 2022-12-31 기간 동안의 '암'관련 진단을 받은 이력을 가져온다.
  - 유방관련 암종코드 : C50 계열.
  - 유방 상피내암 코드: D05 계열.

ABC DIAGCD	ABC DIAGENGNM	ABC DIAGHNGNM	ABC ICD10CD
D0519.000.06	DCIS of breast, unspecified side	유방의 관내 제자리암종, 상세불명 쪽	D0519
D0570.000.00	Other carcinoma in situ of breast, right	기타 유방의 제자리암종, 오른쪽	D0570
C5000.000.01	Malignant neoplasm of nipple and areola, right	유두 및 유륜의 악성신생물, 오른쪽	C5000
C5099.000.05	Breast cancer, unspecified, unspecified side	상세불명의 유방암, 상세불명 쪽	C5099
C5041.000.01	Malignant neoplasm of upper-outer quadrant of breast, left	유방의 상외사분의 악성신생물, 왼쪽	C5041

Figure 12. 유방관련 암종 진단 정보 샘플

- 유방암 대상 환자의 '암'관련 진단 데이터 추출을 위한 CTE (Common Table Expression).

```

WITH PTINF AS (
  SELECT T.*
  FROM (
    SELECT /*유방암 진단을 받은 환자들의 모든 암 진단 + 유방암 관련 질병에 대한 정보 수집 목적의 SELECT*/
      A.PTSBSTNO /*환자등록번호*/
    , A.DIAGDD /*진단일자*/
    , A.DIAGCD /*진단코드*/
    , A.DIAGENGM /*진단영문명*/
    , A.DIAGHNGM /*진단한글명*/
    , A.ICD10CD /*ICD10 CODE --> 여기서 3 자리 문자 자르기 가능*/
    , SUBSTR(A.DIAGCD,1,3) AS DIAG_KIND_CD
    /*환자번호 + 진단코드 기준으로 RN 생성 --> 진단일자가 달라도 코드가 같은 경우 있으므로*/
    , ROW_NUMBER() OVER(PARTITION BY A.PTSBSTNO, A.DIAGCD ORDER BY A.DIAGDD DESC) AS RN
  FROM
    CDW.RG_DIAG_LIST A /*진단정보 TABLE*/
  INNER JOIN (
    SELECT
      PT_SBST_NO AS PTSBSTNO
    FROM
      LCD.PT_CNRG_TRGT
    WHERE 1=1
      AND CKND_CLSF_CD = 'CD0002' /*유방암코드*/
      AND FRDX_YMD BETWEEN '2010-01-01' AND '2022-12-31'
  ) C1 /*유방암 대상 환자만 추출할 목적*/
  USING(PTSBSTNO)
  WHERE
    A.DIAGCD LIKE 'C%' /*유방암 대상 환자의 모든 암 정보 추출*/
    OR A.DIAGCD LIKE 'D05%' /*유방암 대상 환자의 기타질환(제자리암종) 정보 추출*/
  ) T
  WHERE T.RN = 1 /*중복방지 1 개만 구하기*/
  AND T.DIAGDD BETWEEN '2010-01-01' AND '2022-12-31' /*진단일자 기준 2010~2022 년*/
)

```

Figure 13. 유방암 대상 환자의 '암'관련 진단 정보 데이터 추출 CTE

## 7) 유방암\_검사\_영상 (BRCN\_EXAM\_IMAG)

- 주요 원천 테이블 : CDW.PE\_PACS\_RSLT (PACS검사결과 Table).
- 비정형 데이터 추출은 주로 'Mammography' 검사의 소견에서 이루어진다. 이는 약어로 'MMG'에 해당한다.

### A. '#14 #15 유방밀도분류' 데이터 추출

- Keyword : Mammography 결과에서 breast density에 대한 소견.
- Extremely dense | Fatty breast | Heterogeneously dense | Scattered fibroglandular | 기타

```
Breast Composition : Scattered fibroglandular tissue.

Breast composition : Heterogeneous dense, may lower the
sensitivity of mammography

Breast composition : Extremely dense, which could obscure a
lesion

Breast composition : Almost entirely fat
```

Figure 14. 유방밀도 소견에 대한 예시

```
CASE WHEN R2.EXAMKINDNM = 'MMG' THEN /* MMG: Mammography 의 소견에서 검사해야 한다. */
CASE
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'Extremely dense', 'i') THEN '1' /*Extremely dense*/
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'entirely fat', 'i') THEN '2' /*Fatty breast*/
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'Heterogen(e)?ous(ly)?( dense)?', 'i') THEN '3' /*Heterogeneously dense*/
  WHEN REGEXP_LIKE(R2.READCONCLUSION, '(Scattered )?fibroglandular', 'i') THEN '4' /*Scattered fibroglandular*/
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'iso( )?dens', 'i') THEN '9' /*iso-dense*/
END
END AS BRST_DENS_CLCD /*#14 유방밀도분류코드*/
```

Figure 15. #14, #15 유방밀도분류 데이터 추출 목적의 Query

### B. '#16 #17 영상검사결과진단분류' 데이터 추출.

- Keyword : 영상검사 결과 진단적 분류 (diagnostic classification) 코드.
  - negative | benign | probably benign | suspicious | highly suggestive | proven malignancy | 기타
- 
- BI-RADS 5<sup>th</sup> Edition 기준의 **Assessment Category** 분류.
  - 0 : **Incomplete** (Need additional imaging evaluation and/or prior examination for comparison). Not available
  - 1 : **Negative** (Essentially 0%)
  - 2 : **Benign** (Essentially 0%)
  - 3 : **Probably Benign** ( >0% but <= 2% )
  - 4 : **Suspicious** ( >2% but 95% ) Tissue diagnosis.
    - 4A = Low suspicion for malignancy ( >2% to <= 10% )
    - 4B = Moderate suspicion ( >10% to <= 50% )
    - 4C = High suspicion ( >50% to <= 95% )
  - 5 : **Highly suggestive** of malignancy ( >=95% )
  - 6 : Known **biopsy-proven** malignancy (Surgical Excision)

```

Category : 3 ( probable benign )

Category : 0      Category : 5      Category : 4C

Category 2 : Benign finding (Grade IV pattern of breast.)

```

Figure 16. 영상검사결과진단분류 소견 예시

```

CASE
  WHEN REGEXP_LIKE(DIAG_CLCD, '0') THEN '99'          /*기타(incomplete)*/
  WHEN REGEXP_LIKE(DIAG_CLCD, '1') THEN '01'          /*Negative*/
  WHEN REGEXP_LIKE(DIAG_CLCD, '3') THEN '03'          /*Probably benign*/
  WHEN REGEXP_LIKE(DIAG_CLCD, '2') THEN '02'          /*Benign*/
  WHEN REGEXP_LIKE(DIAG_CLCD, '4') THEN '04'          /*Suspicious*/
  WHEN REGEXP_LIKE(DIAG_CLCD, '5') THEN '05'          /*Highly suggestive*/
  WHEN REGEXP_LIKE(DIAG_CLCD, '6') THEN '06'          /*Proven malignancy*/
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'incomplete', 'i') THEN '99'
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'Proven malignancy|malignant', 'i') THEN '06'
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'negative( left| right)? (breast|finding|for)', 'i') THEN '01'
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'Probabl[ey] benign|probabl[ye]', 'i') THEN '03'
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'Highly (suggestive|suspicion)', 'i') THEN '05'
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'Suspicious|suspicion', 'i') THEN '04'
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'Benign', 'i') THEN '02'
END AS IMEX_RSLT_DIAG_CLCD /*#16 영상검사결과진단분류코드*/

```

Figure 17. #16 #17 영상검사결과진단분류 데이터 수집 목적의 Query

### C. '#18 #19 미세석회화여부' 데이터 추출.

- Keyword : Mammography 결과에서 microcalcification 여부 코드. (Yes | No | Unknown)

```

Findings :
Right UOQ에 clustered plemorphic microcalcification 보이며
malignancy로 진단된 병변임. 또한 subareolar 방향으로 nipple에서 3.9cm

Left UIQ에 clustered microcalcification이 있는데 주로 round, lucent
centered calcification들이 보이며, benign으로 생각됨.

microcalcification 은 보이지 않음.
양측 axilla 에 significant lymph node enlargement 는 보이지 않음.

```

Figure 18. 미세석회화(Microcalcification) 소견 예시

```

CASE WHEN R2.EXAMKINDNM = 'MMG' THEN
CASE
  WHEN REGEXP_LIKE(R2.READCONCLUSION,
    'microcalcification\s?은 (보이지|동반되어 있지|관찰되지) 않음|'
    || 'microcalcification(등의|이나)[\s\S]*(소견은 )?보이지 않음|'
    || 'no residual microcalcification|'
    || 'microcalcification 이나.*관찰되지 않음', 'i') THEN 'N'
  WHEN REGEXP_LIKE(R2.READCONCLUSION, 'microcalcification', 'i') THEN 'Y'
    /*위의 경우를 제외하고 microcalcifications 언급되면 YES*/
  ELSE 'Z' /*Unknown*/
END
END

```

Figure 19. #18, #19 미세석회화여부 데이터 수집 목적의 Query

#### D. '#20 #21 비대칭여부확인불가구분' 데이터 추출.

- Keyword : Mammography 결과에서 asymmetry 여부 코드. (Yes | No | Unknown)

```
Right mid outer에 약간의 infiltration과 focal asymmetry가 관찰되며

[결론]
1. Postoperative left breast.
2. Benign calcifications in both breasts.
3. Focal asymmetry, right LIQ. probable benign

[결론]
Asymmetry, right upper breast
```

Figure 20. 비대칭(Asymmetry)여부 소견 예시

```
CASE WHEN R2.EXAMKINDNM = 'MMG' THEN
CASE
WHEN REGEXP_LIKE(R2.READCONCLUSION, 'asymmetry 는 (현재 뚜렷하지|이전보다 분명치|관찰되지) 없음|'
|| 'asymmetry.{1,30}없음', 'i') THEN 'No'
WHEN REGEXP_LIKE(R2.READCONCLUSION, 'asymmetry', 'i') THEN 'Yes' /*위의 경우를 제외하고 asymmetry 언급되면 YES*/
ELSE 'Unknown' /*Unknown*/
END
END
```

Figure 21. #20, #21 비대칭여부확인불가구분 데이터 추출 Query

#### E. '#22 #23 전신전이여부확인불가구분' 데이터 추출.

- Keyword : Distant Metastasis

```
Multiple hypermetabolic lesions in above described sites
-- Left breast cancer with multiple distant metastases and
peritoneal carcinomatsis, more likely 예 (Y)
-- Double primary cancer (breast and pancreas) with
multiple distant metastases
and peritoneal carcinomatsis, less likely

- 그 외 lymph node나 distant metastasis를 의심할 abnormal FDG
uptake 소견은 보이지 않음. 아니오(N)

3) No evidence of other metabolically active lesion suggesting
axillary nodal or distant metastasis in relation to breast
cancer. 아니오(N)

-- Multiple metastases of known breast cancer, more
likely
-- Lung cancer in RLL with multiple distant
metastases, less likely
Rec) tissue confirmation
```

Figure 22. 전신전이(Distant metastasis) 소견 예시

```
CASE
WHEN REGEXP_LIKE(R2.READCONCLUSION, '(no (definite)?evidence of|without)[a-zA-Z, ()0-9]*distant meta|'
|| 'distant metastasis[가-힣 a-zA-Z ]*관찰되지 않음', 'i') THEN 'No'
WHEN REGEXP_LIKE(R2.READCONCLUSION, 'distant meta', 'i') THEN 'Yes'
ELSE 'Unknown'
END
```

Figure 23. #22, #23 전신전이여부확인불가구분 데이터 추출 Query

## 8) 유방암\_검사\_생식세포변이검사 (BRCN\_EXAM\_GMVX)

- 주요 원천 테이블 : CDW.PE\_PAEX\_RSLT (병리검사결과 Table).
- 병리검사 중 NGS (Next-Generation Sequencing) 또는 Sanger Sequencing 검사의 데이터를 추출한다.

```
/*병리검사 중 NGS 와 Sanger 키워드를 포함하는 검사 종류 찾기 */
SELECT *
FROM CDW.CC_LIS_LPCMTEST /* 병리검사코드 Table */
WHERE REGEXP_LIKE(TESTENGM, 'NGS|SANGER', 'i')
```

ABC TESTCD	ABC TESTENGM	ABC TESTHNGNM
PMO13001	NGS Technology base Genetic Panel Test(Solid malignant tumor/Level II)	NGS차세대염기서열분석 기반 유전자패널검사(비유전성 유전자검사/고형암/Level II)
PMO13002	(NGS)CancerSCAN Level 1	(NGS)캔서스캔 레벨1
PMO13003	(NGS)CancerSCAN Level 2	(NGS)캔서스캔 레벨 2
PMO13004	(NGS)HemaSCAN	(NGS)헤마스캔
PMO13005	NGS-Solid tumor panel 1	비유전성 고형암 유전자 패널1검사
PMO13006	NGS-Solid tumor panel 2	비유전성 고형암 유전자 패널2검사
PMO13007	NGS-Lymphoma panel	악성림프종 유전자 패널검사
PMO13008	(위탁)NGS-Solid tumor panel 2(161)	(위탁)비유전성 고형암 유전자 패널2검사(161)
PMO13009	(위탁)NGS-Solid tumor panelII(523gene, RNA포함)	(위탁)비유전성 고형암 유전자 패널 II검사(523gene, RNA포함)
PMO13010	(위탁)[RNA 미포함]NGS-Solid tumor panel II(523gene,RNA미포함)	(위탁)[RNA 미포함]비유전성 고형암 유전자패널 II(523gene,RNA미포함)

Figure 24. NGS 검사 종류 데이터 확인

### A. 유방암 환자 대상의 검사 결과 데이터 추출.

- NGS관련 검사 결과 데이터를 CTE로 가져온다. 이는 추후 데이터 정제에 활용한다.

```
PAEX_T AS ( /* NGS 병리검사 결과 데이터 추출 CTE */
```

```
SELECT
```

```
PTSBSTNO
```

```
, ACPTDD /*접수일자*/
```

```
, SPCCD /*검체코드*/
```

```
, SPCNM /*검체명*/
```

```
, TESTCD /*검사코드*/
```

```
, TESTENGM/*검사영문명*/
```

```
, RSLTCNTS3/*판독결과*/
```

```
FROM CDW.PE_PAEX_RSLT
```

```
INNER JOIN PTINF
```

```
USING(PTSBSTNO)
```

```
WHERE ACPTDD BETWEEN '2010-01-01' AND '2022-12-31'
```

```
AND TESTCD IN ('PMO13002', 'PMO13003', 'PMO13004', 'PMO13008', 'PMO13009',
```

```
'PMO13010', 'PMO13001', 'PMO13007', 'PMO13005', 'PMO13006')
```

```
)
```

ABC SPCNM	ABC TESTCD	ABC TESTENGM	ABC RSLTCNTS3
breast	PMO13006	NGS-Solid tumor panel II (41303059)	<검체정보>†Specimen ID: S-2019-18599
brain	PMO13006	NGS-Solid tumor panel 2	<검체정보>†Specimen ID: S-2019-6245 A
ovary	PMO13008	(위탁)NGS-Solid tumor panel 2(161)	<검체정보>†Specimen ID: S-2019-10087
breast	PMO13003	(NGS)CancerSCAN Level 2	<검체정보>†장기/검체종류: Breast/Rese
breast	PMO13006	NGS-Solid tumor panel 2	<검체정보>†Specimen ID: S-2019-23606
breast	PMO13003	(NGS)CancerSCAN Level 2	<검체정보>†장기/검체종류: Breast/Rese
lung	PMO13009	(위탁)NGS-Solid tumor panelII(TMB/M	<검체정보>†Specimen ID: S-2021-8809 (
세포 lymph node, neck	PMO13007	NGS-Lymphoma panel	<검체정보>†Specimen ID: S-2021-10066

Figure 25. NGS 관련 검사 결과 데이터 추출 CTE

<검사정보>

검사방법 : Next Generation Sequencing (NGS)

검사패널: Solid Tumor Panel II

검사수탁 기관 : 녹십자

검사시행 기관 : 녹십자

<정도관리결과>

Sample (DNA) QC: PASS

Library QC: PASS

Sequencing QC: PASS

Mapped Reads: 11072935

On Target: 95.28 %

Mean Depth: 3220

Uniformity: 96.89 %

<Result>

Variants found (Tier I or II- clinically actionable alteration)

SNV/Indel

Tier (%)	Gene	RefSeq	AA	Nucleotide	*T.D.	VAF
2	TP53	NM_000546.5	p.Tyr234Cys	c.701A>G	2000	72

\*Total depth

Reference: 링크된 보고서 참조

CNV

Tier	Gene	CNV	Copy No. (5%- 95% CI)
2	CCNE1	amplification	4.79

Variants not found in essential genes as follow: ALK, BRAF, BRCA1, BRCA2, EGFR, ERBB2,

IDH1, IDH2, KIT, KRAS, MYC, MYCN, NRAS and PDGFRA

Figure 26. NGS 관련 검사 소견 예시

#### B. '#16 #17 #18 #19 #20 DNA변이정보' 데이터 추출

- DNA 변이정보 A값 도메인 : NUMERIC(10,3) - 소수를 포함하는(소수점 3자리까지) 수
- DNA 변이정보 B값 도메인 : A|G|C|T >
- DNA 변이정보 C값 도메인 : A | G | C | T | Ins | Del | Dup | Unknown
- DNA 변이정보 소견이 포함된 위치 : "Nucleotide" 소견 내용.  
- ex) "c.701A>G"에서 A값: 701, B값: A>, C값: G

#### C. '#21 #22 #23 #24 #25 아미노산변이정보' 데이터 추출

- 아미노산 변이정보 A값 도메인  
- Ala(A) | Arg(R) | Asn(N) | Asp(D) | Cys(C) | Gln(Q) | Glu(E) | Gly(G) | His(H) | Ile(I) | Leu(L) | Lys(K) | Met(M) | Phe(F) | Pro(P) | Ser(S) | Thr(T) | Trp(W) | Tyr(Y) | Val(V) | 기타
- 아미노산 변이정보 B값 도메인 : NUMERIC(10,3)
- 아미노산 변이정보 C값 도메인  
- Ala(A) | Arg(R) | Asn(N) | Asp(D) | Cys(C) | del | dup | fs(Frame shift) | Gln(Q) | Glu(E) | Gly(G) | His(H) | Ile(I) | ins | Leu(L) | Lys(K) | Met(M) | Phe(F) | Pro(P) | Ser(S) | Thr(T) | Trp(W) | Tyr(Y) | Val(V) | \*(Ter/Stop) | 기타
- 아미노산 변이정보 소견이 포함된 위치 : "AA" 소견 내용.  
- ex) "p.Tyr234Cys"에서 A값: Tyr(Y), B값: 234, C값: Cys(C)



#### D. '#26 참고순번' 데이터 추출

- 참고순번 도메인 : NM\_999999.9 (예시)
- 검사소견 중 "RefSeq" 항목의 결과를 추출한다.

#### E. 검사 소견 데이터 추출 방식.

- 하나의 검사 소견에 N개의 유전자 항목이 포함되어 있다. 이를 구분하면서 데이터를 추출하기 위해서는 유전자 항목마다의 SubQuery를 생성해야하며 이들을 UNION ALL로 결합해 하나의 Temporary Table을 만들어야 한다.

Tier (%)	Gene	RefSeq	AA	Nucleotide	*T.D.	VAF
2	BCL10	NM_003921	p.Thr91Hisfs*5	c.271del	632	43
3	HIST1H1E	NM_005321	p.Ala111Val	c.332C>T	115	63
3	KMT2D	NM_003482	p.Ser4251Phe	c.12752C>T	130	41

  

ABC GENE	ABC REFSEQ	ABC AA	ABC NUCLEOTIDE
BCL10	NM_003921	Thr91Hisfs*5	271del
HIST1H1E	NM_005321	Ala111Val	332C>T
KMT2D	NM_003482	Ser4251Phe	12752C>T

Figure 27. 검사한 유전자 항목 개별의 Row 데이터 추출 결과

```

SELECT PTSBSTNO, ACPTDD, TESTCD, TESTENGM, RSLTCNTS3
, 'BCL10' AS GENE /*항목*/
, LTRIM(REGEXP_SUBSTR(RSLTCNTS3, '(?<=BCL10)\s*NM_\d*(\.\d*)?')) AS REFSEQ
, REGEXP_SUBSTR(REGEXP_SUBSTR(RSLTCNTS3, 'BCL10\s*NM_\d*(\.\d*)?\s*p\.[a-zA-Z0-9*>_]*\s*c\.[a-zA-Z0-9*>_]*'),
'(?<=p\.)[a-zA-Z0-9*>_]*') AS AA
, REGEXP_SUBSTR(REGEXP_SUBSTR(RSLTCNTS3, 'BCL10\s*NM_\d*(\.\d*)?\s*p\.[a-zA-Z0-9*>_]*\s*c\.[a-zA-Z0-9*>_]*'),
'(?<=c\.)[a-zA-Z0-9*>_]*') AS NUCLEOTIDE
FROM PAEX_T
WHERE REGEXP_LIKE(RSLTCNTS3, 'BCL10\s*NM_\d*(\.\d*)?\s*p\.[a-zA-Z0-9*>_]*\s*c\.[a-zA-Z0-9*>_]*')
/* UNION ALL */
/*아래에 다른 유전자 검사 항목을 같은 방식으로 붙여서 결합한다. */

```

Figure 28. 특정 유전자 검사 항목(BCL10)의 소견 데이터 추출 Query

#### F. 추출 결과

- 각 수집 도메인에 맞게 CASE-WHEN-THEN 구문을 반영 후 아래와 같은 결과 테이블을 도출한다.

ABC GMTE_MTST_EXAM_RSLT_CONT	123 DNA_1	ABC	ABC DI	ABC DI	ABC DNA_VA	ABC AMSN	ABC	123 AMSN	ABC AM	ABC AMSN_1	ABC REF_SEQ
-GENE: BCL10 -DNA Change: 271del -AA Change: Thr91Hisfs*5	271	[NULL]	[NULL]	06	Del	17	Thr(T)	91	12	His(H)	NM_003921
-GENE: HIST1H1E -DNA Change: 332C>T -AA Change: Ala111Val	332	3	C>	04	T	01	Ala(A)	111	24	Val(V)	NM_005321
-GENE: KMT2D -DNA Change: 12752C>T -AA Change: Ser4251Phe	12752	3	C>	04	T	01	Ser(S)	4251	18	Phe(F)	NM_003482
-GENE: TP53 -DNA Change: 267delC -AA Change: Ser90Profs*33	267	[NULL]	[NULL]	06	Del	16	Ser(S)	90	19	Pro(P)	NM_000546.5
-GENE: PTEN -DNA Change: 499_500insA -AA Change: Thr167Asnfs*13	499	[NULL]	[NULL]	05	Ins	17	Thr(T)	167	03	Asn(N)	NM_000314.6
-GENE: WT1 -DNA Change: 1139G>A -AA Change: Arg380Gln	1139	2	G>	01	A	02	Arg(R)	380	09	Gln(Q)	NM_024426.4
-GENE: AKT1 -DNA Change: 49G>A -AA Change: Glu17Lys	49	2	G>	01	A	07	Glu(E)	17	16	Lys(K)	NM_001014431.1
-GENE: TP53 -DNA Change: 701A>G -AA Change: Tyr234Cys	701	1	A>	02	G	19	Tyr(Y)	234	05	Cys(C)	NM_000546.5

- "GMTE\_MTST\_EXAM\_RSLT\_CONT" Column 생성 Query

```

: '-GENE: '||GENE||' -DNA Change: '||NUCLEOTIDE||' -AA Change: '||AA
REF_SEQ 값까지 추가하면 캐시 테이블 용량 초과로 생성 불가.

```

## 9) 유방암\_검사\_유전자발현검사 (BRCN\_EXAM\_GNRX)

- 주요 원천 테이블 : CDW.PE\_PAEX\_RSLT (병리검사결과 Table).
- 검사 종류 도메인 : BCI | Endopredict | GenesWell BCT | Mammaprint | Oncofree | Oncotype Dx

<b>SELECT *</b> <b>FROM</b> CDW.CC_LIS_LPCMTST <b>WHERE</b> REGEXP_LIKE(TESTENGM, 'BCI Endopredict GenesWell BCT Mammaprint Oncofree Oncotype Dx' , 'i')		
ABC TESTCD	ABC TESTENGM	ABC TESTHNGNM
PMO02003	(위탁)Oncotype DX Breast Recurrence Score	(위탁)온코타입 디엑스 유방암 재발스코어
PMO02004	(위탁)GenesWell BCT	(위탁)조기 유방암 예후예측검사

Figure 29. 도메인 기준의 검사항목 코드 확인

- 원내에서는 관련 검사가 2가지 있는 것으로 확인. 해당 코드(TESTCD)와 매칭되는 검사 결과 데이터 추출.

### A. 유방암 환자 대상의 유전자발현 병리검사 소견 데이터 추출.

<b>SELECT</b> PTSBSTNO , ACPTDD /*접수일자*/ , SPCCD /*검체코드*/ , SPCNM /*검체명*/ , TESTCD /*검사코드*/ , TESTENGM /*검사영문명*/ , RSLTCNTS3 /*판독결과*/ <b>FROM</b> CDW.PE_PAEX_RSLT <b>INNER JOIN</b> PTINF <b>USING</b> (PTSBSTNO) <b>WHERE</b> ACPTDD <b>BETWEEN</b> '2010-01-01' <b>AND</b> '2022-12-31' <b>AND</b> TESTCD <b>IN</b> ('PMO02004', 'PMO02003')				
ABC SPCCD	ABC SPCNM	ABC TESTCD	ABC TESTENGM	ABC RSLTCNTS3
T01500	breast	PMO02003	(위탁)Oncotype DX Breast Recurrence Score	<검체정보>TSpec
T01500	breast	PMO02003	(위탁)Oncotype DX Breast Recurrence Score	<검체정보>T1 . Sp

Figure 30. 검사 소견 데이터(Row) 추출 결과 예시

```

<검사정보>
검사방법 : Oncotype DX
검사수탁 기관 : EXACT SCIENCES Genomic Health, Inc
검사시행 기관 : EXACT SCIENCES Genomic Health, Inc

<Result>
Recurrence Score: 15
Distant Recurrence Risk at 9 years: 4%
Group Average Absolute Chemotherapy (CT) Benefit: < 1%
  
```

Figure 31. 'Oncotype DX' 검사 소견 원문 예시

### B. '#9 유전자발현검사결과내용' 데이터 추출

- 검사 종류마다 결과 소견이 다르다. Oncotype DX의 경우 'Oncotype Recurrence Score (RS)'.
- Oncotype DX의 경우, 정규표현식으로 "Recurrence Score: 15"에서 "15"만 추출할 수 있도록 작성.

```
LTRIM(REGEXP_SUBSTR(RSLTCNTS3, '(?<=Recurrence Score[\\:;])\\s\\d{1,3}'))
```



## 10) 유방암\_치료\_수술 (BRCN\_TRTM\_OPRT)

- 주요 원천 테이블 : CDW.PE\_OP\_LIST (수술내역 Table).

### A. '#30 #31 수술종양위치' 데이터 추출.

- 도메인 : Left | Right | Bilateral | 기타
  - 수술 시 확인한 종양의 위치이므로 종양의 주제와는 무관하다.

```
Malignant neoplasm of breast unspecified, left
Rt. breast ca
Malignant neoplasm of upper-inner quadrant of breast, right
bilateral breast ca
```

Figure 32. 수술 시 종양의 위치 데이터 예시

```
LEFT OUTER JOIN ( /*수술부위 데이터 수집 JOIN*/
SELECT
    PTSBSTNO
    , ORDDD
    , RTRIM(CASE WHEN REGEXP_LIKE(GRP_TMP, 'breast', 'i') THEN 'Breast|' ELSE '' END||
    CASE WHEN REGEXP_LIKE(GRP_TMP, '\bLN|lymph', 'i') THEN 'LN|' ELSE '' END||
    CASE WHEN REGEXP_LIKE(GRP_TMP, '\bliver', 'i') THEN 'Liver|' ELSE '' END||
    CASE WHEN REGEXP_LIKE(GRP_TMP, '\blung', 'i') THEN 'Lung|' ELSE '' END||
    CASE WHEN REGEXP_LIKE(GRP_TMP, 'stomach', 'i') THEN 'Stomach|' ELSE '' END||
    CASE WHEN REGEXP_LIKE(GRP_TMP, 'Thyroid', 'i') THEN 'Thyroid|' ELSE '' END||
    CASE WHEN REGEXP_LIKE(GRP_TMP, 'colon', 'i') THEN 'Colon|' ELSE '' END||
    CASE WHEN REGEXP_LIKE(GRP_TMP, 'cervic', 'i') THEN 'Cervical|' ELSE '' END, '|')
AS GRP_CONT
    , NVL(REGEXP_SUBSTR(GRP_TMP, 'Bilateral', 1,1, 'i'),
    REGEXP_SUBSTR(GRP_TMP, 'LT\.|LEFT', 1,1, 'i'),
    REGEXP_SUBSTR(GRP_TMP, 'RT\.|RIGHT', 1,1, 'i'), '') AS LOCA_CONT
FROM (
    SELECT
        PTSBSTNO
        , ORDDD
        , LISTAGG(RECTERMNM, '|') AS GRP_TMP
    FROM
        CDW.MR_FORM_REC
    INNER JOIN PTINF P
    USING(PTSBSTNO)
    WHERE
        /*수술명, Tumor 위치, 진단명, */
        DEGNITEMLEVLCD IN (
            '0000509.0000509.0000509',
            '0000512.0000917',
            '0000521.0000960', '0000521.0000960.0000960',
            '0000960.0000960', '0000960.0000960.0000960',
            '0000933.0000509.0000509',
            '0000931.0000931.0000931',
            '0011529.0000960.0000960',
            '0007983.0000960',
            '0011449.0000960.0000960',
            '0000512.0000359'
        )
        AND RECTERMNM IS NOT NULL
        AND REGEXP_LIKE(RECTERMNM, 'breast|lymph|\bLN|\bliver|\blung|stomach|thyroid|colon|'
        ||'prostate|cervic|\b[lr]t|left|right|bilateral', 'i')
    GROUP BY PTSBSTNO, ORDDD
    ) TMP
) B6
ON A.PTSBSTNO = B6.PTSBSTNO
AND A.INDD = B6.ORDDD
```

```

CASE
WHEN REGEXP_LIKE(B6.LOCA_CONT, 'LT|LEFT', 'i') THEN '1' /*Left*/
WHEN REGEXP_LIKE(B6.LOCA_CONT, 'RT|RIGHT', 'i') THEN '2' /*Right*/
WHEN REGEXP_LIKE(B6.LOCA_CONT, 'Bilateral', 'i') THEN '3' /*Bilateral*/
END AS OPRT_TUMR_LOCA_CD /*#30 수술종양위치코드*/

```

Figure 33. #30, #31 수술종양위치 데이터 추출 Query

## B. '#32 #33 수술종양상세위치' 데이터 추출.

- 도메인 : Central | Lower inner | Lower outer | Upper inner | Upper outer | 기타

환자의 Rt breast upper outer quadrant 에 약 3.0 x 3.0 cm 크기로 necrotic change 동반된  
상태가 확인되었으며, eschar를 debridement 한 뒤 동반되는 결손부위를 limberg flap으로 coverage  
Lt.10H mass (main tumor에서 upper inner 방향): benign  
op.cavity의 inner & lower breast tissue를

```

LEFT OUTER JOIN ( /* 수술종양상세위치 데이터 수집 JOIN*/
SELECT
PTSBSSTNO
,
OPRSRVNO
,
LISTAGG(NVL(CAST(RECCNTS AS VARCHAR), ''), '|') AS GRP_CONT
FROM
CDW.MR_FORM_REC
INNER JOIN PTINF P
USING(PTSBSSTNO)
WHERE
DEGNITEMLEVLCD IN ('0000507.0000146', '0000053', '0000507') /*수술결과, 수술과정, 수술소견 데이터*/
AND OPRSRVNO IS NOT NULL
AND RECCNTS IS NOT NULL
AND REGEXP_LIKE(RECCNTS, 'CENTRAL|bINNER|bOUTER|UPPER (OUTER|INNER)|LOWER (INNER|OUTER)|UPPER|LOWER', 'i')
GROUP BY PTSBSSTNO, OPRSRVNO
) B9
ON A.PTSBSSTNO = B9.PTSBSSTNO
AND A.OPRSRVNO = B9.OPRSRVNO

CASE
WHEN REGEXP_LIKE(B9.GRP_CONT, 'CENTRAL', 'i') THEN '01' /*Central*/
WHEN REGEXP_LIKE(B9.GRP_CONT, 'Lower inner|inner\s?(\, \&)\s?lower', 'i') THEN '02' /*Lower inner*/
WHEN REGEXP_LIKE(B9.GRP_CONT, 'Lower outer|outer\s?(\, \&)\s?lower', 'i') THEN '03' /*Lower outer*/
WHEN REGEXP_LIKE(B9.GRP_CONT, 'Upper inner|inner\s?(\, \&)\s?upper', 'i') THEN '04' /*Upper inner*/
WHEN REGEXP_LIKE(B9.GRP_CONT, 'Upper outer|outer\s?(\, \&)\s?upper', 'i') THEN '05' /*Upper outer*/
ELSE '99' /*기타*/
END AS OPRT_TUMR_DETL_LOCA_CD /*#32 수술종양상세위치코드*/

```

Figure 34. #32, #33 수술종양상세위치 데이터 추출 Query

### C. '#34 #35 #36 #37 재건술분류/방법' 데이터 추출.

- 분류 도메인 : Immediate reconstruction | Delayed reconstruction | Oncoplastic | Not applicable | 기타
- 방법 도메인 : Advanced glandular | Direct to implant | LD flap | Tissue expander | TRAM/DIEP | Vicryl mesh

```
Rt breast reconstruction c glandular flap & mastopexy
breast reconstruction c silicone implant insertion
Lt breast reconstruction c tissue expander insertion
우측 latissimus dorsi myocutaneous flap elevation을 위하여
```

Figure 35. 소견 내용에서 재건술 방법 데이터 확인

- 서식내용 (CDW.MR\_FORM\_REC) 중에서 'Immediate Reconstruction' 여부를 확인하는 항목이 별도로 존재.
- 값이 'Yes'일 경우 재건술 방법 항목까지 같이 확인 가능하다.

ABC DEGNITEMLEVLCD	ABC RECTERMNM	ABC DEGNITEMNM
0000507.0009608	Yes	Immediate reconstruction
0000507.0009608.0024517	TRAM	If yes

Figure 36. Immediate reconstruction 여부와 재건술 방법 데이터 확인

- 수술소견 등의 데이터에서는 정규표현식으로 키워드를 찾아내며, 서식 항목으로 존재하는 경우는 별도의 JOIN을 반영해 데이터를 확인할 수 있도록 SQL 작성

```
LEFT JOIN ( /* 재건술분류 데이터 수집 목적의 JOIN*/
SELECT
PTSBSSTNO
, OPRSRVNO
, LISTAGG(
NVL(REGEXP_SUBSTR(RECCNTS, '(immediate (breast ))?reconstruction', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, 'Oncoplastic', 1,1, 'i'), '')||'+'||
NVL(
REGEXP_SUBSTR(RECCNTS, 'glandular', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, 'latissimus dorsi', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, 'implant', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, '\bTRAM\bDIEP', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, '\bLD( flap)?', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, '(Tissue )?expander', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, 'STAPLER', 1,1, 'i'),
REGEXP_SUBSTR(RECCNTS, 'Vicryl(\s?(d(-d)?))?', 1,1, 'i'),
), '|') AS GRP_CONT
, LISTAGG(CAST(RECCNTS AS VARCHAR)) AS T
FROM
CDW.MR_FORM_REC
INNER JOIN PTINF P
USING(PTSBSSTNO)
WHERE
DEGNITEMLEVLCD IN ('0000045', '0000507.0000146', '0000053', '0000507')
AND OPRSRVNO IS NOT NULL
AND RECCNTS IS NOT NULL
AND REGEXP_LIKE(RECCNTS, 'breast', 'i')
AND REGEXP_LIKE(RECCNTS, 'reconstruction|latissimus dorsi|oncoplastic', 'i')
GROUP BY PTSBSSTNO, OPRSRVNO
) B7
ON A.PTSBSSTNO = B7.PTSBSSTNO
AND A.OPRSRVNO = B7.OPRSRVNO
LEFT JOIN ( /*Immediate Reconstruction 시행 여부*/
SELECT
PTSBSSTNO, OPRSRVNO
, 1 AS IM_CONT_YES /*WHERE 조건에서 'Yes'를 확인하므로 추출된 결과는 1로 고정.
```

<pre> FROM     CDW.MR_FORM_REC INNER JOIN PTINF P USING(PTSBSTNO) WHERE     DEGNITEMLEVLCD = '0000507.0009608' /*Immediate Reconstruction Yes 여부 확인 */     AND RECTERMNM = 'Yes' GROUP BY PTSBSTNO, OPRSRVNO ) B77 ON A.PTSBSTNO = B77.PTSBSTNO AND A.OPRSRVNO = B77.OPRSRVNO LEFT JOIN ( /*재건술 방법 데이터 추가 추출*/ SELECT     PTSBSTNO     , OPRSRVNO     , LISTAGG(RECTERMNM) AS REPR_CONT FROM     CDW.MR_FORM_REC INNER JOIN PTINF P USING(PTSBSTNO) WHERE     DEGNITEMLEVLCD = '0000507.0009608.0024517' /*Immediate Reconstruction If Yes */ GROUP BY PTSBSTNO, OPRSRVNO ) B777 ON A.PTSBSTNO = B777.PTSBSTNO AND A.OPRSRVNO = B777.OPRSRVNO CASE     WHEN REGEXP_LIKE(B7.GRP_CONT, 'immediate', 'i') THEN '1' /*Immediate reconstruction*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'reconstruction') THEN '2' /*Delayed reconstruction*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'Oncoplastic') THEN '3' /*Oncoplastic*/     WHEN B77.IM_CONT_YES = 1 THEN '1'     ELSE '4' /*Not applicable*/ END AS REPR_CLCD /*#34 재건술분류코드*/  CASE     WHEN REGEXP_LIKE(B7.GRP_CONT, 'glandular', 'i') THEN '01' /*Advanced glandular*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'implant', 'i') THEN '02' /*Direct to implant*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'sewing', 'i') THEN '03' /*Hand-sewing*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'LATISSIMUS DORSI'    '\bLD( flap)?', 'i')         THEN '04' /*LD flap*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'Stapler', 'i') THEN '06' /*Stapler*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'expander', 'i') THEN '07' /*Tissue expander*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'TRAM DIEP', 'i') THEN '08' /*TRAM/DIEP*/     WHEN REGEXP_LIKE(B7.GRP_CONT, 'Vicryl', 'i') THEN '09' /*Vicryl mesh*/     WHEN B777.REPR_CONT = 'Direct implant' THEN '02'     WHEN B777.REPR_CONT = 'LD flap' THEN '04'     WHEN B777.REPR_CONT = 'Tissue expander' THEN '07'     WHEN B777.REPR_CONT = 'TRAM' THEN '08'     ELSE '05' /*Not applicable*/ END </pre>
--

Figure 37. #34, #35, #36, #37 재건술 분류/방법 데이터 추출 Query

#### D. ‘#38 #39 감시림프절검출방법’ 데이터 추출.

- 도메인 : Dye | ICG | Indigocarmine | Radioisotope | 기타
- 도메인 설명 : 감시 림프절 생검술 시 node identification을 위해 사용된 방법(Sentinel node detection method).
- 감시 림프절(Sentinel Node) 여부가 ‘Yes’면 나타나는 검사 방법 항목 확인.

ABC DEGNITEMLEVLCD	ABC RECTERMNM	RECCNTS	ABC DEGNITEMNM
0000507.0000440	Yes	[NULL]	Sentinel Node Biopsy
0000507.0000440.0009600	Radio-isotope(Agent: TC-tin-colloid)	[NULL]	Radio-isotope(Agent:
0000507.0000440.0002714	[NULL]	08:30	Injection time
0000507.0000440.0015247	Vital Dye	[NULL]	Vital Dye
0000507.0000440.0000627	[NULL]	indigo carmine	Agent

Figure 38. 감시림프절검출 관련 서식 항목 확인

```

LEFT OUTER JOIN ( /*감시 림프절 생검술 시 node identification 을 위해 사용된 방법 명 JOIN*/
SELECT
    PTSBSTNO
    , OPRSRVNO
    , LISTAGG(NVL(CAST(RECCNTS AS VARCHAR), RECTERMNM, ''), '|') AS GRP_CONT
FROM CDW.MR_FORM_REC
INNER JOIN PTINF P
USING(PTSBSTNO)
WHERE
    DEGNITEMLEVELCD IN ('0000507.0000440.0000627', /*Agent*/
                        '0000507.0000440.0009600', /*Radio-isotope*/
                        '0000507.0000440.0015247') /*Vital DYE*/
    AND NOT (RECCNTS IS NULL AND RECTERMNM IS NULL)
GROUP BY PTSBSTNO, OPRSRVNO
) B8
ON A.PTSBSTNO = B8.PTSBSTNO
AND A.OPRSRVNO = B8.OPRSRVNO

CASE
WHEN REGEXP_LIKE(B8.GRP_CONT, 'INDI(G)?0', 'i') THEN '3' /*Indigocarmine*/
WHEN REGEXP_LIKE(B8.GRP_CONT, 'RADIO\-', 'i') THEN '4' /*Radioisotope*/
WHEN REGEXP_LIKE(B8.GRP_CONT, 'DYE', 'i') THEN '1' /*Dye*/
WHEN B8.GRP_CONT IS NOT NULL THEN '9' /*기타*/
END

```

Figure 39. #38 #39 감시림프절검출방법 데이터 추출 Query

## 11) 유방암\_검사\_외과병리 (BRCN\_EXAM\_SGPT)

- 주요 원천 테이블 : CDW.PE\_PAEX\_RSLT (병리검사결과) Table.
- 주요 데이터 : 테이블 내의 'RSLTCNTS3' Column에 결과 소견을 저장한다.

```

/* 유방암 대상 환자의 외과병리 검사 소견 데이터 추출을 위한 CTE 구성 */
WITH PTINF AS (
SELECT
    PT_SBST_NO AS PTSBSTNO
FROM LCD.PT_CNRG_TRGT
WHERE 1=1
    AND CKND_CLSF_CD = 'CD0002' /*유방암코드*/
    AND FRDX_YMD BETWEEN '2010-01-01' AND '2022-12-31' /*최초진단일시 2010~2022 년*/
),
A1 AS (
SELECT PTSBSTNO, TO_CHAR(ACPTDD, 'YYYYMMDD') AS ACPTDD, READD, LOWER(RSLTCNTS3) AS RSLTCNTS3
    , RSLTCNTS3 AS MAINR
    , CASE
        WHEN REGEXP_LIKE(REGEXP_SUBSTR(RSLTCNTS3, 'Blocks[\:\;\;][\s\S]*?\n'), 'tumor', 'i') THEN 'Y'
        WHEN REGEXP_LIKE(RSLTCNTS3, 'tumor (site|size)|carcinoma|size of tumor', 'i') THEN 'Y'
        WHEN REGEXP_LIKE(RSLTCNTS3, '(no|neurogenic|residual) tumor|ulcer|abscess|no visible tumor', 'i') THEN 'N'
        ELSE 'N'
    END AS CAYN /*암종이 있는 소견만 추출하기 위함*/
FROM CDW.PE_PAEX_RSLT /*병리검사결과 TABLE*/
INNER JOIN PTINF
USING(PTSBSTNO)
WHERE 1=1
    AND INSTCD = '101'
    AND READD IS NOT NULL
    AND TESTCD IN ( 'PSU00016', 'PSU00018', 'POU15650', 'POU15660', 'POU15670', 'PSU56050', 'PSU56051', 'PSU56060'
        , 'PSU56061', 'PSU56070', 'PSU56071' )
    AND ACPTDD BETWEEN '2010-01-01' AND '2022-12-31'
)

```

Figure 40. 외과병리 검사 소견 데이터 추출 CTE

**A. '#6 #7 #8 #9 외과병리체취부위, 검사부위편측성' 데이터 추출.**

- 체취부위 도메인 : Breast | LN | 기타
- 검사부위편측성 도메인 : Left | Right | Bilateral | 기타

```
diagnosis:
  breast, left, mastectomy:
    invasive ductal carcinoma
  lymph node, axillar, left, biopsy:
    no metastasis (0/2)

diagnosis:
  breast, skin near tumor, right, biopsy:
```

Figure 41. 체취부위, 검사부위편측성 데이터 확인

- 소견 내의 'diagnosis' 부분에 값이 기록되어 있으므로 해당 부분을 정규표현식으로 필터링 수행.

```
TRIM(REGEXP_SUBSTR(TRIM(REGEXP_SUBSTR(RSLTCNTS3,
'(?<=diagnosis\s?:)\s*([^\s:]+)')), '(?<=(\n|\s)).*'))
```

Figure 42. diagnosis 소견 부분 필터링 정규표현식

**A. '#19 #20 림프혈관침윤여부' 데이터 추출.**

- 관련 키워드 : Lymphovascular invasion 여부

```
lymphovascular invasion: absent
```

```
REGEXP_SUBSTR(RSLTCNTS3, 'lymphovascular invasion\s?[\:\;]\s{1,3}(absent|present|not)', 1,1, 'i')
```

Figure 43. lymphovascular invasion 관련 필터링 정규표현식

**B. '#21 #22 림프절주위침윤여부' 데이터 추출.**

- 관련 키워드 : Perinodal extension 여부
- 현재 원내 소견에서는 perinodal extension에 대한 내용 확인 불가.
- extranodal extension 소견으로 반영.

```
extranodal extension: no
```

```
REGEXP_SUBSTR(RSLTCNTS3, 'perineural invasion\s?[\:\;]\s{1,3}(absent|present|not|([+|-]))|'
|| '(\s|\s?)\s?perineural|no perineural|'
|| 'extranodal extension\s{1,3}(present|absent|no)|'
|| 'with extranodal extension', 1,1, 'i')
/*신경주위(perineural) 림프절외(extranodal) 침윤여부 확인 */
```

Figure 44. perineural, extranodal 침윤 관련 필터링 정규표현식

C. '#23 #24 유두침윤존재여부' 데이터 추출.

- 관련 키워드 : Nipple invasion 여부
- Nipple invasion의 명확한 소견 확인 불가. Nipple paget's disease 여부 내용을 추출한다.

nipple paget's disease: present

```
REGEXP_SUBSTR(RSLTCNTS3, 'nipple paget(''|')s disease\s?[\:|;]\s?(present|absent|not)')
```

Figure 45. nipple paget's disease 여부 필터링 정규표현식

D. '#25 #26 미세석회화여부' 데이터 추출.

- 관련 키워드 : Microcalcification 여부

microcalcification: present (intratumoral)

```
REGEXP_SUBSTR(RSLTCNTS3, 'Microcalcification(s)?\s?[\:|;]\s{0,3}(present|absent|not)|'
|| '(with|foci of|multiple|change|focal|few|several) microcalcification|'
|| 'microcalcification (around|is observ)', 1,1, 'i')
```

Figure 46. Microcalcification 여부 필터링 정규표현식

E. '#27 #28 상피내암괴사여부' 데이터 추출.

- 관련 키워드 : necrosis 여부

tumor necrosis (nc): yes | necrosis: present

```
REGEXP_SUBSTR(RSLTCNTS3, 'with(out)?( comedo| central| (multi)?focal| caseous)? necrosis|'
|| 'necrosis( \s20\~3962\))?(\:|;)\s?(present|absent|not)|'
|| 'show(s)? fat necrosis|'
|| 'tumor necrosis \s(nc)\s?(\:|;)\s?(no|yes)', 1,1, 'i')
```

Figure 47. Necrosis 여부 필터링 정규표현식

F. '#29 #30 상피내암동반여부' 데이터 추출.

- 관련 키워드 : EIC (Extensive Intraductal Component) 여부

eic: present 99 % | eic: absent 1 % | eic: not applied

```
REGEXP_SUBSTR(RSLTCNTS3, '(?<=\.\seic).*', 1,1, 'i')
```

Figure 48. EIC 여부 필터링 정규표현식

#### G. '#31 #32 상피내암핵등급' 데이터 추출.

- 도메인 : G1 | G2 | G3 | Gx | High | Intermediate | Low | 기타
- Histologic grade(score) 또는 Nuclear grade 소견 데이터를 참조한다.

```
4. histologic grade (nottingham histologic score)
   grade 3 (tf: 3 np: 3 mc: 3)

histologic grade 1 (nottingham histologic score)

nuclear grade : 11 (intermediate)

g3 nuclei very irregular,
```

```
NVL(REGEXP_SUBSTR(RSLTCNTS3, 'nuclear grade [1-9il]+', 1,1, 'i'),
REGEXP_SUBSTR(RSLTCNTS3, 'total nottingham score(\;|\:)\s?.*', 1,1, 'i'),
REGEXP_SUBSTR(RSLTCNTS3,
'nuclear (grade|pleomorphism)\s?(\(s20\~3962\))?\s?(\:|\;)\s?.*', 1,1, 'i'),
REGEXP_SUBSTR(RSLTCNTS3,
'histologic grade( \(\nottingham histologic score\))?\s*(grade\s{1,2}[1-9il]+|not (grade|app))',
1,1, 'i'),
REGEXP_SUBSTR(RSLTCNTS3,
'histologic grade( \(\nottingham histologic score\))?\(\:|\;)\s?.*', 1,1, 'i'),
REGEXP_SUBSTR(RSLTCNTS3, 'g[1-5]\:\s?nuclei', 1,1, 'i'),
REGEXP_SUBSTR(RSLTCNTS3, 'grade [1-9il]+', 1,1, 'i')
)
```

Figure 49. 상피내암핵등급 데이터 필터링 정규표현식

#### H. '#35 #36 수술절제면잔존' 데이터 추출.

- 도메인 : R0 | R1 | R2 | 기타
- 수술 절제면 관련 R0, R1, R2 소견 기준 참고.
- R0 : No residual tumor
- R1 : Microscopic residual tumor
- R2 : Macroscopic residual tumor (The area of primary tumor, regional lymph nodes, distant site)

```
a few small residual invasive lesion (0.3 cm) and dcis foci
two microscopic residual dcis lesions around previous exci
no residual invasive cancer
```

```
REGEXP_SUBSTR(RSLTCNTS3, '(no|without) residual( tumor)?|'|'microscopic residual dcis|'
|'|'residual tumor present|'|'residual( solid)? dcis|'
|'|'lymph node(s)? show residual tumor|(situ|carcinoma), residual|'
|'|'(small )?residual (invasive|infil)|'
|'|'residual micro|(microinvasive|microscopic) residual carcinoma', 1,1, 'i')
```

Figure 50. residual 관련 내용 필터링 정규표현식



#### I. '#37 전이암크기' 데이터 추출.

- 관련 키워드 : Size of metastasis

```
size of largest metastasis 1 cm,  
[metastatic tumor size, 0.7 cm; extranodal tumor extension]
```

```
REGEXP_SUBSTR(RSLTCNTS3, 'size of ( largest)? metastasis \d{1,2}\s?(cm|mm)|'  
|| '\d{1,2}(\.\d{1,2})?(x\d{1,2}(\.\d{1,2})?)?\scm\)' with metastasis|'  
|| 'metastatic tumor size, \d{1,2}(\.\d{1,2})? cm', 1, 1, 'i')
```

Figure 51. 전이암크기 데이터 필터링 정규표현식

#### J. '#40 #41 잔여종양' 데이터 추출.

- 관련 키워드 : Residual tumor 내용
- 도메인 : No residual | CIS only | Residual breast tumor | Residual LN meta | Residual breast tumor and LN meta
- 정규표현식은 'H.'의 수술절제면잔존과 같은 것으로 반영.

```
the lymph nodes show residual tumor with necrosis  
excisional biopsy: no residual tumor  
residual dcis lesions surrounding previous excision
```

#### K. '#44 #45 #46 #47 감시|비감시 림프절 수' 데이터 추출.

- 관련 키워드 : Sentinel Lymph(감시 림프), Non-sentinel Lymph(비감시 림프)

```
regional lymph nodes (pn0); (0/1)      11. lymph node ( 5 / 12 )  
sentinel (0/1)                        sentinel l ln: 1 / 1 (macro)  
                                       sentinel ll ln: 1 / 1 (macro)  
                                       non-sentinel lns: 3 / 10  
  
11. lymph node (4/13) (pn2)  
sentinel l ln: 4/5 (macro)  
non-sentinel lns: 0/8  
  
11. lymph node (0/5) (pn0)  
sentinel ln: 0 / 5
```

/\*감시 림프절 수 데이터 추출\*/

```
REGEXP_SUBSTR(RSLTCNTS3, 'SENTINEL( L+)? LN(S)?\:\s{1,3}\d{1,2}\s?\s?\s?\d{1,2}|'  
|| 'SENTINEL\s?\s?\s?\d{1,2}\s?\s?\s?\d{1,2}', 1, 1, 'i')
```

/\* 비감시 림프절 수 데이터 추출\*/

```
REGEXP_SUBSTR(RSLTCNTS3, 'NON-SENTINEL( L+)? LN(S)?\:\s{1,3}\d{1,2}\s?\s?\s?\d{1,2}',  
1, 1, 'i')
```

Figure 52. Sentinel, Non-sentinel Lymph 수 데이터 필터링 정규표현식

## L. '#49 #50 #51 #52 TNM 병기 값' 데이터 추출.

- 관련 키워드: Pathologic TNM stage

### • T병기 분류 기준 (Primary Tumor)

- Tis : 소엽 또는 유관 상피내암 (lobular carcinoma, DCIS)
- T1 : 최대 직경 2cm 이하 (  $\leq 2\text{cm}$  )
- T2 : 최대 직경 2초과 5이하 (  $2\text{cm} < X \leq 5\text{cm}$  )
- T3 : 최대 직경 5cm 초과 (  $> 5\text{cm}$  )
- T4 : 흉벽이나 유방의 피부를 침범한 경우
- 

### • N병기 분류 기준 (Regional Lymph Node)

- N0 : 림프절 전이 없음
- N1 : 전이 림프절 수 1~3개
- N2 : 전이 림프절 수 4~9개
- N3 : 전이 림프절 수 10개 이상

- M병기(Distant Metastasis)는 소견에서 명확히 나타나지 않아 생략 - 'Mx'로 대체.

```
microscopic tumor size (pt2)
histologic type: (ptis)
ductal carcinoma in situ
depth of tumor invasion: se, pt4
breast, left, mastectomy: invasive lobular carcinoma
```

Figure 53. T병기 소견 데이터 예시

#### ① T 병기 소견 정규표현식

```
REGEXP_SUBSTR(RSLTCNTS3, 'pt\s?(is|0|1|2|3|4)|no tumor|'
|| '(no residual )?ductal carcinoma in situ|'
|| 'invasive (ductal|lobular) carcinoma', 1,1, 'i')
```

Figure 54. T병기 데이터 추출 정규표현식

```
'T' || CASE
  WHEN REGEXP_LIKE(B1.STAG_T, 'no (residual|tumor)', 'i') THEN '0'
  WHEN REGEXP_LIKE(B1.STAG_T, '(ductal|lobular) carcinoma', 'i') THEN 'is'
  WHEN REGEXP_LIKE(B1.STAG_T, 'pt\s?(is|0|1|2|3|4)', 'i')
  THEN TRIM(REGEXP_SUBSTR(B1.STAG_T, '(?<=pt)\s?(is|0|1|2|3|4)'))
  ELSE 'x'
END
||
'N' || CASE
  WHEN B1.PSTV_LN_CNT = 0 THEN '0'
  WHEN B1.PSTV_LN_CNT >= 10 THEN '3'
  WHEN B1.PSTV_LN_CNT >= 4 THEN '2'
  WHEN B1.PSTV_LN_CNT >= 1 THEN '1'
  ELSE 'x'
END
|| 'Mx'
```

Figure 55. TNM 병기 데이터 결합 Query

## 12) 유방암\_치료\_약제성분 (BRCN\_TRTM\_DRIN)

• 주요 원천 테이블 : CDW.PE\_ORD\_LIST (처방정보 Table).

• 품질진단 시 중점으로 보는 '약제성분'에 대한 정리.

- 세포독성, 항암표적, 항암면역, 기타항암의 분류로 나뉘며 성분 분류의 근거로 '약학정보원'의 '항암제' 문서를 참고한다.

### ▶ 세포독성약제

Azacitidine	Ifosfamide	Bendamustine	Melphalan	Cisplatin	Carboplatin
Dacarbazine	Oxaliplatin	Busulfan	Temozolomide	Fluorouracil	Capecitabine
Doxifluridine	Tegafur	Enocitabine	Clofarabine	Cladribine	Cytarabine
Decitabine	Pemetrexed	Pralatrexate	Doxorubicin	Fludarabine	Cyclophosphamide
Irinotecan	Idarubicin	Epirubicin	Daunorubicin	Etoposide	Methotrexate
Topotecan	Paclitaxel	Docetaxel	Vinblastine	Cabazitaxel	Mercaptopurine
Vincristine	Vinorelbine	Bleomycin	Hydroxyurea	Mitomycin C	Mitoxantrone

### ▶ 표적항암제

Axitinib	Alemtuzumab	Alectinib	Cetuximab	Bevacizumab	Brentuximab
Rituximab	Ibritumomab	Elotuzumab	Erlotinib	Gefitinib	Vandetanib
Apatinib	Lapatinib	Pazopanib	Sunitinib	Sorafenib	Regorafenib
Imatinib	Dasatinib	Nilotinib	Ceritinib	Crizotinib	Trastuzumab

### ▶ 면역항암제

Atezolizumab	Blinatumomab	Pembrolizumab	Nivolumab	Ipilimumab
--------------	--------------	---------------	-----------	------------

### ▶ 기타항암제

Anastrozole	Bicalutamide	Exemestane	Enzalutamide	Fulvestrant
Tamoxifen	Letrozole			

- 데이터 수집은 대장암 라이브러리와 동일하므로 설명 생략.

### 13) 유방암\_치료\_항암요법 (BRCN\_TRTM\_ANTP)

- 관리하는 별도의 테이블이 없으므로 서식기록내역 (CDW.MR\_FORM\_REC) Table에서 수집한다.
- 서식분류 코드 중 FORMCD = '1200005394'는 '유방암 환자 항암화학요법 기록지 1'이며 '1200005398'은 '유방암 환자 항암화학요법 기록지 2'다. 이 두 서식의 내용을 바탕으로 SQL을 작성한다.

ABC DEGNITEMLEVLCD	RECCNTS	ABC RECTERMNM	ABC DEGNITEMNM
0002076	[NULL]	Adjuvant	Chemotherapy type
0002965.0003124.0000627	[NULL]	[NULL]	Drug
0002965.0003124.0000627.0012276	5-FU	[NULL]	약명
0002965.0003124.0000627.0003124	980mg	[NULL]	용량
0002965.0003124.0000627.0005781	iv	[NULL]	용법
0002965.0003124.0000627.0012276	Epizin	[NULL]	약명
0002965.0003124.0000627.0003124	98mg	[NULL]	용량
0002965.0003124.0000627.0005781	iv	[NULL]	용법
0002965.0003124.0000627.0012276	Endoxan	[NULL]	약명
0002965.0003124.0000627.0003124	980mg	[NULL]	용량
0002965.0003124.0000627.0005781	iv	[NULL]	용법
0002965	[NULL]	[NULL]	Scheduled Regimen
0002965.0002965	FEC #6 cycles	[NULL]	Regimen
0002965.0002076	2013-09-17	[NULL]	항암치료 시작일

Figure 56. 유방암 환자 항암화학요법 기록지 1 서식 내용 일부 (1200005394)

ABC DEGNITEMLEVLCD	RECCNTS	ABC RECTERMNM	ABC DEGNITEMNM
0001931	[NULL]	1. 일상적 생활가능	ECOG Scale Performance Status
0000358	[NULL]	Anorexia	부작용(해당사항에 √해 주세요.)
0000358	[NULL]	Nausea	부작용(해당사항에 √해 주세요.)
0002965	AC	[NULL]	Used Regimen
0002965.0011432	AC 2	[NULL]	Regimen type and Cycle
0002965.0003467	2013-05-14	[NULL]	항암제 투약일
0002965.0003489	[NULL]	No	약제 용량 변경
0000627	[NULL]	[NULL]	Drug
0000627.0012276	Doxorubin	[NULL]	약명
0000627.0003124	90mg	[NULL]	용량
0000627.0012276	Endoxan	[NULL]	약명
0000627.0003124	900mg	[NULL]	용량
0018570	[NULL]	[NULL]	Prechemo Tx. Lab
0018570.0000333	6400	[NULL]	WBC
0018570.0002912	3200	[NULL]	ANC
0018570.0002551	413000	[NULL]	PLT
0018570.0002885	24	[NULL]	OT/PT
0018570.0000283	26	[NULL]	/

Figure 57. 유방암 환자 항암화학요법 기록지 2 서식 내용 일부 (1200005398)

#### A. 주요 서식항목 내용 가져오는 CTE 작성.

<pre> CTX1 AS ( SELECT PTSBSTNO /*유방암 환자 항암화학요법 기록지 1 */ , ORDDD /*환자대체번호*/ , DEGNITEMLEVLCD /*디자인서식레벨코드*/ , RECCNTS /*기록내용 1*/ , RECTERMNM /*기록내용 2*/ , DEGNITEMNM /*기록항목(주제)*/ FROM CDW.MR_FORM_REC INNER JOIN PTINF USING(PTSBSTNO) WHERE FORMCD = '1200005394' ) </pre>	<pre> CTX2 AS (/*유방암 환자 항암화학요법 기록지 2 */ SELECT PTSBSTNO , ORDDD , DEGNITEMLEVLCD , RECCNTS , RECTERMNM , DEGNITEMNM FROM CDW.MR_FORM_REC /*서식기록내역 Table*/ INNER JOIN PTINF USING(PTSBSTNO) WHERE FORMCD = '1200005398' ) </pre>
---	---

## B. '#6 항암요법명' 데이터 추출.

- 관련 키워드: 약제 조합으로서의 항암요법
  - 약제 정보 관련 디자인코드 : 0002965.0003124.0000627.0012276 (CTX1)
  - 0000627.0012276 (CTX2)
- 항암치료에 사용한 약제 정보를 '+'문자로 결합해서 하나의 Row로 추출한다.
- CTX1, CTX2 Table에서 모두 추출할 수 있으므로 UNION ALL로 결합한다.

<pre>SELECT PTSBSTNO, ORDDD, /*약제 정보에서 공백문자 제거 과정을 거쳐야 하므로 아래의 SQL 코드 정렬 금지*/ LISTAGG(RTRIM(TRIM(CAST(RECCNTS AS VARCHAR))), '), '+' ) AS DRUG FROM CTX1 WHERE DEGNITEMLEVLCD = '0002965.0003124.0000627.0012276' GROUP BY PTSBSTNO, ORDDD</pre>	<pre>SELECT PTSBSTNO, ORDDD, LISTAGG(LTRIM(TRIM(RTRIM(REPLACE(CAST(RECCNTS AS VARCHAR), '약명', ''))), '), '+' ) AS DRUG FROM CTX2 WHERE DEGNITEMLEVLCD = '0000627.0012276' GROUP BY PTSBSTNO, ORDDD</pre>								
<table><tr><th>ABC ANTP_NM</th></tr><tr><td>5-FU+A.D. mycin+Endoxan</td></tr><tr><td>5-FU+A.D. mycin+Endoxan</td></tr><tr><td>5-FU+A.D.MYCIN+ENDOXAN</td></tr><tr><td>5-FU+A.D.MYCIN+ENDOXAN</td></tr><tr><td>5-FU+Doxorubin+Endoxan</td></tr><tr><td>5-FU+Doxorubin+Endoxan</td></tr><tr><td>5-FU+EPIZIN+ENDOXAN</td></tr></table>		ABC ANTP_NM	5-FU+A.D. mycin+Endoxan	5-FU+A.D. mycin+Endoxan	5-FU+A.D.MYCIN+ENDOXAN	5-FU+A.D.MYCIN+ENDOXAN	5-FU+Doxorubin+Endoxan	5-FU+Doxorubin+Endoxan	5-FU+EPIZIN+ENDOXAN
ABC ANTP_NM									
5-FU+A.D. mycin+Endoxan									
5-FU+A.D. mycin+Endoxan									
5-FU+A.D.MYCIN+ENDOXAN									
5-FU+A.D.MYCIN+ENDOXAN									
5-FU+Doxorubin+Endoxan									
5-FU+Doxorubin+Endoxan									
5-FU+EPIZIN+ENDOXAN									

Figure 58. 항암요법명 데이터 추출 (약제 조합)

## C. '#7 항암요법주기수' 데이터 추출.

- 관련 키워드: 항암요법 치료 주기 수 (Cycle)
  - CTX1 디자인코드 : 0002965.0002965 (#N Cycle 값으로 기록되어 있다.)
  - CTX2 디자인코드 : 0002965.0012039 (Cycle수가 그대로 기록되어 있다.)

<pre>SELECT PTSBSTNO, ORDDD, REGEXP_SUBSTR(LISTAGG(CAST(RECCNTS AS VARCHAR)), '\d{1,2}(?=\s?cycle)', 1,1, 'i') AS CYC FROM CTX1 WHERE DEGNITEMLEVLCD = '0002965.0002965' GROUP BY PTSBSTNO, ORDDD</pre>	<pre>SELECT PTSBSTNO, ORDDD, LISTAGG(CAST(RECCNTS AS VARCHAR), ' ') AS CYC FROM CTX2 WHERE DEGNITEMLEVLCD = '0002965.0012039' GROUP BY PTSBSTNO, ORDDD</pre>
---	--

Figure 59. 항암요법주기수 데이터 추출 Query

## D. '#16 #17 #18 #19 항암요법부작용' 데이터 추출.

- 부작용 도메인 : Alopecia | Anorexia | Constipation | Diarrhea | Dry skin | Fatigue | Hand-foot syndrome | Leukopenia | Mucositis | Nausea/vomiting | Neuropathy | Skin rash | 기타
- 부작용등급 도메인 : (Grade) I | II | III | IV | V

- CTX2에는 부작용과 관련된 서식 항목이 존재하지만, CTX1에는 존재하지 않는다.
- 하지만, FORMCD = '1600011944' 서식에서 부작용에 대한 기록을 확인할 수 있다.
- 1600011944 FORMCD = 동일 항암화학요법 부작용평가 서식.

ABC DEGNITEMLEVLCD	ABC RECTERMNM	ABC DEGNITEMNM
0015952	[NULL]	General disorders and administration site conditions
0015952.0017680	Fatigue	[NULL]
0015952.0017680.0019385	1	Grade
0015941	[NULL]	Blood and lymphatic system disorders
0015941.0017545	Anemia	[NULL]
0000484	[NULL]	Gastrointestinal disorders
0000484.0017679	Diarrhea	[NULL]
0000484.0017679.0003453	1	Grade
0015952	[NULL]	General disorders and administration site conditions
0015952.0017680	Fatigue	[NULL]
0015952.0017680.0019385	1	Grade

Figure 60. FORMCD 1600011944 서식 항목의 내용

- 서식내용 중 'General disorders and administration site conditions' 항목의 부작용을 추출해서 활용한다.
- 디자인코드 : 0015952.0017680 (부작용 내용)
- 디자인코드 : 0015952.0017680.0019385 (부작용 등급)

<pre> SELECT PTSBSTNO, ORDDD , LISTAGG(RECTERMNM, ':') AS SEFF FROM CDW.MR_FORM_REC WHERE FORMCD = '1600011944' AND DEGNITEMLEVLCD LIKE '%0015952%' /*General disorders and administration site conditions 주제*/ GROUP BY PTSBSTNO, ORDDD </pre>	<pre> SELECT PTSBSTNO, ORDDD, LISTAGG(RECTERMNM, ' ') AS SEFF FROM CTX2 WHERE DEGNITEMLEVLCD = '0013574' GROUP BY PTSBSTNO, ORDDD </pre>
---	--

Figure 61. 부작용 내용과 등급 데이터 수집 Query

#### E. '#20 #21 항암치료목적' 데이터 추출.

- 도메인 : Neoadjuvant | Adjuvant | Palliative | Concurrent | Induction | Maintenance | Salvage | Consolidation
- CTX1 : 0002076.
- CTX2 : 치료목적에 대한 항목 없음.

```

SELECT PTSBSTNO, ORDDD, LISTAGG(CAST(RECTERMNM AS VARCHAR), '|') AS PRPS
FROM CTX1
WHERE DEGNITEMLEVLCD = '0002076'
GROUP BY PTSBSTNO, ORDDD

```

Figure 62. 항암치료목적 데이터 수집 Query