

## III-1. 'HIRA 빅데이터 분석 가이드' 란?

- 본 분석가이드는 건강보험심사평가원의 데이터를 활용을 원하는 연구자들에게 도움이 되고자 작성되었습니다. 심평원 데이터를 처음 접하거나, 익숙치 않은 분들도 연구에 자유롭게 활용할 수 있도록 자세하고 쉽게 설명하고자 합니다.
- 이번 호에서는 동반질환 분석에 대한 설명을 담고 있으며, 다음 호에서는 대조군 설정에 대해 소개할 예정입니다.

**[표 1]** HIRA 빅데이터 분석 가이드 계획(2018년)

호	주 제
2018년 1분기 (제2권 1호)	HIRA 보건 의료 빅데이터 소개
2018년 2분기 (제2권 2호)	입원 에피소드 구축
2018년 3분기 (제2권 3호)	동반질환 분석
2018년 4분기 (제2권 4호)	대조군 설정





## III-2. HIRA 빅데이터 분석 - 동반질환 분석

※ 출처: 건강보험 청구자료 분석 매뉴얼(2017, 건강보험심사평가원)

### 1. 동반질환의 정의

- 동반질환(Comorbidity)은 환자가 가지고 있는 질환 중 주진단 질환이 아닌 나머지 질환을 뜻함
  - ▶ 합병증의 개념과 비슷하지만 주진단 질환과 관련이 없다는 점에서 차이가 있음
  - ▶ 환자에 따라 여러 가지 동반질환을 가지고 있을 수 있음
- 건강보험 청구자료를 활용한 연구에서는 객관적인 인과성 추론을 위하여 중증도 보정 과정을 거치는데 동반질환은 중증도 보정 시 중요한 변수임
  - ▶ 중증도 보정은 환자 결과에 영향을 줄 수 있는 교란변수들을 통제하는 과정으로 연령, 성별, 의료보장형태, 응급실 방문 등의 변수를 활용함
  - ▶ 동반질환은 주진단 질환과 관련은 없으나 합병증, 사망, 재원일수, 진료비 등에 영향을 준다는 점에서 중요한 중증도 변수임
  - ▶ 예를 들어 동일한 질환, 동일한 치료를 받은 환자임에도 연령, 성별, 동반질환 유무 등에 따라 환자의 결과는 차이가 있을 수 있으므로, 중증도 보정을 통해 동일한 조건으로 보정이 필요함

## 2. 동반질환 구축 과정

### 2.1 동반질환 측정 도구

- 동반질환 보정 방법에는 기존 발표된 연구 결과를 근거로 임상적 견해 혹은 통계적 검정 결과에 따라 유의한 질환만을 선택적으로 보정하는 방법(ad hoc selection)과 동반질환 측정도구를 사용하는 방법이 있음
  - ▶ 선택적 보정 방법은 동반질환에 대한 서로 다른 견해 또는 통계적 접근 방법으로 연구들 간의 비교가능성이 낮다는 제한점이 있음
  - ▶ 동반질환을 종합적으로 평가하고, 다양한 질환에 적용할 수 있는 Charlson 동반질환 지수(Charlson comorbidity index, CCI)<sup>1)</sup>와 Elixhauser의 동반질환(Elixhauser's comorbidity measure, ECM)<sup>2)</sup> 등이 주로 활용됨

#### 1) Charlson 동반질환 지수(CCI)

- CCI는 동반질환의 중증도가 반영되어 있다는 점에서 행정자료를 자료원으로 하는 연구에서 널리 사용되는 측정도구임
  - ▶ Charlson 등은 1984년 1개월 동안 뉴욕병원에 입원한 604명 환자의 의무기록을 토대로 1년 사망을 잘 예측하는 19개 질환을 선정함. 그리고 각 질환별 상대위험도를 근거로 1, 2, 3, 6점의 가중치를 부여하였고, 이 가중치의 합이 CCI에 해당됨

ID	MI (1)	CHF (1)	...	Renal disease(2)	...	Tumor (6)	CCI
1	1 +	0 +		1 +		0 =	3
2	0	1		1		0	3
3	0	1		0		0	1
4	0	0		0		1	6
...							

[그림 1] Charlson 동반질환 지수 산출 방법

1) Charlson ME, et al. A new method of classifying prognostic comorbidity in longitudinal studies development and validation. J Chron Dis 1987;40(5):373-383.  
 2) Elixhauser A, et al. Comorbidity measures for use with administrative data. Med Care 1998;36(1):8-27.



- 의무기록자료를 토대로 개발된 CCI를 행정자료에 적용하기 위하여 ICD-9 혹은 ICD-10 코드로 전환함
  - ▶ CCI를 개발할 당시에는 19개의 동반질환이 선정되었으나, Deyo 등<sup>3)</sup>이 ICD-9 코드로 전환하면서 백혈병과 림프종을 비전이형 고형암에 포함시켜 17개 질환으로 축소됨. 그 뒤 Quan 등(2005)<sup>4)</sup>은 Deyo의 알고리즘을 토대로 ICD-10 코드로 전환하여 최종적으로 17개 질환으로 구성됨

Comorbidities	International Classification of Disease, 10th revision code	Original weight	Updated weight
Myocardial infarction	I21.x, I22.x, I25.2	1	0
Congestive heart failure	I08.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0	1	-2
Peripheral vascular disease	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9	1	0
Cerebrovascular disease	G45.x, G46.x, H34.0, I60.x-I69.x	1	0
Dementia	F00.x-F03.x, F05.1, G30.x, G31.1	1	2
Chronic pulmonary disease	I27.8, I27.9, J40.x-J47.x, J60.x-J67.x, J68.4, J70.1, J70.3	1	1
Rheumatologic disease	M05.x, M06.x, M31.5, M32.x-M34.x, M35.1, M35.3, M36.0	1	1
Peptic ulcer disease	K25.x-K28.x	1	0
Mild liver disease	B18.x, K70.0-K70.3, K70.9, K71.3-K71.5, K71.7, K73.x, K74.x, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4	1	2
Diabetes without chronic complication	E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9	1	0
Diabetes with chronic complication	E10.2-E10.5, E10.7, E11.2-E11.5, E11.7, E12.2-E12.5, E12.7, E13.2-E13.5, E13.7, E14.2-E14.5, E14.7	2	1
Hemiplegia or paraplegia	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9	2	2
Renal disease	I12.0, I13.1, N03.2-N03.7, N05.2-N05.7, N18.x, N19.x, N25.0, Z49.0-Z49.2, Z94.0, Z99.2	2	1
Any malignancy including leukemia and lymphoma	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C81.x-C85.x, C88.x, C90.x-C97.x	2	2
Moderate or severe liver disease	I85.0, I85.9, I86.4, I88.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7	3	4
Metastatic solid tumor	C77.x-C80.x	6	6
Acquired immune deficiency syndrome/ human immunodeficiency virus	B20.x-B22.x, B24.x	6	4

[그림 2] Charlson 동반질환과 가중치(Quan 등, 2005)

- CCI는 일반적으로 0점, 1점, 2점, 3점 이상으로 구분하여 범주형 변수로 사용되나, CCI 분포에 따라 범주를 구분하여 사용할 수 있음(예: 0점, 1점, 2점 이상)

- 3) Deyo RA, et al. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45(6):613-619.
- 4) Quan H, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005;43(11):1130-1139.

## 2) Elixhauser 동반질환(ECM)

- Elixhauser 등은 행정자료를 사용하여 재원기간, 진료비, 병원 내 사망 등을 잘 예측하는 30개 동반질환을 선정하고, 각 질환별로 ICD-9 코드를 제시함
  - ▶ 30개 동반질환은 합병증을 동반한 고혈압과 동반하지 않는 고혈압을 통합한 개수로, 이를 분리하면 31개 질환임
- ECM은 CCI처럼 하나의 지수로 환산하지 않고, 각각의 동반질환을 이분형(binary) 형태로 사용되나, CCI와 같이 하나의 점수로 환산하는 연구들이 보고됨
  - ▶ Dominick 등<sup>5)</sup>은 30개 질환에 각각 1점을 부여한 후 합산하는 방법을 제시하였고, van Walraven 등<sup>6)</sup>은 병원 내 사망을 예측하는 다중 로지스틱 회귀모형에서 얻은 회귀계수를 사용하여 점수 시스템을 개발함
- ECM에는 CCI와 비교하여 다양한 정신질환(알코올, 약물 남용 등)이 포함되어 있으며, 비만과 체중 감소와 같이 건강보험 청구자료에는 과소입력될 가능성이 있는 질환이 포함되어 있음

5) Dominick KL, et al. Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis. *Arthritis Rheum* 2005;53(5):666-672.

6) van Walraven, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care* 2009;47(6):626-633.



Comorbidities	International Classification of Disease, 10th revision code	Weight
Congestive heart failure	I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5-I42.9, I43.x, I50.x, P29.0	7
Cardiac arrhythmias	I44.1-I44.3, I45.6, I45.9, I47.x, I49.x, R00.0, R00.1, R00.8, T82.1, Z45.0, Z95.0	5
Valvular disease	A52.0, I05.x-I08.x, I09.1, I09.0, I34.x-I39.x, I73.0-I73.3, Z95.2-Z95.4	-1
Pulmonary circulation disorders	I26.x, I27.x, I28.0, I28.8, I28.9	4
Peripheral vascular disorders	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9	2
Hypertension, uncomplicated	I10.x	0
Hypertension, complicated	I11.x, I13.x, I15.x	0
Paralysis	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9	7
Neurodegenerative disorders	G10.x-G13.x, G20.x-G22.x, G25.4, G25.5, G31.2, G31.8, G31.9, G32.x, G35.x-G37.x, G40.x, G41.x, G93.1, G93.4, I47.0, I456.x	6
Chronic pulmonary disease	I27.8, I27.9, J40.x-J47.x, J60.x-J67.x, J68.4, J70.1, J70.3	3
Diabetes, uncomplicated	E10.0, E10.1, E10.9, E11.0, E11.1, E11.9, E12.0, E12.1, E12.9, E13.0, E13.1, E13.9, E14.0, E14.1, E14.9	0
Diabetes, complicated	E10.2-E10.8, E11.2-E11.8, E12.2-E12.8, E13.2-E13.8, E14.2-E14.8	0
Hypothyroidism	E00.x-E03.x, E89.0	0
Renal failure	I12.0, I13.1, N18.x, N19.x, N25.0, Z49.0, Z49.2, Z94.0, Z99.2	5
Liver disease	B18.x, I85.x, I86.4, I98.2, K70.x, K71.1, K71.3, K71.5, K71.7, K72.x, K74.x, K76.0, K76.2, K76.9, Z94.4	11
Peptic ulcer disease, no bleeding	K25.7, K25.9, K26.7, K26.9, K27.7, K27.9, K28.7, K28.9	0
Acquired immune deficiency syndrome/ human immunodeficiency virus	B20.x-B22.x, B24.x	0
Lymphoma	C81.x-C85.x, C88.x, C96.x, C90.0, C90.2	9
Metastatic cancer	C77.x-C80.x	12
Solid tumour without metastasis	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C97.x	4
Rheumatoid arthritis/ collagen vascular diseases	L94.0, L94.1, L94.3, M05.x, M06.x, M08.x, M12.0, M12.3, M30.x, M31.0-M31.3, M32.x-M35.x, M45.x, M46.1, M46.8, M46.9	0
Coagulopathy	D65-D68.x, D69.1, D69.3-D69.6	3
Obesity	E66.x	-4
Weight loss	E40.x, E46.x, R63.4, R64	6
Fluid and electrolyte disorders	E22.2, E86.x, E87.x	5
Blood loss anaemia	D50.0	-2
Deficiency anaemia	D50.0, D50.9, D51.x-D53.x	-2
Alcohol abuse	F10, E52, G62.1, I42.6, K29.2, K70.0, K70.3, K70.9, I51.x, Z50.2, Z71.4, Z72.1	0
Drug abuse	F11.x, F16.x, F18.x, F19.x, Z71.5, Z72.2	-7
Psychosis	F20.x, F22.x-F25.x, F28.x, F29.x, F30.2, F31.2, F31.5	0
Depression	F20.4, F31.3-F31.5, F32.x, F33.x, F34.1, F41.2, F43.2	-3

[그림 3] Elixhauser의 동반질환과 가중치(Quan 등, 2005)

## 2.2 동반질환 관찰기간(look back period) 결정

- 동반질환 관찰기간은 환자의 이전 의료이용을 추적하는 기간을 의미하며, 이는 동반질환 유병률과 모형 예측력에 영향을 줌
  - ▶ 관찰기간이 길어질수록 동반질환을 관찰할 기회는 증가하지만, 동반질환 보정 효과가 감소하여 비용-효과적이지 않다고 보고됨
- 건강보험 청구자료에는 환자 고유 식별자가 포함되어 있어 자료 접근 범위 내에서 장기간 추적이 가능하지만, 대부분의 연구에서는 자료 접근성과 임상적 판단을 고려하여 기준 일자(index date)를 기준으로 이전 1년 정도를 동반질환 관찰기간으로 설정함

- 건강보험 청구자료만으로는 입원 중에 발생한 질환인지의 여부를 구분할 수 없기 때문에 기준 시점에 입력된 상병을 제외한 이전에 기록된 상병만을 관찰할 것을 제안함

## 2.3 자료 범위 결정

- 동반질환 분석 시 결정해야 하는 자료의 범위는 크게 의료 이용과 상병임
  - ▶ 의료 이용은 입원과 외래 진료를 뜻하며, 분석 질환 등에 따라 모두 사용하거나 한쪽을 선택하여 사용함
  - ▶ 상병의 범위는 i) 주상병, ii) 주상병과 제1부상병, iii) 모든 상병으로 구분할 수 있으며, 연구자의 판단에 따름
- 건강보험 청구자료를 사용한 분석에서는 입원과 외래 자료를 모두 사용하여 동반질환을 관찰함
  - ▶ 비교적 진단의 정확성이 높은 입원자료만을 사용하여 동반질환을 관찰할 수 있으나, 당뇨와 고혈압과 같은 외래 중심 질환을 관찰하지 못한다는 제한점이 있음
- 주상병은 진료기간 중 최종적으로 환자가 진단받은 병태로서 치료나 검사에 대한 요구가 가장 큰 병태이며, 부상병은 진료기간 중 주상병과 함께 있었거나, 진료기간 중 발생되어 환자 진료에 영향을 준 병태를 의미함
  - ▶ 배제 상병은 병태가 의심되어 검사 등을 수행하였으나 최종 진단을 받지 않은 경우에 해당되므로, 제외하고 동반질환을 구축함



### 3. 동반질환 분석(CCI) 예시

- 환자 기준 시점(index date)을 기준으로 과거 1년 동안 동반질환을 관찰하여 CCI를 산출하는 프로그램을 제시
  - ▶ 동반질환을 합병증과 구분하기 위하여 기준 시점에 청구된 상병명을 CCI에 포함하지 않고, Quan 등(2005)이 제시한 ICD-10 알고리즘과 Original 가중치를 적용함
  - ▶ 주상병과 제1부상병만(200테이블)을 관찰하는 경우, 모든 상병(400테이블)을 관찰하는 경우로 구분하여 프로그램을 제시함

/\*[1단계] 환자의 기준 시점(index date) 설정: 가장 처음 진료받은 건 기준\*/

환자 고유 식별자(pat\_hpin), 요양개시일자(recu\_fr\_dd), 요양종료일자(recu\_to\_dd)를 기준으로 정렬하고, 첫 번째 자료를 기준으로 설정\*/

```
proc sort data=temp; by pat_hpin recu_fr_dd recu_to_dd; run;
data index(keep=pat_hpin index_start_date index_end_date); set temp; by pat_hpin;
format index_start_date index_end_date yymmdd10.;
index_start_date = input(recu_fr_dd, yymmdd8.);
index_end_date = input(recu_to_dd, yymmdd8.);
if first.pat_hpin=1 then output; run;
```

/\*[2단계] Charlson 동반질환 해당 여부 코딩을 위한 DATA 변환\*/

/\* 200테이블의 주진단과 제1부진단만 사용하는 경우(st200) → DATA를 400테이블 형태로 변경\*/

```
data temp1; set st200;
sick_cd=msick_cd; output;
sick_cd=ssick_cd; output; run;
data temp2; set temp1; if sick_cd=' ' or sick_cd='$' then delete;
keep pat_hpin recu_fr_dd recu_to_dd sick_cd spjkey; run;
```

※ DATA형태 전환

[전] wide form

[후] long form

pat_hpin	msick_cd	ssick_cd	recu_fr_dd	recu_to_dd	→	pat_hpin	sick_cd	recu_fr_dd	recu_to_dd
1	I20	I21	20160101	20160112		1	I20	20160101	20160112
2	I22		20160108	20160201		1	I21	20160101	20160112
						2	I22	20160108	20160201

/\* 400테이블의 상병명을 사용하는 경우(st400) → 배제진단만 제외\*/

```
data temp2; set st400; if sick_div_ty_cd='3' then delete;
keep pat_hpin recu_fr_dd recu_to_dd sick_cd spjkey2; run;
```



```

/*[3단계] 청구명세서 단위로 Charlson 동반질환 코딩*/
proc sql;
create table comorbidity as select SPJKEY,
/*01. MI*/
max(case when substr(SICK_CD,2,3) in ('I21', 'I22') or substr(SICK_CD,2,4) in ('I252') then 1 end)
as count_cci01,
/*02. CHF 심부전*/
max(case when substr(SICK_CD,2,3) in ('I43', 'I50') or substr(SICK_CD,2,4) in ('I099', 'I110', 'I130',
'I132', 'I255', 'I420', 'I425', 'I426', 'I427', 'I428', 'I429', 'P290') then 1 end) as count_cci02,
/*03. peri, 말초혈관질환*/
max(case when substr(SICK_CD,2,3) in ('I70', 'I71') or substr(SICK_CD,2,4) in ('I731', 'I738', 'I739',
'I771', 'I790', 'I792', 'K551', 'K558', 'K559', 'Z958', 'Z959') then 1 end) as count_cci03,
/*04. CERE, 뇌혈관질환*/
max(case when substr(SICK_CD,2,3) in ('G45', 'G46', 'I60', 'I61', 'I62', 'I63', 'I64', 'I65', 'I66', 'I67', 'I68',
'I69') or substr(SICK_CD,2,4) in ('H340') then 1 end) as count_cci04,
/*05. Deme, 치매*/
max(case when substr(SICK_CD,2,3) in ('F00', 'F01', 'F02', 'F03', 'G30') or substr(SICK_CD,2,4) in
('G311', 'F051') then 1 end) as count_cci05,
/*06. COPD 만성폐질환*/
max(case when substr(SICK_CD,2,3) in ('J40', 'J41', 'J42', 'J43', 'J44', 'J45', 'J46', 'J47', 'J60', 'J61',
'J62', 'J63', 'J64', 'J65', 'J66', 'J67') or substr(SICK_CD,2,4) in ('I278', 'I279', 'J684', 'J701', 'J703')
then 1 end) as count_cci06,
/*07. tiss, Rheumatologic disease 결합조직질환*/
max(case when substr(SICK_CD,2,3) in ('M05', 'M06', 'M32', 'M33', 'M34') or substr(SICK_CD,2,4) in
('M315', 'M351', 'M353', 'M360') then 1 end) as count_cci07,
/*08. ulcer, Peptic ulcer disease 소화궤양*/
max(case when substr(SICK_CD,2,3) in ('K25', 'K26', 'K27', 'K28') then 1 end) as count_cci08,
/*09. liver, Mild Liver Disease 경도의 간질환*/
max(case when substr(SICK_CD,2,3) in ('B18', 'K73', 'K74') or substr(SICK_CD,2,4) in ('K700', 'K701',
'K702', 'K703', 'K709', 'K713', 'K714', 'K715', 'K717', 'K760', 'K762', 'K763', 'K764', 'K768', 'K769', 'Z944') then 1
end) as count_cci09,
/*10. dm, 당뇨*/
max(case when substr(SICK_CD,2,4) in ('E100', 'E101', 'E106', 'E108', 'E109', 'E110', 'E111',
'E116', 'E118', 'E119', 'E120', 'E121', 'E126', 'E128', 'E129', 'E130', 'E131', 'E136', 'E138', 'E139', 'E140',
'E141', 'E146', 'E148', 'E149') then 1 end) as count_cci10,

```



```

/*12. hemi, Hemiplegia or paraplegia 반신마비*/
  max(case when substr(SICK_CD,2,3) in ('G81','G82') or substr(SICK_CD,2,4) in ('G041','G114','G80
1','G802','G830','G831','G832','G833','G834','G839') then 1 end) as count_cci12,
/*13. Renal, Renal diseiase 중등도이상 신장질환*/
  max(case when substr(SICK_CD,2,3) in ('N18','N19') or substr(SICK_CD,2,4) in ('I120','I131','N032','
N033','N034','N035','N036','N037','N052','N053','N054','N055','N056',
'N057','N250','Z490','Z491','Z492','Z940','Z992') then 1 end) as count_cci13,
/*14. lymf, 비전이성고형암, 백혈병, 림프종, 다발성 골수종 */
  max(case when substr(SICK_CD,2,3) in ('C00','C01','C02','C03','C04','C05','C06','C07','C08',
'C09','C10','C11','C12','C13','C14','C15','C16','C17','C18','C19','C20','C21','C22','C23','C24','C25','C26','C
30','C31','C32','C33','C34','C37','C38','C39','C40','C41','C43','C45','C46','C47','C48','C49','C50','C51','C5
2','C53','C54','C55','C56','C57','C58','C60','C61','C62','C63','C64','C65','C66','C67','C68','C69','C70','C71
','C72','C73','C74','C75','C76','C81','C82','C83','C84','C85','C88','C90','C91','C92','C93','C94','C95','C96',
'C97') then 1 end) as count_cci14,
/*15. liver_s, Moerate or severe liver disease 중등도 이상 간질환*/
  max(case when substr(SICK_CD,2,4) in ('I850','I859','I864','I982','K704','K711',
'K721','K729','K765','K766','K767') then 1 end) as count_cci15,
/*16. tumor, Metastatic solid tumor 전이성 고형암*/
  max(case when substr(SICK_CD,2,3) in ('C77','C78','C79','C80') then 1 end) as count_cci16,
/*17. AIDS*/
  max(case when substr(SICK_CD,2,3) in ('B20','B21','B22','B24') then 1 end) as count_cci17

from temp2
group by SPJKEY; quit;

data comorbidity2; set comorbidity;
total=sum(count_cci01, count_cci02, count_cci03, count_cci04, count_cci05, count_cci06,
count_cci07, count_cci08, count_cci09, count_cci10, count_cci11, count_cci12, count_cci13,
count_cci14, count_cci15, count_cci16, count_cci17);
if total=0 then delete; run; /*Charlson 동반질환이 없는 청구건은 삭제*/

```

```

/*[4단계] index date를 기준으로 1년 전 동반질환 내역만 가져오기*/
proc sort data=index; by pat_hpin; run;
proc sort data=comorbidity2; by pat_hpin; run;
data comorbidity3; merge index comorbidity2; by pat_hpin;
format fol_start_date fol_end_date yymmdd10.;
fol_start_date = input(recu_fr_dd, yymmdd8.);
fol_end_date = input(recu_to_dd, yymmdd8.);
if -365<intck('day',index_start_date,fol_end_date)<0; run;
/*기준 시점의 요양개시일자와 과거 진료내역의 요양종료일자의 차이가 1년 내 청구자료만 가져오기*/

```

```

/*[5단계] 청구명세서 단위의 동반질환 내역을 환자 단위로 축소*/
proc sql;
create table comorbidity4 as select *,
max(count_cci1) as ncount_cci1, max(count_cci2) as ncount_cci2,
max(count_cci3) as ncount_cci3, max(count_cci4) as ncount_cci4,
max(count_cci5) as ncount_cci5, max(count_cci6) as ncount_cci6,
max(count_cci7) as ncount_cci7, max(count_cci8) as ncount_cci8,
max(count_cci9) as ncount_cci9, max(count_cci10) as ncount_cci10,
max(count_cci11) as ncount_cci11, max(count_cci12) as ncount_cci12,
max(count_cci13) as ncount_cci13, max(count_cci14) as ncount_cci14,
max(count_cci15) as ncount_cci15, max(count_cci16) as ncount_cci16,
max(count_cci17) as ncount_cci17
from comorbidity3
group by pat_hpin
order by pat_hpin; quit;

proc sort data=comorbidity4 nodupkey out=comorbidity5; by pat_hpin; run;

```



```

/*[6단계] Index Data에 동반질환점수 붙이기*/
proc sort data=comorbidity5; by pat_hpin; run;
proc sort data=index; by pat_hpin; run;
data Final; merge index comorbidity5; by pat_hpin;
if ncount_cci1=. then ncount_cci1=0; if ncount_cci2=. then ncount_cci2=0;
if ncount_cci3=. then ncount_cci3=0; if ncount_cci4=. then ncount_cci4=0;
if ncount_cci5=. then ncount_cci5=0; if ncount_cci6=. then ncount_cci6=0;
if ncount_cci7=. then ncount_cci7=0; if ncount_cci8=. then ncount_cci8=0;
if ncount_cci9=. then ncount_cci9=0; if ncount_cci10=. then ncount_cci10=0;
if ncount_cci11=. then ncount_cci11=0; if ncount_cci12=. then ncount_cci12=0;
if ncount_cci13=. then ncount_cci13=0; if ncount_cci14=. then ncount_cci14=0;
if ncount_cci15=. then ncount_cci15=0; if ncount_cci16=. then ncount_cci16=0;
if ncount_cci17=. then ncount_cci17=0; /*동반질환이 없는 환자는 0으로 코딩*/
CCI=(ncount_cci1+ncount_cci2+ncount_cci3+ncount_cci4+ncount_cci5+ncount_cci6+ncount_cci7+
ncount_cci8+ncount_cci9+ncount_cci10+2*ncount_cci11+2*ncount_cci12+2*ncount_cci13+2*ncou
nt_cci14+3*ncount_cci15+6*ncount_cci16+6*ncount_cci17);
CCI_group=CCI;
if CCI>=3 then CCI_group=3; run;

```

## 4. 맺음말

- 건강보험 청구자료를 분석할 때는 환자의 결과에 영향을 줄 수 있는 변수들을 보정하여야 보다 객관적인 인과 추론이 가능함
  - ▶ 환자의 결과에 영향을 줄 수 있는 교란변수는 주진단 질환의 중증도, 임상적 상태 등 다양하나, 건강보험 청구자료에서 활용할 수 있는 변수는 연령, 성별, 동반질환 등 임
- 본 가이드에서는 일반적으로 많이 활용하고 있는 Charlson 동반질환 지수에 대해 설명하고 있지만 연구자 판단에 따라 다양한 방법들이 활용될 수 있음
  - ▶ 연구 성격에 따라 동반질환 측정 도구뿐만 아니라 적합한 관찰 기간, 자료 범위를 설정하여야 함

## 👉 HIRA 보건 의료 빅데이터 이용 안내

- ‘보건 의료 빅데이터 개방 시스템(opendata.hira.or.kr)’에서는 여러 가지 형태의 빅데이터 서비스가 제공되고 있으며, 용도에 따라 적합한 서비스 선택 · 이용이 가능함

구분	내용
원시자료(raw data) 분석 지원 서비스	연구자 및 산업체가 원하는 맞춤형 원시데이터 분석이 가능하도록 방문 또는 원격 접속으로 빅데이터 분석 시스템 활용
Open API 실시간 제공 서비스	민간기업, 프로그램 개발자, 인터넷 이용자를 대상으로 손쉽게 직접 응용 프로그램과 서비스를 개발할 수 있도록 병원 · 약국 정보 서비스 등 보건 의료 데이터 지원
환자표본 DB 제공 서비스	연구자의 필요에 따라 환자표본 데이터 등 총 4개의 DB 데이터셋을 생성 · 제공하여 연구 및 R&D 사업에 활용하도록 지원

- HIRA 보건 의료 빅데이터 이용 절차(원격 접속)



- ※ 자세한 내용은 ‘보건 의료 빅데이터 개방 시스템(opendata.hira.or.kr)’에 접속 후 ‘의료 빅데이터’ - ‘빅데이터 분석 이용 안내’ 참고

