

'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 입원 에피소드 구축에 대한 설명을 담고 있으며, 다음 호에서는 동반질환 분석에 대해 소개할 예정입니다.

[표 1] HIRA 빅데이터 분석 가이드 계획(2018년)

호	주제
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





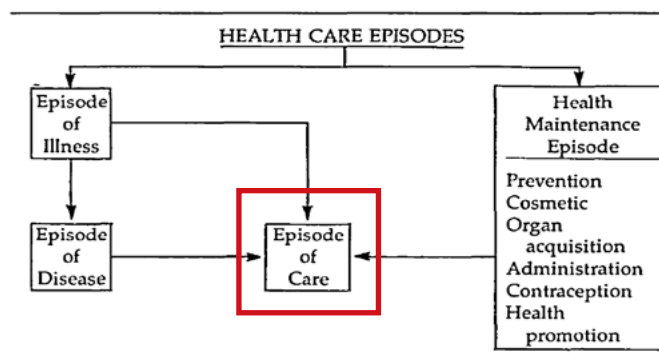
HIRA 빅데이터 분석 - 입원 에피소드 구축

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

1. 에피소드의 정의

- 보건의료 에피소드(Health care episode)란 ‘건강 문제의 시작 시점부터 완료 시점까지’를 하나의 사건으로 묶은 것으로, Hornbrook 등(1985)¹⁾에 의해 소개됨
 - ▶ 건강 문제란 특정 질병·질환의 진행, 건강관리 문제, 치료 과정 등을 의미하며, 각각 하나의 에피소드로 구분할 수 있음
 - ▶ 실제 건강 문제를 인과관계로 규명하기에는 어려움이 있어 시작과 끝을 정의하기 불분명함
 - ▶ 따라서 대안으로 요양기관에서 의료 서비스를 받기 시작한 시점부터 종료된 시점까지를 진료 에피소드(Episode of Care)로 정의하였음

TABLE 1 Types of Health Care Episodes



[그림 1] 보건의료 에피소드의 종류(Hornbrook 등, 1985)

- 환자가 요양기관에서 진료를 받는 경우 진료내역이 명세서 상에 기록되므로 진료를 개시한 날짜와 종료한 날짜를 확인할 수 있음
 - ▶ 우리나라는 건강보험 제도 기반하에 전 국민의 진료내역이 건강보험심사평가원에 청구되므로, 이를 활용하여 진료 에피소드를 정의할 수 있음

1) Hornbrook MC, Hurtado AV, Johnson RE. Health Care Episodes: Definition, Measurement and Use. Med Care Rev 1985; 42(2):163-218.

2. 청구자료 분석 시 에피소드 구축의 필요성

- 건강보험 청구자료는「요양급여비용 청구방법, 심사청구서·명세서서식 및 작성요령」(보건복지부 고시)에 따라 요양기관이 심사평가원으로 청구한 명세서를 청구건 단위로 누적한 자료임
- 따라서 환자 단위 분석을 시도할 시, 청구건 단위 명세서를 동일 수진자 기준으로 정렬·통합할 필요가 있음
- 입원의 경우 동일 수진자에 대한 입원진료기간의 요양급여내역을 동일한 명세서에 통합하여 작성하는 것이 원칙이나, 아래와 같은 '분리 청구'의 경우가 발생할 수 있음
 - ▶ 입원 기간 중 월이 달라질 경우 월별로 구분하여 청구
 - ▶ 30일을 초과하여 입원진료를 하는 경우 월단위로 분리하여 청구
- 입원 '분리 청구'의 경우, 동일 수진자 한 건의 입원사례에서 여러 건의 명세서가 발생할 수 있게 되므로, 이를 통합하기 위해 입원 에피소드 구축이 필요함

1. 일반 청구	2. 분리 청구
----------	----------

환자	요양개시일자	요양종료일자
A	20180101	20180115

환자	요양개시일자	요양종료일자
A	20180101	20180131
A	20180201	20180228
A	20180301	20180315

※ ○ 는 실제 입원 날짜, □ 는 실제 퇴원 날짜 임

[그림 2] 일반 청구와 분리 청구 명세서 비교

※ 위의 분리 청구 사례 외에도, 다양한 경우로 인해 에피소드 구축이 필요한 경우가 있을 수 있으므로 청구형태를 반드시 확인하여야 함²⁾

2) 52p 'Tip 입원 에피소드 구축 시 주의할 점' 참고



3. 입원 에피소드 구축 과정

- 입원 에피소드는 명세서 테이블에 있는 수진자개인식별대체키, 요양기관식별대체키, 요양개시일자, 요양종료일자를 기준으로 정렬한 뒤, '기준 입원 명세서'와 '다음 입원 명세서'를 비교하는 과정을 통해 구축함

※ 용어 정의

1. 기준 입원 명세서: 에피소드 구축 시 기준이 되는 명세서로 동일 환자, 동일 요양기관에서 발생한 명세서 중 요양개시일자가 가장 빠른 명세서
2. 다음 입원 명세서: '기준 입원 명세서'와 환자, 요양기관이 동일하며, 요양개시일자가 '기준 입원 명세서'의 요양종료일자와 연속(1일 이내)되는 명세서

ex) '기준 입원 명세서' ➔ '다음 입원 명세서' ➔ ('다음 입원 명세서') ...

- 본 가이드에서는 요양개시 · 종료일자를 이용하여 연속되는 명세서들을 하나의 에피소드로 구축하는 과정을 소개함

① 자료 정렬

- 동일 환자, 동일 요양기관별로 묶고, 연속되는 명세서를 파악하기 위해 요양개시일자, 요양종료일자별로 정렬함

```
/*[1단계] 환자 고유 식별자(pat_hpin), 요양기관기호(ykoho), 요양개시일자(recu_fr_dd), 요양종료일자(recu_to_dd)를 기준으로 정렬*/
proc sort data=temp1; by pat_hpin ykoho recu_fr_dd recu_to_dd; run;
```

	pat_hpin	ykoho	recu_fr_dd	recu_to_dd	amt
1	a	aa	20180101	20180121	500000
2	b	aa	20180101	20180120	600000
3	a	aa	20180201	20180228	4000000
4	a	bb	20180301	20180315	3000000
5	b	bb	20180302	20180320	5000000
6	b	bb	20180325	20180331	2000000
7	b	cc	20180401	20180415	5000000
8	c	cc	20180401	20180415	7000000
9	d	cc	20180301	20180331	800000
10	d	aa	20180331	20180425	6000000
11	d	aa	20180505	20180531	3500000
12	d	aa	20180601	20180620	2570000
13	a	bb	20180601	20180615	3500000
14	a	cc	20180601	20180630	560000
15	a	cc	20180701	20180731	8110000
16	b	aa	20180805	20180823	3540000
17	b	bb	20180702	20180712	5160000
18	c	bb	20180206	20180228	8552000
19	c	aa	20180401	20180423	9550000
20	c	aa	20180301	20180331	9800000

정렬 전

	pat_hpin	ykoho	recu_fr_dd	recu_to_dd	amt
1	a	aa	20180101	20180121	500000
2	a	aa	20180201	20180228	4000000
3	a	aa	20180301	20180315	3000000
4	a	bb	20180601	20180615	3500000
5	a	cc	20180601	20180630	560000
6	a	cc	20180701	20180731	8110000
7	b	aa	20180101	20180120	600000
8	b	aa	20180805	20180823	3540000
9	b	bb	20180302	20180320	5000000
10	b	bb	20180325	20180331	2000000
11	b	bb	20180702	20180712	5160000
12	b	cc	20180401	20180415	5000000
13	c	aa	20180301	20180331	9800000
14	c	aa	20180401	20180423	5550000
15	c	bb	20180206	20180228	8552000
16	c	cc	20180101	20180115	7000000
17	d	aa	20180331	20180425	6000000
18	d	aa	20180505	20180531	3500000
19	d	aa	20180601	20180620	2570000
20	d	cc	20180301	20180331	800000

정렬 후

② 변수 생성

- 에피소드별로 구분하기 위해 필요한 변수를 생성하고, 생성된 변수를 사용하여 에피소드별로 날짜, 비용 계산
 - ▶ 심사평가원의 날짜변수(요양개시 · 종료일자)는 문자형식으로 되어있어 날짜형식으로 변환하여 명세서간 간격을 구함
 - ▶ 에피소드 구분자를 생성하고, 이를 기준으로 날짜, 비용(누적합)을 계산함

```
/*[2단계] 입원 청구자료에 에피소드 구분자 생성*/
data temp2; set temp1;
by pat_hpjn ykiho;
format start_date end_date start_date2 end_date2 yymmdd10.;
/*문자형식을 날짜형식으로 변환: 요양개시일자(start_date), 요양종료일자(end_date)*/
start_date = input(recu_fr_dd, yymmdd8.); end_date = input(recu_to_dd, yymmdd8.);

/*retain 변수 생성: 요양개시일자(start_date2), 요양종료일자(end_date2), 비용 합계(tot_amt),
에피소드 구분자(eps)*/
retain start_date2 0 end_date2 0 tot_amt 0 eps 0;

/*에피소드 구분자(번호) 부여*/
/*요양기관기호가 다른 경우에는 다른 에피소드로 정의하고, 일자와 비용을 retain변수에 업데이트함*/
if first.ykiho then do;
    eps+1; start_date2=start_date; end_date2=end_date; tot_amt=amt; end;

/*동일 요양기관이고 청구건 간격이 1일 이내인 경우, 동일 에피소드로 정의함.
에피소드 내에서 요양개시일자의 최솟값과 요양종료일자의 최댓값을 retain 변수에 업데이트함.
비용을 누적합산하여 retain 변수에 업데이트함*/

else if start_date-end_date2(<=1 then do;
    eps=eps; start_date2=min(start_date, start_date2);
    end_date2=max(end_date, end_date2); tot_amt+amt; end;

/*동일 요양기관이면서 청구건 간격이 1일을 초과하는 경우에는 다른 에피소드로 정의함.
일자와 금액을 retain 변수에 업데이트함*/
else do; eps+1; start_date2=start_date; end_date2=end_date; tot_amt=amt; end; run;
```

	pat_hpjn	ykiho	recu_fr_dd	recu_to_dd	amt	start_date	end_date	start_date2	end_date2	tot_amt	eps
1	a	aa	20180101	20180131	300000	2018-01-01	2018-01-31	2018-01-01	2018-01-31	300000	1
2	a	aa	20180201	20180228	400000	2018-02-01	2018-02-28	2018-02-01	2018-02-28	400000	2
3	a	bb	20180301	20180315	300000	2018-03-01	2018-03-15	2018-03-01	2018-03-15	300000	3
4	a	bb	20180801	20180815	300000	2018-08-01	2018-08-15	2018-08-01	2018-08-15	300000	4
5	a	cc	20180601	20180630	500000	2018-06-01	2018-06-30	2018-06-01	2018-06-30	500000	5
6	a	cc	20180701	20180731	600000	2018-07-01	2018-07-31	2018-06-01	2018-07-31	607000	6
7	b	aa	20180101	20180120	600000	2018-01-01	2018-01-20	2018-01-01	2018-01-20	600000	7
8	b	aa	20180805	20180823	350000	2018-08-05	2018-08-23	2018-08-05	2018-08-23	350000	8
9	b	bb	20180302	20180320	500000	2018-03-02	2018-03-20	2018-03-02	2018-03-20	500000	9
10	b	bb	20180325	20180331	300000	2018-03-25	2018-03-31	2018-03-25	2018-03-31	2000000	10
11	b	bb	20180702	20180712	500000	2018-07-02	2018-07-12	2018-07-02	2018-07-12	500000	11
12	b	cc	20180401	20180415	500000	2018-04-01	2018-04-15	2018-04-01	2018-04-15	500000	12
13	c	aa	20180301	20180331	600000	2018-03-01	2018-03-31	2018-03-01	2018-03-31	600000	13
14	c	aa	20180401	20180423	550000	2018-04-01	2018-04-23	2018-03-01	2018-04-23	1500000	14
15	c	bb	20180206	20180228	850000	2018-02-06	2018-02-28	2018-02-06	2018-02-28	850000	15
16	c	cc	20180101	20180115	700000	2018-01-01	2018-01-15	2018-01-01	2018-01-15	700000	16
17	d	aa	20180331	20180425	600000	2018-03-31	2018-04-25	2018-03-31	2018-04-25	600000	17
18	d	aa	20180505	20180531	350000	2018-05-05	2018-05-31	2018-05-05	2018-05-31	350000	18
19	d	aa	20180601	20180620	250000	2018-06-01	2018-06-20	2018-05-05	2018-06-20	6070000	19
20	d	cc	20180301	20180331	800000	2018-03-01	2018-03-31	2018-03-01	2018-03-31	800000	20

※ 부분이 새로 생성된 변수이며, 세부 설명은 다음 그림 참고



① 날짜형식으로 변환
- recu_fr_dd ⇒ start_date, recu_to_dd ⇒ end_date

② 동일 에피소드 구분
- 기준에 따라 1번 ~ 16번 에피소드 생성

	pat_hpjn	ykho	recu_fr_dd	recu_to_dd	amt	start_date	end_date	start_date2	end_date2	tot_amt	eps
1	a	aa	20180101	20180731	500000	2018-07-01	2018-07-31	2018-07-01	2018-07-31	500000	1
2	a	aa	20180201	20180228	400000	2018-02-01	2018-02-28	2018-01-01	2018-02-28	4500000	1
3	a	bb	20180301	20180315	3000000	2018-03-01	2018-03-15	2018-03-01	2018-03-15	3000000	2
4	a	bb	20180601	20180615	350000	2018-06-01	2018-06-15	2018-06-01	2018-06-15	350000	3
5	a	cc	20180501	20180530	56000	2018-05-01	2018-05-30	2018-05-01	2018-05-30	56000	4
6	a	cc	20180701	20180731	811000	2018-07-01	2018-07-31	2018-06-01	2018-07-31	867000	4
7	b	aa	20180101	20180120	60000	2018-01-01	2018-01-20	2018-01-01	2018-01-20	60000	5
8	b	aa	20180605	20180623	354000	2018-06-05	2018-06-23	2018-06-05	2018-06-23	354000	6
9	b	bb	20180302	20180320	50000	2018-03-02	2018-03-20	2018-03-02	2018-03-20	50000	7
10	b	bb	20180225	20180331	2000000	2018-02-25	2018-03-31	2018-02-25	2018-03-31	2000000	8
11	b	bb	20180702	20180712	516000	2018-07-02	2018-07-12	2018-07-02	2018-07-12	516000	9
12	b	cc	20180401	20180415	500000	2018-04-01	2018-04-15	2018-04-01	2018-04-15	500000	10
13	c	aa	20180301	20180331	969000	2018-03-01	2018-03-31	2018-03-01	2018-03-31	969000	11
14	c	aa	20180401	20180423	555000	2018-04-01	2018-04-23	2018-03-01	2018-04-23	1524000	11
15	c	bb	20180206	20180228	8552000	2018-02-06	2018-02-28	2018-02-06	2018-02-28	8552000	12
16	c	cc	20180101	20180115	700000	2018-01-01	2018-01-15	2018-01-01	2018-01-15	700000	13
17	d	aa	20180331	20180425	600000	2018-03-31	2018-04-25	2018-03-31	2018-04-25	600000	14
18	d	aa	20180505	20180531	3500000	2018-05-05	2018-05-31	2018-05-05	2018-05-31	3500000	15
19	d	aa	20180601	20180620	2570000	2018-06-01	2018-06-20	2018-06-05	2018-06-20	6070000	15
20	d	cc	20180301	20180331	80000	2018-03-01	2018-03-31	2018-03-01	2018-03-31	80000	16

③ 동일 에피소드 내에서 변수 생성
- start_date2(최초 입력 날짜) = 첫번째 명세서(동일 에피소드 내)의 start_date
- end_date2(최종 퇴원 날짜) = 마지막 명세서(동일 에피소드 내)의 end_date
- tot_amt(입원 비용) = 모든 명세서(동일 에피소드 내)의 amt 합

※ 1. 2번 에피소드는 1번 에피소드와 환자가 동일(ⓐ)하고, 날짜가 연속(2. 28. - 3. 1.)되지만, 요양기관(1번: aa, 2번: bb)이 다르므로 다른 에피소드로 정의
2. 3번 에피소드는 2번 에피소드와 환자, 요양기관이 동일(ⓐ, bb)하지만, 날짜가 연속되지 않아 다른 에피소드로 정의

③ 에피소드 단위의 자료 생성

- 동일 에피소드의 경우 마지막 경우만 선택하여 중복되는 값 제거

```
/*[3단계] 에피소드 단위의 분석자료 생성*/
proc sort data=temp2; by pat_hpin ykiho eps; run;
data temp3; set temp2(drop=start_date end_date); by pat_hpin ykiho eps ;
if last.eps; run;
```

	pat_hpin	ykiho	recu_fr_dd	recu_lo_dd	amt	start_date2	end_date2	totLamt	eps
1	a	aa	20180201	20180228	4000000	2018-01-01	2018-12-31	4000000	1
2	a	bb	20180301	20180315	3000000	2018-03-01	2018-03-15	3000000	2
3	a	bb	20180801	20180815	3500000	2018-08-01	2018-08-15	3500000	3
4	a	cc	20180701	20180731	800000	2018-06-01	2018-07-31	800000	4
5	b	aa	20180101	20180120	600000	2018-01-01	2018-01-20	600000	5
6	b	aa	20180805	20180823	3540000	2018-08-05	2018-08-23	3540000	6
7	b	bb	20180302	20180320	5000000	2018-03-02	2018-03-20	5000000	7
8	b	bb	20180325	20180331	2000000	2018-03-25	2018-03-31	2000000	8
9	b	bb	20180702	20180712	5160000	2018-07-02	2018-07-12	5160000	9
10	b	cc	20180401	20180415	5000000	2018-04-01	2018-04-15	5000000	10
11	c	aa	20180401	20180423	5550000	2018-03-01	2018-04-23	15240000	11
12	c	bb	20180206	20180228	85520000	2018-02-06	2018-02-28	85520000	12
13	c	cc	20180101	20180115	7000000	2018-01-01	2018-01-15	7000000	13
14	d	aa	20180331	20180425	6000000	2018-03-31	2018-04-25	6000000	14
15	d	aa	20180601	20180620	25700000	2018-06-05	2018-06-20	60700000	15
16	d	cc	20180301	20180331	800000	2018-03-01	2018-03-31	800000	16

4. 맺음말

- 건강보험 청구자료를 분석할 때는 입원 자료의 청구 특성을 이해한 뒤 에피소드 단위로 분석하여야 보다 의미있는 결과를 얻을 수 있음
 - ▶ 에피소드를 구축하지 않고 분석할 경우에는 분리 청구, 요양기관기호 변경 등의 상황으로 인해 입원건수, 입원기간, 비용 등이 과대·과소 산출될 수 있음
- 본 가이드에서는 청구건 간 간격을 고려하여 입원 에피소드를 구축하는 방법을 설명하고 있지만, 연구자의 판단에 따라 다양한 구축 방법들이 설계될 수 있음
 - ▶ 특정 상병, 청구건 간 간격, 진찰료 청구 여부, 동일 요양기관 정의 등 여러 사항들을 고려하여 연구 성격에 맞는 에피소드를 구축하여야 함



tip 입원 에피소드 구축 시 주의할 점



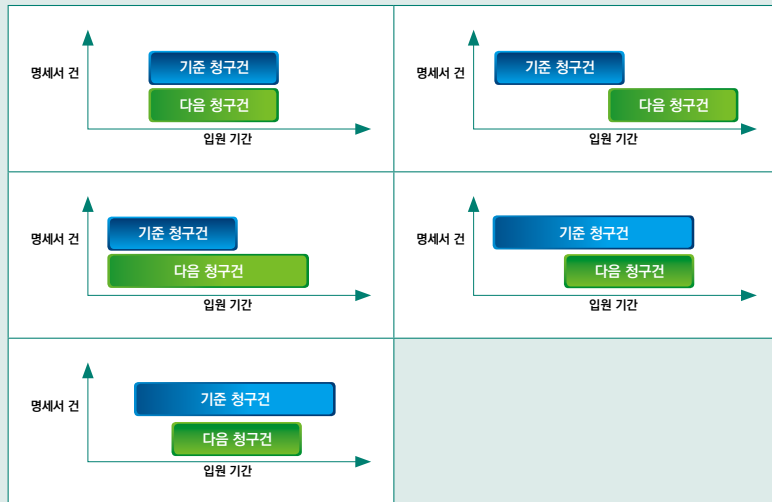
1. 특정 상병

- 동일한 입원이지만 입원 중 다른 질환(합병증, 전이 등)을 진료받을 경우 다른 주상병이 기재된 명세서가 포함될 수 있음
 - ▶ 다른 상병이 기재된 명세서를 포함한 에피소드를 구축할 것인지, 특정 상병의 명세서만으로 구축할 것인지 판단해야 함



2. 청구건 간격

- '기준 입원 명세서'와 '다음 입원 명세서'가 연속되지 않고, 입원 형태에 따라 중복, 포함되는 등 다양한 경우가 발생할 수도 있음
 - ▶ 연속되는 입원 형태의 경우 청구건 간 간격을 0 또는 1일로 설정하여 에피소드를 구축하지만, 아래와 같은 다른 형태의 명세서가 존재할 수 있으므로 사전에 고려할 필요가 있음



3. 재입원 여부

- 연속되는 입원을 재입원으로 정의할 경우 별도의 확인이 필요함
 - ▶ 재입원의 경우 진찰료(초진, 재진)가 청구되므로 진료내역에서 진찰료 청구여부를 확인한 후 에피소드 구축



4. 동일 요양기관의 요양기호 변경 여부

- 실제로 동일한 요양기관이지만 행정상의 이유(요양기관 종별, 대표자, 소재지 등의 변동)로 요양기관 기호가 변경되는 경우가 있음

👉 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

