



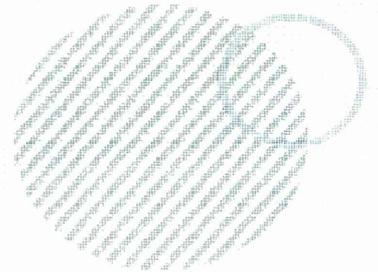
건강보험심사평가원 데이터 현황 및 활용

2014. 12



06

HIRA-NPS (HIRA National
Patient Sample)
건강보험심사평가원 환자표본자료



환자 표본자료 개요

- 개발 배경

- 건강보험 청구자료를 활용한 연구 수요(요구)의 급증
- 건강보험진료정보의 가치화 및 사회적 활용 증대(접근성, 편의성 확대) 요구
- 보건의료분야 국가정책 수립 및 학술연구 활성화를 위한 진료정보제공 인프라에 대한 필요성

- 개발 목적

- 건강보험 진료정보의 활용도 증대 및 표본자료를 활용한 보건의료 연구의 활성화
- 사전정보공개의 일환으로 일반연구자에게 중복된 통계자료 제공업무를 줄이고 자료를 효율적으로 사용할 수 있도록 사전에 가공, 공표된 자료제공(표본자료)

- 기대 효과

- 연구 활용으로 국가 및 사회적으로 편익과 효용성 제고

표본자료 필요성 및 외국의 예

- 대만과 미국의 경우 실제 청구 자료나 환자의 퇴원자료의 표본자료를 연구자들에게 제공하여 일반 연구자가 접근하기 힘든 실제 청구 자료에 대한 갈증을 해소하는데 일조하고 있음
- 우리나라는 아직까지 실제 청구데이터에 대한 자료제공을 하고 있지 않음
- 따라서 연구자로 하여금 청구데이터에 대한 접근성을 높여 다양한 연구분야에 활용할 수 있도록 “표본 자료 제공 방안”을 마련

표본자료 필요성 및 외국의 예

■ 국가별 표본자료 비교

국가별비교	우리나라(HIRA-NIS)	미국(NIS)	대만(NHIRD)
제공 대상	모든 연구자	모든 연구자	국가 연구기관 및 연구자 (일반인은 학습용 데이터셋이용)
추출방식	인구학적 층화 추출 (환자기준)	병원의 특성과 위치를 고려한 병원 추출 (병원기준)	인구학적 랜덤추출 (환자기준)
장 점	-분산에 따라 입원,외래의 비율 고려하여 자료의 효율성 증대	- 비급여 포함	-월별 입원 외래자료 및 다양한 세부 데이터 제공
단 점	-세부데이터 부재(개발 중) -비급여 부재	-표본 병원의 모든 퇴원자료이며 환자 구분코드가 존재하지 않음 -입원환자로 제한	-입원과 외래를 구분하지 않아 입원의 비율이 상대적으로 적어 비효율적임 -비급여 부재
샘플빈도	-입원 환자 70만 명(외래진료 포함) -외래 환자 50만 명	-약 700만건 정도의 기관 에피소드	-전체 등록자의 100만 명

환자 표본자료 종류 및 산출 기준

표본자료 종류	산출 기준
HIRA-NIS (2009-2011)	1년 단위 입원환자 약 70만 명(13%), 외래환자 약 40만 명(1%)
HIRA-NPS (2010-2011)	1년 단위 전체환자 약 140만 명(3%)
HIRA-APS (2010-2011)	1년 단위 65세 이상 환자 약 100만 명(20%)
HIRA-PPS (2010-2011)	1년 단위 20세 미만 환자 약 110만 명(10%)

※ 각 환자표본자료의 표본 한계치는 환자 수 150만 명 또는 영역별 20%이내를 기준으로 함

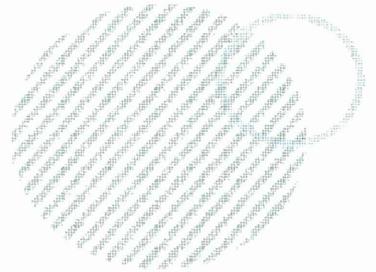
장·단점 및 제한점

- 장점
 - 분산에 따라 입원/외래 비율 고려하여 자료의 효율성 증대
 - 환자 인구학적 부분을 고려하여 다양한 에피소드 구축 가능
 - 청구자료에 대한 트레이닝
- 제한점
 - 표본자료이므로 추출빈도가 적은 희귀 질환은 설명력 감소
 - 1년간 표본자료이므로 유병기간이 긴 만성질환 연구 시 제한 존재



07

심평원 환자표본자료 검증



타당도 검증

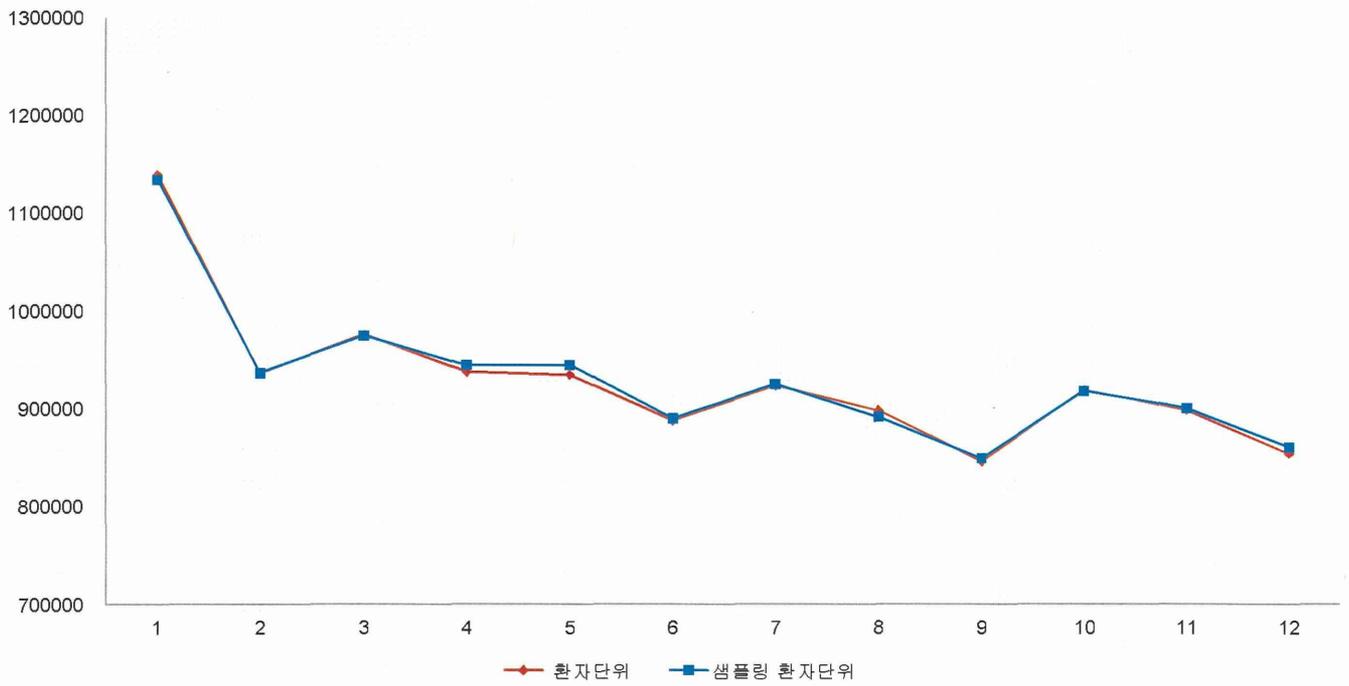
- 내부 타당도 평가 결과
 - 입원환자 100개 다빈도 질환에서 표본자료와 모집단의 비율이 동일
 - 표본자료와 모집단의 월별 환자당 평균 진료비 추이가 거의 일치하며 상대분산 0.98의 높은 일치율을 보임
 - 분산분석(일원분산분석)을 사용해 95%신뢰구간에서 표본자료와 모집단의 평균과 분산의 집단 간 차이가 없다는 가설 모두 채택
 - 표본자료를 이용한 만성신부전증(N18)환자의 모집단 추정 결과 95%신뢰구간 내에 참값이 존재

모집단 대비 표본 상병빈도

상병순위	상병코드	모집단	표본빈도	백분율
1	O800	188840	24484	12.97
2	J189	139603	18025	12.91
3	A09	127301	16373	12.86
4	I841	126508	16510	13.05
5	H259	79074	10243	12.95
6	O820	78332	10298	13.15
7	K359	74155	9581	12.92
8	M511	69381	9052	13.05
9	I639	59495	7847	13.19
10	I10	53087	6860	12.92
11	S335	52243	6681	12.79
12	H251	45861	5903	12.87
13	O821	43856	5739	13.09
14	N10	42737	5519	12.91
15	H258	42503	5645	13.28
16	J459	41512	5431	13.08
17	N393	39918	5257	13.17
18	J209	38260	4958	12.96
19	J180	36630	4675	12.76
20	J039	33588	4257	12.67
21	D259	32834	4311	13.13
22	M4806	32383	4122	12.73
23	I209	30714	4018	13.08
24	M512	29983	3866	12.89
25	I200	29760	3858	12.96
26	C169	29430	3819	12.98
27	J157	28704	3714	12.94
28	S3200	27184	3479	12.80
29	M170	26832	3497	13.03
30	E119	26500	3469	13.09

상병순위	상병코드	모집단	표본빈도	백분율
31	K409	25795	3319	12.87
32	C220	25289	3247	12.84
33	C73	25045	3309	13.21
34	I842	24851	3267	13.15
35	N390	24298	3183	13.10
36	I839	24094	3223	13.38
37	J219	23499	3000	12.77
38	K603	23485	3102	13.21
39	K635	23039	2992	12.99
40	C349	22936	2965	12.93
41	I638	22929	2919	12.73
42	K529	22593	2867	12.69
43	O814	20744	2579	12.43
44	O808	20468	2535	12.39
45	J342	20185	2697	13.36
46	S134	19930	2550	12.79
47	N40	19911	2556	12.84
48	J350	19600	2551	13.02
49	H250	19455	2566	13.19
50	N201	19411	2459	12.67
51	J159	18520	2418	13.06
52	O470	18495	2392	12.93
53	S832	18059	2322	12.86
54	I848	17786	2261	12.71
55	I849	17544	2340	13.34
56	I844	17292	2228	12.88
57	K210	16837	2165	12.86
58	K350	16315	2131	13.06
59	O828	15968	2097	13.13
60	D27	15904	1997	12.56

모집단과 표본자료의 환자당 월 청구액의 평균 비교 (입원환자 기준)



- 상대분산 0.9802 = 상대 효율 98%

모집단 추정 (예시)

▪ 만성신부전증(N18)

• 표본인원

연령구간	입원	외래
1	6	
2	7	
3	18	
4	26	2
5	25	4
6	106	10
7	167	15
8	260	30
9	348	44
10	559	66
11	719	65
12	793	68
13	856	68
14	1100	78
15	1063	62
16	1388	75
Total	7439	587

x 입원 7.692
x 외래 100

• 모집단 추정인원

연령구간	추정 입원	추정 외래	합계
1	46	0	46
2	54	0	54
3	138	0	138
4	200	200	400
5	192	400	592
6	815	1000	1815
7	1285	1500	2785
8	2000	3000	5000
9	2677	4400	7077
10	4300	6600	10900
11	5531	6500	12031
12	6100	6800	12900
13	6584	6800	13384
14	8461	7800	16261
15	8177	6200	14377
16	10676	7500	18176
Total	57236	58700	115936

모집단 추정

• 모집단 추정인원

연령구간	환자빈도
1	46
2	54
3	138
4	400
5	592
6	1815
7	2785
8	5000
9	7077
10	10900
11	12031
12	12900
13	13384
14	16261
15	14377
16	18176
Total	115936



• 모집단 인원

연령구간	환자 빈도	상대효율
1	48	0.96
2	88	0.61
3	192	0.72
4	430	0.93
5	625	0.95
6	1635	1.11
7	2802	0.99
8	4918	1.02
9	6696	1.06
10	9897	1.10
11	12298	0.98
12	11796	1.09
13	12808	1.04
14	14940	1.09
15	13858	1.04
16	17323	1.05
Total	110354	1.05

모집단 추정 (추정오차구간)

▪ 추정오차구간 계산(모비율 추정 사용)

표본의 비율은 모비율 p 에 대한 불편 추정량 이므로

$$\frac{\text{표본에서 해당 질환의 대상이 되는 환자수}}{\text{표본 전체 환자수}} \approx \text{해당 질환의 추정모비율 } (\hat{p})$$

모비율의 신뢰구간을 계산하기 위해 표본비율 \hat{p} 의 분산 추정량을 계산해 보면, 환자표본자료는 비복원 추출이므로

$$\hat{V}(\hat{p}) = \frac{\hat{p}\hat{q}}{n-1} \left(\frac{N-n}{N} \right)$$

이를 이용하여 다음과 같이 \hat{p} 의 95% 신뢰구간을 계산할 수 있다.

$$\hat{p} \pm 1.96 \sqrt{\hat{V}(\hat{p})}$$

모집단 추정 (추정오차구간)

- 모집단 추정오차구간을 구하기 위해 필요한 정보

- 전체 모집단 환자수 : 45,969,893 명

- 모집단 입원 환자 : 5,472,670명

- 표본입원 환자 : 711,457명

- 모집단 외래 환자 : 40,497,223명

- 표본 외래 환자 : 404,583명

※ 엑셀 참조

모집단 추정(추정오차구간)

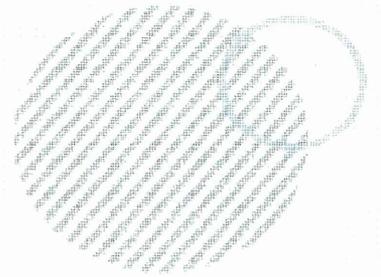
- 95%신뢰수준 모집단 추정오차구간 산출 결과
- 각 구간별로 모두 신뢰 구간 내에 참값이 존재

연령구간	95% 신뢰하한	95% 신뢰상한	모집단 환자 빈도
1	11	81	48
2	16	92	88
3	78	199	192
4	127	749	430
5	123	1,053	625
6	1,034	2,574	1,635
7	1,820	3,715	2,802
8	3,657	6,275	4,918
9	5,454	8,600	6,696
10	8,888	12,763	9,897
11	9,986	13,928	12,298
12	10,797	14,849	11,796
13	11,266	15,349	12,808
14	13,960	18,386	14,940
15	12,290	16,323	13,858
16	15,854	20,330	17,323
Total	109,263	121,252	110,354



08

심평원 환자표본자료 활용



환자표본자료 불러오기

이름	크기
nps_20_01.txt	1,859,877KB
nps_20_02.txt	1,834,448KB
nps_20_03.txt	1,588,722KB
nps_30_01.txt	1,809,742KB
nps_30_02.txt	1,815,078KB
nps_30_03.txt	1,820,283KB
nps_30_04.txt	1,842,506KB
nps_30_05.txt	1,842,708KB
nps_30_06.txt	1,845,508KB
nps_30_07.txt	1,842,300KB
nps_30_08.txt	1,841,623KB
nps_30_09.txt	1,844,902KB
nps_30_10.txt	1,844,240KB
nps_30_11.txt	1,840,156KB
nps_30_12.txt	1,312,997KB
nps_40_01.txt	1,853,752KB
nps_53_01.txt	1,816,214KB
nps_53_20.txt	1,746,090KB
ykiho.txt	4,323KB



NPS_2009	
nps_20_01.sas7bdat	
nps_20_02.sas7bdat	
nps_20_03.sas7bdat	
nps_30_01.sas7bdat	
nps_30_02.sas7bdat	
nps_30_03.sas7bdat	
nps_30_04.sas7bdat	
nps_30_05.sas7bdat	
nps_30_06.sas7bdat	
nps_30_07.sas7bdat	
nps_30_08.sas7bdat	
nps_30_09.sas7bdat	
nps_30_10.sas7bdat	
nps_30_11.sas7bdat	
nps_30_12.sas7bdat	
nps_40_01.sas7bdat	
nps_53_01.sas7bdat	
nps_53_02.sas7bdat	
ykiho.sas7bdat	



nps_20.sas7bdat
nps_30.sas7bdat
nps_40.sas7bdat
nps_53.sas7bdat
ykiho.sas7bdat

뇌졸중 환자 분석

- 조작적 정의 :

- 상병정의 : 진단코드 주,부상병기준 (I60, I61, I62, I63, I64, I67, I68, I69)
- 1단계 : 입원 청구자에 국한
- 2단계 : CT,MRI 촬영
- 3단계 : 입원 30일 이내에 뇌졸중 치료약제(항 혈전제 등) 사용

※ 건강보험청구자료를 이용한 급성심근경색 및 뇌졸중의 발생율·유병률 산출 방안 (2011) : 김수경, 유경호, 고민정, 홍재석, 김
록영, 이임봉, 김경훈

뇌졸중 환자 분석 1단계

- 상병정의 : 진단코드 주,부상병기준 (I60, I61, I62, I63, I64, I67, I68, I69)
- 1단계 : 입원 청구자에 국한

성별	구분	2009년			
		모집단 추정 값 (표본 X 가중치)	95% 신뢰하한	95%신뢰상한	모집단 환자수
남성	0-9세	354	257	451	358
	10-19세	377	276	477	460
	20-29세	969	808	1,130	847
	30-39세	2,977	2,695	3,259	3,079
	40-49세	10,523	9,993	11,053	10,981
	50-59세	20,623	19,881	21,364	20,285
	60-69세	26,892	26,046	27,738	27,192
	70-79세	29,084	28,204	29,964	28,871
	80-89세	11,984	11,419	12,550	11,894
90세이상	923	766	1,080	1,017	
여성	0-9세	331	237	425	303
	10-19세	400	297	503	368
	20-29세	692	556	828	671
	30-39세	1,592	1,386	1,799	1,743
	40-49세	6,700	6,277	7,123	6,750
	50-59세	12,969	12,381	13,558	12,944
	60-69세	22,192	21,423	22,961	22,588
	70-79세	40,707	39,667	41,747	41,323
	80-89세	25,192	24,373	26,011	25,355
90세이상	3,023	2,739	3,307	3,037	
Total		218,505	216,135	220,874	220,066