

HIRA-NPS: 患者データセットの定義及び紹介

□ 患者データセットの種類及び算出基準

患者データセットの種類及び算出基準

標本資料種類	算出基準
HIRA-NIS	1年単位入院患者約70万人(13%)、外来患者約40万人(1%)
HIRA-NPS	1年単位全体患者約140万人(3%)
HIRA-APS	1年単位65歳以上患者約100万人(20%)
HIRA-PPS	1年単位20歳未満患者約110万人(10%)

※ 各患者データセットの標本限界値は患者数150万人または領域別20%以内を基準とする。

□ 抽出方法：層化抽出(32区間)

○ 層化：性別(2区間) X 年齢群(5歳間隔 16区間)

○ 最大分散を持つ総療養給与費用に対して必要標本数算出

(汎用性と代表性を考慮して標本のサイズ算出)

参照) 入院患者標本開発に関する研究：国民健康保険請求資料を中心に、キム・ノギョン他、保健行

政学会誌 2013;23(2):152-161

提供年度

入院患者データセット(HIRA-NIS) : 2009 - 2013

全体患者データセット(HIRA-NPS) : 2010 - 2013

高齢患者データセット(HIRA-APS) : 2010 - 2013

小児青少年患者データセット(HIRA-PPS) : 2010 - 2013

※ 毎年 9 月末患者データセット年度更新

- 請求資料特性上 6 ヶ月の審査年月基準適用

提供対象

個人、一般研究者、機関など全て公開

- 審査評価院ホームページ、ファックス、電子メールを通じて申込書及びセキュリティ確約書受付

提供所要期間: 申込書及びセキュリティ確約書受付後、約 14 日以内行政
処理及び資料発送(CD)

各患者データセット用量 (テキストファイル基準) 及び手数料

手数料: 各年度別、種類別 300,000 ウォン

※ 手数料は追って変更されることがある。

連番	データセット名称	ハングル名称	用量(GB)	圧縮用量(GB)	手数料
1	HIRA-NIS-2009	入院患者データ セット	30.7	3.56	300,000 ウォン
2	HIRA-NIS-2010		27.4	3.21	300,000 ウォン
3	HIRA-NIS-2011		28.1	3.35	300,000 ウォン
4	HIRA-NIS-2012		25.0	3.25	300,000 ウォン
5	HIRA-NIS-2013		24.9	3.20	300,000 ウォン
6	HIRA-NPS-2010	全体患者データ セット	18.4	2.07	300,000 ウォン
7	HIRA-NPS-2011		18.8	2.14	300,000 ウォン
8	HIRA-NPS-2012		15.5	1.99	300,000 ウォン
9	HIRA-NPS-2013		16.3	2.00	300,000 ウォン
10	HIRA-APS-2010	高齢患者データ セット	31.9	3.35	300,000 ウォン
11	HIRA-APS-2011		33.7	3.56	300,000 ウォン
12	HIRA-APS-2012		27.9	3.49	300,000 ウォン
13	HIRA-APS-2013		29.1	3.63	300,000 ウォン
14	HIRA-PPS-2010	小児青少年 患者データセット	7.45	0.98	300,000 ウォン
15	HIRA-PPS-2011		10.2	1.13	300,000 ウォン
16	HIRA-PPS-2012		8.01	0.99	300,000 ウォン
17	HIRA-PPS-2013		6.27	0.78	300,000 ウォン
合計			359.63	42.68	5,100,000 ウォン

○ 口座番号: 100-200900-76804(ハナ銀行、預金主: 健康保険審査評価院)

□ 注意事項

○ 必ずシリアルナンバーを付与しなければならず、患者データセット申込書
上の研究責任者だけが使用可能(貸し出し及び再販売など禁止)

- 研究報告書、論文など申込書上の研究責任者氏名が記載されていなければ
ならない

- 申込書及びセキュリティ確約書保管、患者データセット提供管理台帳* 作
成

* 管理台帳には研究責任者情報、シリアルナンバー、提供日付などを含み

○ 患者データセットを活用した分析結果は、保険福祉部及び健康保険審査評
価院と関係がない

○ 購入に対する計算書を申し込む時、一般計算書発行及び提供

〈公共データ項目要請レイアウト〉

1. 製薬会社自社製品(他社製品) 要請項目明細

他社医薬品資料提供同意書可否 はい いいえ

- 提供同意書を受け取った会社のコード及び統合分類コードその他の事項に記載
- 同意書保有時、申し込み時ごとに添付資料として提出)

◎ 業社名[業社識別コード(4桁)]:					
◎ 診療期間: [提供可能な期間: 2013.01.01. - 2014.09.30.(のうち、最大で6ヶ月申し込み可能)]					
順番	コラム名	含むかどうか	順番	コラム名	含むかどうか
1	療養開始年月		13	書式コード	
2	処方診療区分コード		14	統合分類コード(製品名)	
3	DW 主傷病コード		15	薬効分類コード	
4	DW 副傷病コード		16	ATC コード	
5	受診者一連番号		17	主成分コード(一般名コード)	
6	受診者年齢(5歳単位)		18	診療科目コード	
7	性別		19	内科詳細科目コード	
8	保険者種別		20	来院日数	
9	療養機関一連番号		21	合計投与日数	
10	種別コード		22	総使用量または実施回数	
11	地域コード		23	金額	
12	総病床数(範疇化)		24	単価	
その他の事項 - 資料提供を同意した会社名: - 統合分類コード(製品コード):					

2. 成分別要請項目明細(一般提供用)

成分別に分析時、統合分類コード(製品名) 変数はマスキングされます。

◎ 診療期間: [提供可能な期間: 2013.01.01. - 2014.09.30.(のうち、最大6ヶ月申し込み可能)]					
順番	コラム名	含むかどうか	順番	コラム名	含むかどうか
1	療養開始年月		13	書式コード	
2	処方診療区分コード		14	統合分類コード(製品名)	
3	DW 主傷病コード		15	薬効分類コード	
4	DW 副傷病コード		16	ATC コード	
5	受診者一連番号		17	主成分コード(一般名コード)	
6	受診者年齢(5歳単位)		18	診療科目コード	
7	性別		19	内科詳細科目コード	
8	保険者種別		20	来院日数	
9	療養機関一連番号		21	総投与日数	
10	種別コード		22	総使用量又は実施回数	
11	地域コード		23	金額	
12	総病床数(範疇化)		24	単価	

抽出基準(択1)

- 薬効分類コード(最大3個申し込み可能):
- ATCコード(最大3個申し込み可能): → 3,4段コードで申請時、下の基準満足する下位5ダンコード提供
- 主成分コード(最大10個申し込み可能):

○「公共機関の情報公開に関する法律」と「個人情報保護法」に基づいて薬効分類だね、ATCコード、成分単位分析で下のような場合には提供することができません。

- 薬品目数または業社数3個以下の場合
- 算出資料の中で特定製品の金額が総請求金額の80%を越える場合

参考資料7



健康保険審査評価院 データ現況及び活用

2014年12月



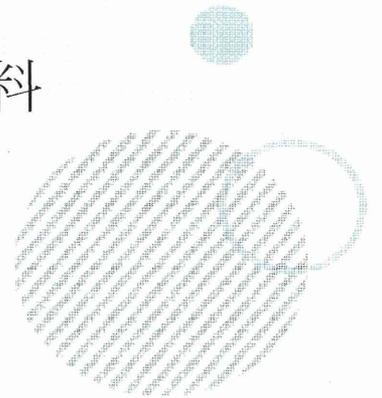
健康保険審査評価院

Health Insurance Review & Assessment Service



06

HIRA-NPS (HIRA National
Patient Sample)
健保審査評価院患者標本資料



患者標本資料概要

- **開発背景**

- 健保請求資料を活用した研究需要(要求)の急増
- 健保診療情報の価値化及び社会的活用増大(接近性、便宜性拡大) 要求
- 保健医療分野の国家政策樹立及び学術研究活性化のための診療情報提供インフラに対する必要性

- **開発目的**

- 健保診療情報の活用度増大及び標本資料を活用した保健医療研究の活性化
- 事前情報公開の一環で、一般研究者に重複した統計資料提供業務を減らし、資料を効率的に使うように前もって加工公表された資料提供(標本資料)

- **期待効果**

- 研究活用で国家及び社会的に便益と効用性向上

標本資料の必要性及び外国の例

- 台湾とアメリカの場合、実際の請求資料や患者の退院資料の標本資料を研究者たちに提供して一般研究者がアクセスしにくい実際の請求資料に対する渴望を解消するのに一助になっている
- 韓国はいまだに実際請求データに対する資料提供をしていない
- したがって、研究者にとって請求データに対するアクセス性を高めて多様な研究分野に活用できるように "標本資料提供案"を用意

標本資料必要性及び外国の例

■国家別標本資料比較

国家別備校	韓国(HIRA-NIS)	アメリカ(NIS)	台湾(NHIRD)
提供対象	すべての研究者	すべての研究者	国家研究機関及び研究者 (一般人は学習用データセット利用)
抽出方式	人口学的層化抽出 (患者基準)	病院の特性と位置を考慮した病院抽出 (病院基準)	人口学的ランダム抽出 (患者基準)
長所	-分散によって入院、外来の割合を考慮して資料の効率性増大	-非給与を含む	-月別入院外来資料及び多様な詳細データ提供
短所	-詳細データ不在(開発中) -非給与不在	-標本病院のすべての退院資料であり、患者区分コードが存在しない -入院患者に制限	-入院と外来を区分せず、入院の割合が相対的に少なく非効率的である -非給与不在
サンプル頻度	-入院患者70万人(外来診療含む) -外来患者50万人	-約700万件ほどの機関エピソード	-全体登録者の100万名

患者標本資料の種類及び算出基準

標本資料種類	算出基準
HIRA-NIS (2009-2011)	1年単位入院患者約70万人(13%), 外来患者約40万人(1%)
HIRA-NPS (2010-2011)	1年単位全体患者約140万人(3%)
HIRA-APS (2010-2011)	1年単位65歳以上の患者約100万人(20%)
HIRA-PPS (2010-2011)	1年単位20歳未満患者約110万人(10%)

※ 各患者標本資料の標本限界値は患者数150万人または領域別20%以内を基準とする

長所・短所及び制限点

- 長所

- 分散によって入院/外来の割合考慮して資料の効率性増大
- 患者人口学的部分を考慮して多様なエピソード構築可能
- 請求資料に対するトレーニング

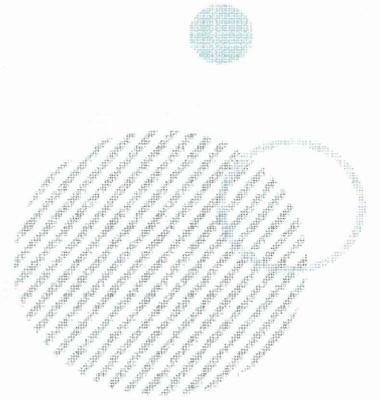
- 制限点

- 標本資料なので、抽出頻度が少ない珍しい疾患は説明力減少
- 1年間の標本資料なので、有病期間が長い慢性疾患、研究時制限存在



07

審評院患者標本資料検証



妥当度検証

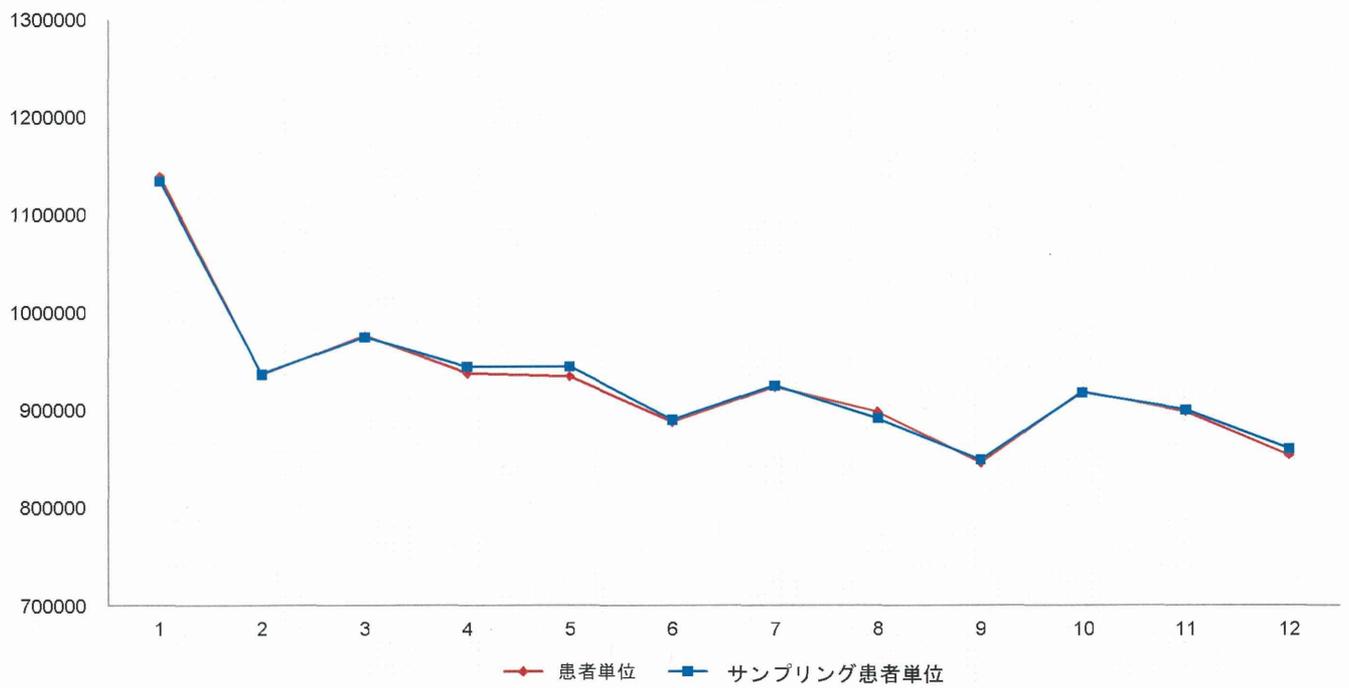
- 内部妥当度の評価結果
 - 入院患者100個多頻度疾患で標本資料と母集団の割合が同一
 - 標本資料と母集団の月別患者当たりの平均診療費の推移がほとんど一致し、相手分散0.98の高い一致率が見られる
 - 分散分析(一元分散分析)を使って95%信頼区間で標本資料と母集団の平均と分散の集団間の差がないという仮説を全て採択
 - 標本資料を利用した慢性腎不全症(N18)患者の母集団推定結果、95%信頼区間内に本当の値が存在

母集団に対する標本傷病頻度

傷病順位	傷病コード	母集団	標本頻度	百分率
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

傷病順位	傷病コード	母集団	標本頻度	百分率
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

母集団と標本資料の患者当りの月請求額の平均比較 (入院患者基準)



- 相対分散0.9802 = 相対効率98%

母集団推定 (例示)

慢性腎不全症(N18)

•標本人数

年齢区間	入院	外来
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



 × 入院7.692
 × 外来100

•母集団推定人数

年齢区間	推定入院	推定外来	合計
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

母集団推定

・母集団推定人数

年齢区間	患者頻度
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



・母集団人数

年齢区間	患者頻度	相手効率
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