

국민건강보험공단 빅데이터

자료구조와 활용사례

김형수

건국대학교 예방의학교실

0. 시작하면서

- 연구자 입장에서
- 국민건강보험공단의 빅데이터를 이용하여
- 부인종양학 분야의 연구를 수행할 수 있을까?

I. 서론

- 건강보험제도 소개
- 국민건강보험 빅데이터 구조

I. 서론 - 건강보험제도 소개

건강보험의 특징

모든 의료 공급자에 대한 **당연지정제도 운영***

· 의료 서비스 기관은 제공자로 당연지정(1979)

* 민간이 전체 의원의 97%, 병원의 94%, 병상의 91%를 차지

단일보험자가 **전 국민의 건강보험 보장**

· 전 국민 의료보험 실현(1989)

· 조합방식의 보험자가 단일보험자인 국민건강보험으로 통합(2000)

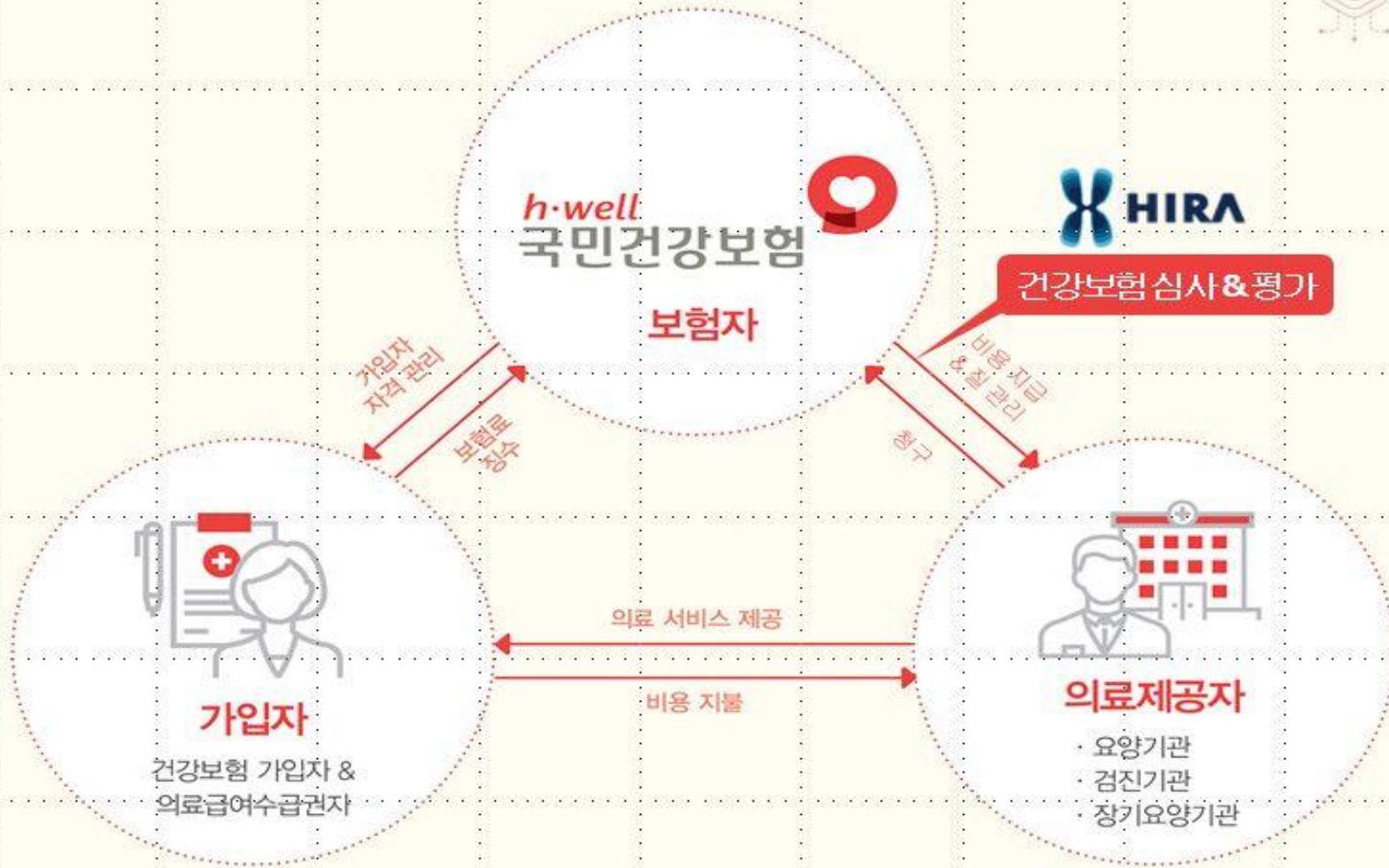
행위별 수가제가 주된 **지불보상제도**

생애주기별 **국가건강검진 제공**



I. 서론 - 건강보험제도 소개

건강보험제도 운영 총괄표



I. 서론 - 건강보험빅데이터 구성

자격 자료원



I. 서론 - 건강보험빅데이터 구성

보험료 징수 자료원



I. 서론 - 건강보험빅데이터 구성

의료이용 및 평가 자료원



요양기관



청구 정보



요양기관 정보



검진기관 정보

건강보험
심사평가원



심사 정보



의료서비스
평가 정보

근로
복지공단



산업재해 정보



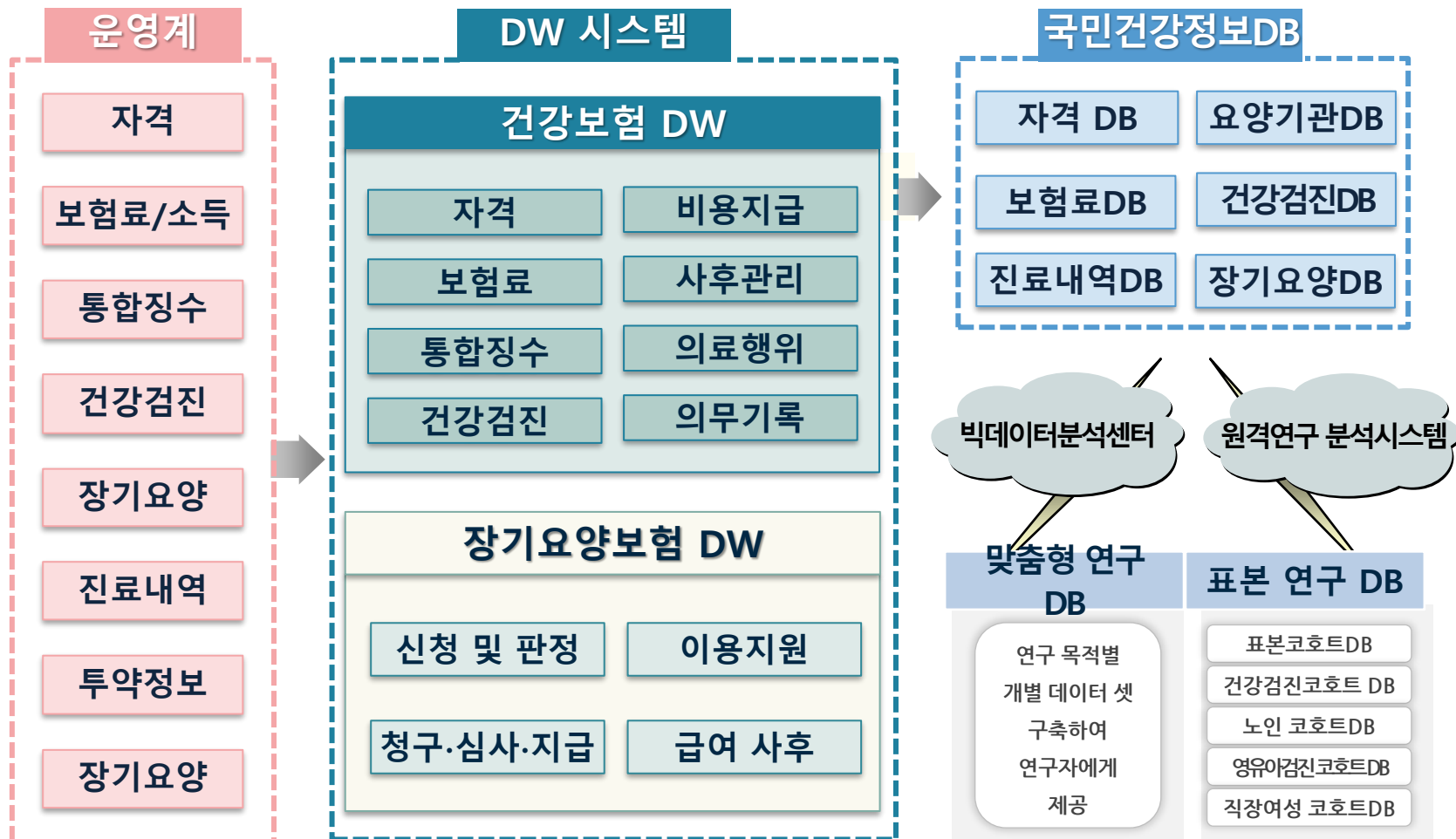
I. 서론 - 건강보험빅데이터 구성

전산화된 청구 시스템



I. 서론 - 건강보험빅데이터 구성

연구용 DB 시스템



I. 서론 - 건강보험빅데이터 변수

건강보험 빅데이터의 주요 변수 - 1



인구사회학적 변수



성별

- 남 / 녀



나이

- 수진시점/ 연도말



출생

- 출생일자



사망

- 사망일자



장애

- 장애유형(15개)
- 장애등급(6개)
- 등록일자



거주지역

- 환자, 의료기관, 사업장
- 시군구 수준/지오코딩



가입자 유형

- 직장가입자 및 피부양자
- 지역세대주 및 세대원
- 의료급여수급권자



보험료

- 금액
- 10분위
- 조정 금액



사업장 유형

- 산업 분야
- 사업장 규모

I. 서론 - 건강보험빅데이터 변수

건강보험 빅데이터의 주요 변수 - 2



건강행태 및 검사

	건강 행태
흡연	흡연력
	양 / 기간
음주	음주 이력
	양 / 기간
신체활동	걷기 / 보통 정도 / 격렬
질병과거력	심혈관계 질환 과거력 및 가족력 암 과거력 및 가족력
정신건강	인지 / 우울

	생의학적 검사
신체계측	신장 / 체중 / 허리둘레
	시각 & 청각 측정
	혈압 (수축기 / 이완기)
혈액검사	공복혈당 / 콜레스테롤 / HDL / LDL / 중성지방
	간 기능 / 신장 기능 검사
암	영상자료 (조영검사, 초음파 등)
	내시경검사
	생검

I. 서론 - 건강보험빅데이터 변수

건강보험 빅데이터의 주요 변수 - 3



의료이용 및 요양기관 변수

	의료이용
질병	ICD-10 코드
의료이용	내원 / 외래
	응급실 / 중환자실
	재원기간
	처방일수
비용	공단부담 / 본인부담
진료상세	의료 행위 및 치료재료
약물	약물 주성분
	약물 제품명
	투약량

	요양기관
인력	의사, 간호사, 치료사의 수
	의료인력 현황 및 정보
시설	병실 / 병상 / 중환자실 / 응급실 등 정보
장비	CT / MRI / PET 등 보유현황

I. 서론 - 건강보험빅데이터 변수

건강보험 빅데이터의 주요변수 - 4

영유아 건강검진 변수

	영유아 건강검진
출생 현황	생년월일 / 미숙아 여부 / 분만예정일 / 출생시 체중
예방접종	BCG / B형간염 / DPT / 소아마비 / 폐구균 / 헤모필루스 / MMR / 수두 / 일본뇌염 / A형간염
기본문진	시각 / 청각 / 안전 / 영양 / 수면 / 돌연사예방 / 구강 / 대소변가리기 / 정서및사회성 / 전자미디어노출 / 개인위생 / 취학준비 / 간접흡연 / 감염
신체측정	키 / 몸무게 / 체질량지수 / 머리둘레 / 시진 / 시력 / 청각
발달평가	K-ASQ / DENVER-II / K-DST
구강	

I. 서론 - 건강보험빅데이터 변수

건강보험 빅데이터의 주요변수 - 5

포괄간호 및 차등제등급 변수

	요양기관 현황		차등제 등급유형
포괄간호 서비스	서비스여부	일반	건강보험 / 의료급여
	서비스시작일자	중환자실	성인소아
	서비스종료일자		신생아
	기타항목간호등급코드	요양병원	간호인력 & 간호인력가산 여부
		의사인력 & 의사가산여부	
		필요인력가산여부	
		완화병동	간호인력
			사회복지가산여부

I. 서론 - 건강보험빅데이터 변수

건강보험 빅데이터의 주요변수 - 6

장기요양보험 변수

	장기요양보험
인정관리	인정조사 / 의사소견서 / 등급판정 등
이용지원	이용지원상담 / 지역사회자원연계 등
자원 및 평가	장기요양기관 유형 / 종사자 / 복지용구제품 / 수용인원 등
사전급여	복지용구조사 등
청구/심사	청구 / 심사 명세서 / 수가상세내역 등
지급	지급 명세서 / 수가상세내역 등
급여사후	체납 / 현지조사 등

건강보험자료 공유서비스

근거중심의 보건 의료 정책 및
학술연구 지원을
선도합니다.



로그인

[회원가입 >](#)
[아이디/비밀번호찾기 >](#)

건강보험 빅데이터 표본연구 DB

국민건강정보자료를 활용하여 구축한 표본
코호트 DB 제공



자세히보기 +

건강보험빅데이터 맞춤형연구 DB

맞춤형 건강정보자료를 열람 및 연구 분석
할 수 있도록 맞춤형연구 DB 제공



자세히보기 +

원격연구포털

국민건강정보자료를 활용하여 구축한 표본
코호트 DB 제공



바로가기 +

의료이용지표

만성질환 발생 전후를 관리할 수 있는 모니터링 지표



바로가기 >

통계

지역별, 연령별, 성별, 기간별로 분류된 다양한 통계정보 제공



조회하기 >

의료이용지표순위

- 1 흡연율
- 2 음주율
- 3 비만율
- 4 권장신체활동수행률
- 5 복부비만율

한국인비만지수참조표준

한국인의 비만지수, 체질량지수, 허리둘레의 평균과 표준불확도 및 분포내역



자세히보기 +

공지사항 >

- [홈페이지이용] (매뉴얼)빅데이터분석센터 PC예약 매뉴얼 ...
2019-04-24 15:44:49
- [홈페이지이용] ★빅데이터 분석센터 연구자로 이관 관련 ...
2019-04-16 16:22:24
- [기타] 2019년 2차수 건강보험 빅데이터 활용 ...
2019-03-20 09:59:18
- [기타] 1차 빅데이터 활용 교육 실습 자료
2019-02-26 13:38:01
- [기타] 자료 분석결과 반출기한 변경 안내
2019-01-31 17:39:24

문의하기

- [표본연구DB]** (033)736-2431, 2433
[맞춤형연구DB] (033)736-2432, 2433
[통계] (033)736-2444
[의료이용지표] (033)736-2430, 2434

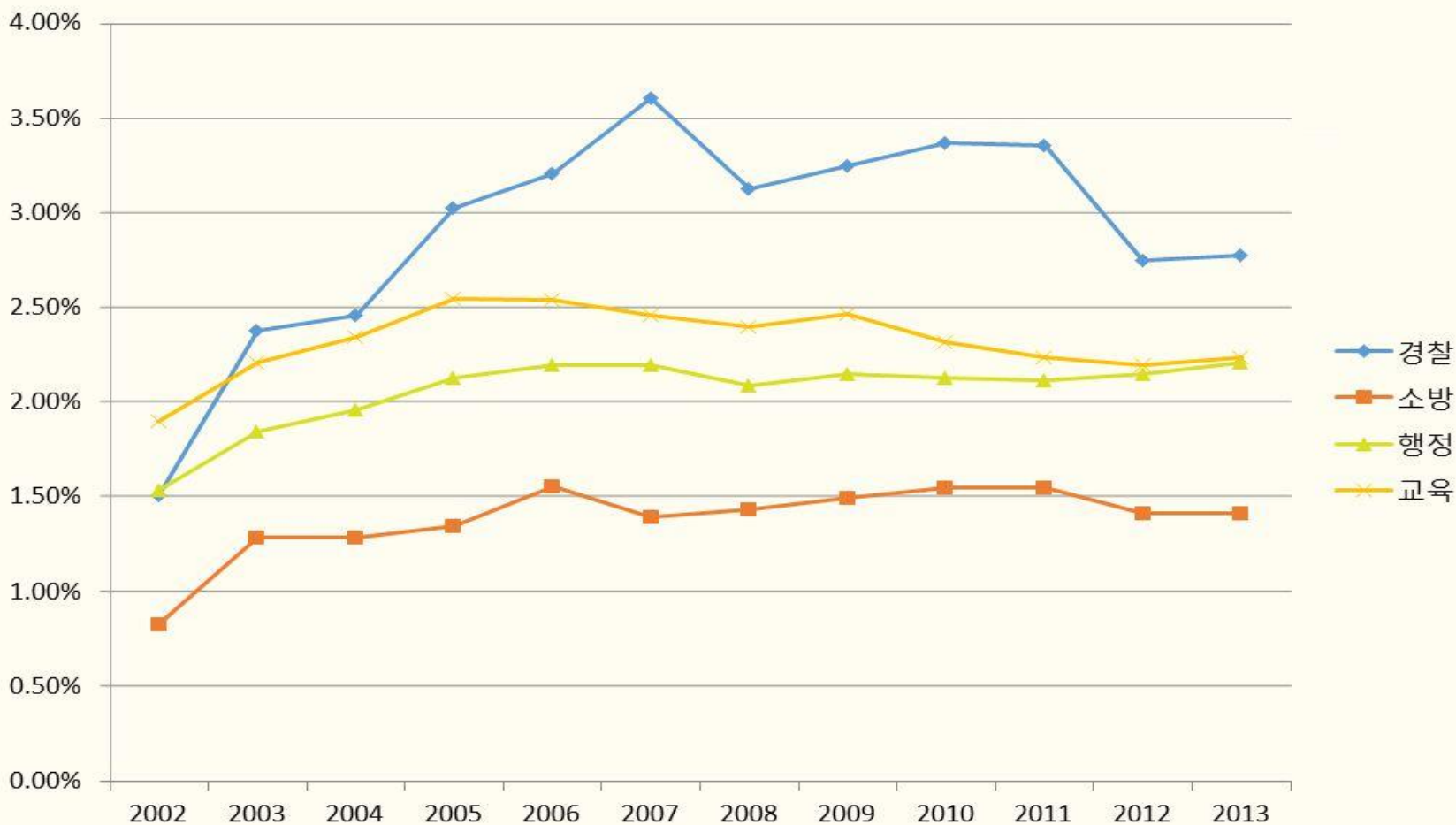
II. 본론; 빅데이터의 활용사례

- 정책연구; 국가, 시도, 시군구
- 개별 연구자의 관심

정책연구 사례 1

산업보건: 직종별 건강 차이

업종별 협심증 유병률



정책연구 사례 2

일-가정 양립정책 모니터링

임신 → 분만 → 산후 직장가입 유지율



임신 연도	2006	2014	비고
직장가입자 수	147,963	186,846	26.3% ↑
분만까지 유지	67.7%	73.8%	9.0% ↑
분만 후 3개월까지 유지	65.2%	71.2%	9.3% ↑
분만 후 6개월까지 유지	64.1%	70.4%	9.7% ↑
분만 후 12개월까지 유지	62.7%	69.8%	11.3% ↑

- 사업장 규모가 크거나 소득수준이 높은 산모일수록
분만 후 12개월까지 직장 자격을 유지하는 경향이 커짐
→ 산모의 직장소득이 일-가정 양립에 영향을 미침

정책연구 사례 3

분만취약지: 강원도 자체충족률

거주지	총 분만(명)	시군구 내 분만(명)	시도 내 분만(명)	시군구 자체충족률	시도 자체충족률
춘천시	1,815	1,422	1,505	78%	83%
원주시	2,272	1,755	1,833	77%	81%
강릉시	1,155	881	965	76%	84%
속초시	493	294	376	60%	76%
동해시	597	339	454	57%	76%
강원도	9,246	4,849	6,780	52%	73%
전국	380,405	145,327	289,209	38%	76%
삼척시	359	115	264	32%	74%
홍천군	364	30	253	8%	70%
태백시	229	13	156	6%	68%
횡성군	171	0	127	0%	74%
양양군	118	0	87	0%	74%
평창군	180	0	121	0%	67%
고성군	162	0	106	0%	65%
정선군	197	0	120	0%	61%
양구군	204	0	115	0%	56%
화천군	234	0	121	0%	52%
인제군	225	0	115	0%	51%
영월군	141	0	32	0%	23%
철원군	330	0	30	0%	9%

정책연구 사례 4

분만취약지: 강원도 산모 거주지~분만기관 거리

거주지	거리 평균값(km)	거리 중앙값(km)
속초시	37.91	2.11
강릉시	26.97	2.73
춘천시	20.50	2.78
원주시	20.73	3.38
동해시	38.87	4.18
전국	21.39	4.35
강원도	36.40	4.68
삼척시	48.90	9.70
양양군	50.94	15.79
횡성군	34.54	17.90
고성군	68.38	23.00
홍천군	48.77	24.46
영월군	48.91	24.72
화천군	73.44	29.18
양구군	77.70	36.15
평창군	53.73	40.14
인제군	78.54	48.64
태백시	70.82	49.20
철원군	74.08	52.31
정선군	66.73	57.74

부인종양학 영역 활용사례 1_1

Asian Nursing Research 12 (2018) 9–16

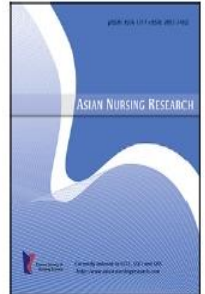


Korean Society of
Nursing Science

Contents lists available at [ScienceDirect](#)

Asian Nursing Research

journal homepage: www.asian-nursingresearch.com



Research Article

Contributions and Limitations of National Cervical Cancer Screening Program in Korea: A Retrospective Observational Study



Jung Hyun Lee, MPH,¹ Hyeongsu Kim, M.D., PhD,^{1,*}☆ Heejung Choi, RN, Ph.D.,²
Hyoseon Jeong, Ph.D.,¹ Young Ko, RN, Ph.D.,³ Seung-Hyuk Shim, M.D.,⁴
Eunjoo Lee, Ph.D.,⁵ Su Hyun Chae, M.D.⁴

¹ Department of Preventive Medicine, School of Medicine, Konkuk University, Seoul, Republic of Korea

² Department of Nursing, College of Nursing, Konkuk University, Seoul, Republic of Korea

³ Department of Nursing, College of Nursing, Gachon University, Incheon, Republic of Korea

⁴ Department of Obstetrics and Gynecology, School of Medicine, Konkuk University, Seoul, Republic of Korea

⁵ Bigdata Steering Department, National Health Insurance Service, Wonju, Republic of Korea

부인종양학 영역 활용사례 1_2

the number of patients with detected cervical cancer (true positive) per 100 whose screening results were positive (total number of positive in screening).

In a previous study [22], the interval cancer referred to cervical cancer detected ≤ 12 months after a negative screening result. However, in the present study, the interval cancer refers to cancer treatment ≤ 6 months after a negative result, based on our definition of cervical cancer “detection”.

Positive rate of the screening test (%)

$$= \frac{\text{Positive on screening (a + b)}}{\text{Cervical cancer screening participants (N)}} \times 100$$

Crude detection rate of the screening test

$$= \frac{\text{Tested positive and detected cervical cancer (a)}}{\text{Cervical cancer screening participants (N)}} \times 100,000$$

Incidence rate of interval cancer

$$= \frac{\text{Detected cervical cancer (c)}}{\text{Negative on screening (c + d)}} \times 100,000$$

Sensitivity of the screening test(%)

$$= \frac{\text{Positive on screening(a)}}{\text{Detected cervical cancer(a + c)}} \times 100$$

Specificity of the screening test(%)

$$= \frac{\text{Negative on screening(d)}}{\text{Not detected cervical cancer(b + d)}} \times 100$$

Positive predictive value of the screening test(%)

$$= \frac{\text{Detected cervical cancer(a)}}{\text{Positive on screening(a + b)}} \times 100$$

Data analysis

Data analyses were carried out using SAS software (ver. 9.4, SAS Institute, Cary, NC, USA). Using cervical cancer screening data and health insurance claim data, the positive rate of the screening test, crude detection rate, and the incidence rate of interval cancer, the sensitivity, specificity, and positive predictive value of the screening test were calculated.

Results

Positive rate of the screening test for cervical cancer

The positive rate of the screening test increased from 1.6% in 2009 to 2.4% in 2014 (Table 1). During the same period, the positive rate of the screening test increased for all ages. Among age group, the positive rate of the screening test was highest for women in their 40s (1.8% in 2009 and 2.8% in 2014) and was lowest for women in their 60s (1.3% in 2009 and 1.8% in 2014).

Crude detection rate of the screening test on cervical cancer

The crude detection rate of the screening test per 100,000 participants increased from 100.7 in 2009 to 102.1 in 2014 (Table 2). Dividing the cervical cancer into CIS and malignant neoplasm, while the crude detection rate on CIS increased from 76.7 to 79.6, the crude detection rate on malignant neoplasm of the cervix

부인종양학 영역 활용사례 1_3

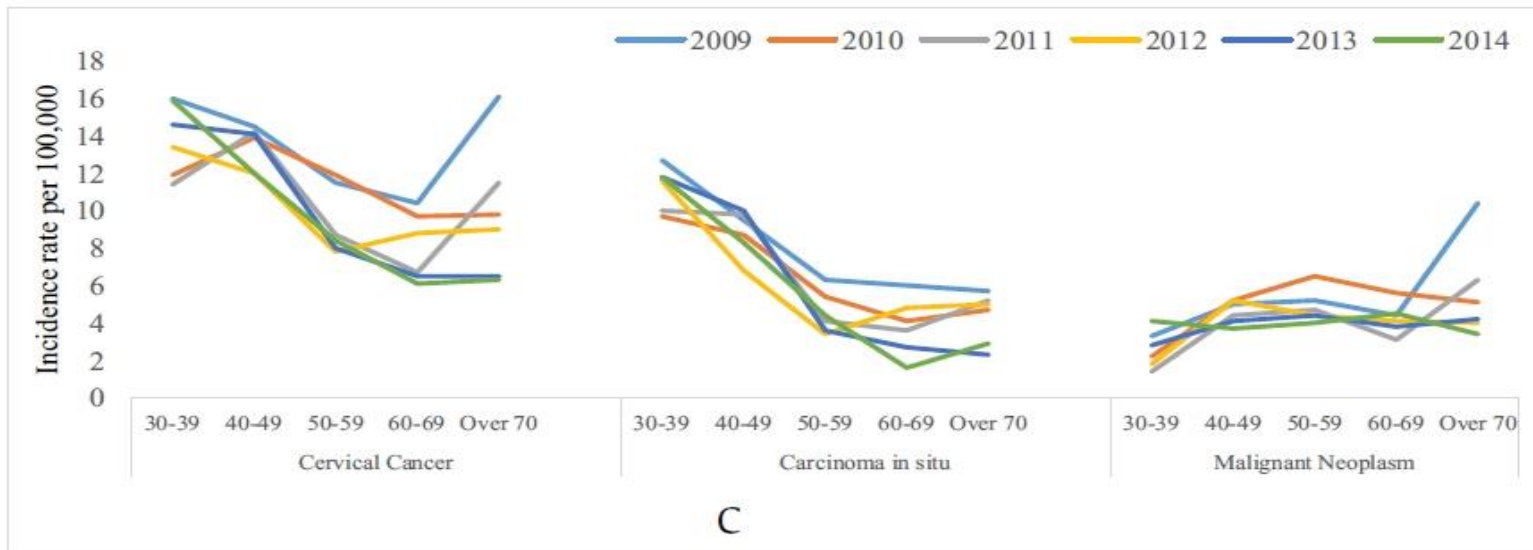
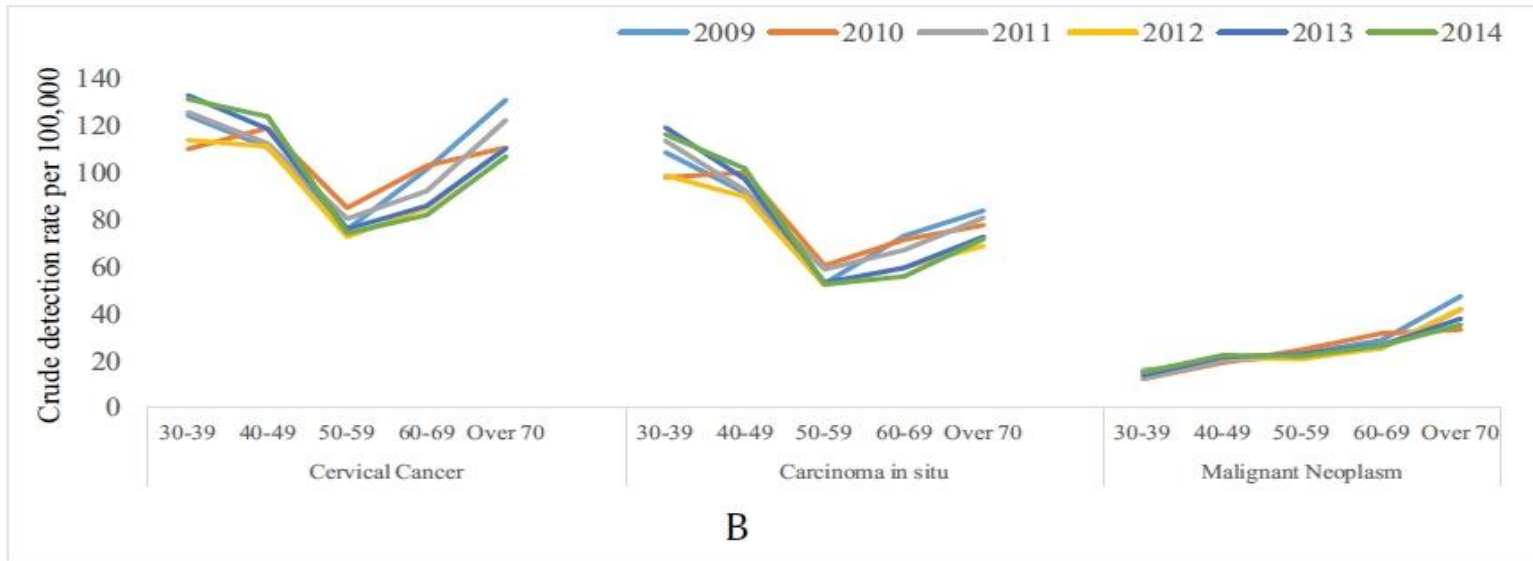


Figure 1. (A) Result matrix of detection, interval cancer, and screening test. (B) Change of crude detection rate of cervical cancer, carcinoma *in situ*, and malignant neoplasm by age-group. (C) Change of incidence rate of interval cancer by age-group.

부인종양학 영역 활용사례 1_4

Table 4 Sensitivity, Specificity and Positive Predictive Value of Screening Test for Cervical Cancer from 2009 to 2014.

	2009	2010	2011	2012	2013	2014
Sensitivity ^a	88.7	89.8	90.7	90.5	90.9	91.2
Specificity ^b	98.5	98.3	98.1	98.2	97.9	97.7
Positive predictive value ^c	5.6	5.1	5.0	4.8	4.3	4.3

^a Sensitivity = the number of positive cervical cancer screening results among 100 detected cases of cervical cancer.

^b Specificity = the number of negative results on cervical cancer screening among 100 cases where cervical cancer was not detected.

^c Positive predictive value = the number of patients with detected cervical cancer per 100 whose screening results were positive.

J Gynecol Oncol. 2017 Sep;28(5):e63
https://doi.org/10.3802/jgo.2017.28.e63
pISSN 2005-0380·eISSN 2005-0399

JGO JOURNAL OF
GYNECOLOGIC
ONCOLOGY

Original Article



Nationwide cervical cancer screening in Korea: data from the National Health Insurance Service Cancer Screening Program and National Cancer Screening Program, 2009–2014

OPEN ACCESS

Seung-Hyuk Shim ,¹ Hyeongsu Kim ,² In-Sook Sohn ,¹ Han-Sung Hwang,¹ Han-Sung Kwon,¹ Sun Joo Lee,¹ Ji Young Lee,¹ Soo-Nyung Kim,¹ Kunsei Lee,² Sounghoon Chang ²

Received: Jan 31, 2017
Revised: Apr 20, 2017
Accepted: May 17, 2017

¹Department of Obstetrics and Gynecology, Konkuk University School of Medicine, Seoul, Korea
²Department of Preventive Medicine, Konkuk University School of Medicine, Seoul, Korea

부인종양학 영역 활용사례 2_2

Table 1. Pap smear results by timeline from the NHISCSP and the NCSP for cervical cancer (2009–2014)

Variables	Year					
	2009	2010	2011	2012	2013	2014
Target population	6,523,862	6,182,627	7,710,765	7,314,299	7,535,004	7,527,179
Total Pap test	2,680,986	2,694,573	3,356,852	3,454,471	3,514,542	3,877,774
Participation rate (%)	41.10 (41.06–41.13)	43.58 (43.54–43.62)	43.53 (43.50–43.57)	47.23 (47.19–47.27)	46.64 (46.61–46.68)	51.52 (51.48–51.55)
Results						
Negative for intraepithelial Epithelial cell abnormality	2,583,683 (96.37)	2,592,248 (96.20)	3,237,865 (96.45)	3,338,264 (96.63)	3,387,482 (96.38)	3,731,970 (96.24)
Squamous cell	54,351 (2.02)	59,804 (2.22)	74,941 (2.23)	66,636 (1.93)	75,800 (2.16)	91,828 (2.37)
ASC-US	26,196 (0.98)	31,636 (1.17)	43,315 (1.29)	49,627 (1.44)	60,783 (1.73)	73,149 (1.89)
ASC-H	6,950 (0.26)	3,722 (0.14)	4,538 (0.14)	4,212 (0.12)	4,412 (0.13)	4,872 (0.13)
LSIL	5,658 (0.21)	6,133 (0.23)	9,506 (0.28)	7,595 (0.22)	8,595 (0.24)	9,421 (0.24)
HSIL	3,996 (0.15)	4,149 (0.15)	4,809 (0.14)	4,620 (0.13)	4,989 (0.14)	5,472 (0.14)
Squamous cell cancer	604 (0.02)	548 (0.02)	514 (0.02)	449 (0.01)	444 (0.01)	414 (0.01)
Glandular cell						
AGC	2,574 (0.10)	2,404 (0.09)	2,667 (0.08)	2,699 (0.08)	3,534 (0.10)	4,013 (0.10)
AIS	81 (0.00)	27 (0.00)	56 (0.00)	59 (0.00)	63 (0.00)	75 (0.00)
Adenocarcinoma	42 (0.00)	45 (0.00)	40 (0.00)	47 (0.00)	71 (0.00)	75 (0.00)
Infection/reactive*	977,727 (36.47)	1,093,153 (40.57)	1,513,532 (45.09)	1,654,322 (47.89)	1,764,418 (50.20)	1,987,713 (51.26)
Total No. of abnormal results	17,331	14,624	19,463	16,982	1,857	20,329
Abnormal rate [†] (%)	0.650 (0.637–0.656)	0.540 (0.534–0.551)	0.580 (0.572–0.588)	0.490 (0.484–0.499)	0.530 (0.521–0.536)	0.520 (0.517–0.531)
ASC-US:SIL [‡]	2.71	3.08	3.03	4.06	4.47	4.91

Values are presented as number (%) or median (95% CI).

AGC, atypical glandular cell; AIS, adenocarcinoma in situ; ASC-H, atypical squamous cells-cannot exclude high-grade squamous intraepithelial lesion; ASC-US, atypical squamous cells of undetermined significance; CI, confidence interval; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; NCSP, National Cancer Screening Program; NHISCSP, National Health Insurance Service Cancer Screening Program; Pap, Papanicolaou; SIL, squamous intraepithelial lesion.

*The presence of organisms such as *Trichomonas vaginalis* or fungal organisms consistent with *Candida* species will be included as a comment in this category. Components that are optionally listed in the category include atrophy, radiation, and inflammation; [†]The rates of abnormal results in cervical cancer screening were calculated by dividing the number of abnormal Pap test results by the number of participants, with a 95% CI; [‡]The ratio of ASC-US to SIL (ASC-US:SIL ratio) was calculated by dividing the number of ASC-US cases by the number of all SIL cases (including LSIL and HSIL).

부인종양학 영역 활용사례 3_1 (학술지 발표 준비중)

자궁암 검진

- 1) 1회, 2회, 3회차에서 연속 정상 판정을 받은 집단 (G1)
- 2) 1회차 미실시, 2회, 3회차에서 정상 판정을 받은 집단 (G2)
- 3) 1회, 2회차 미실시, 3회차서 정상 판정을 받은 집단 (G3)
- 4) 1회, 2회, 3회차 모두 미실시한 집단 (G4)

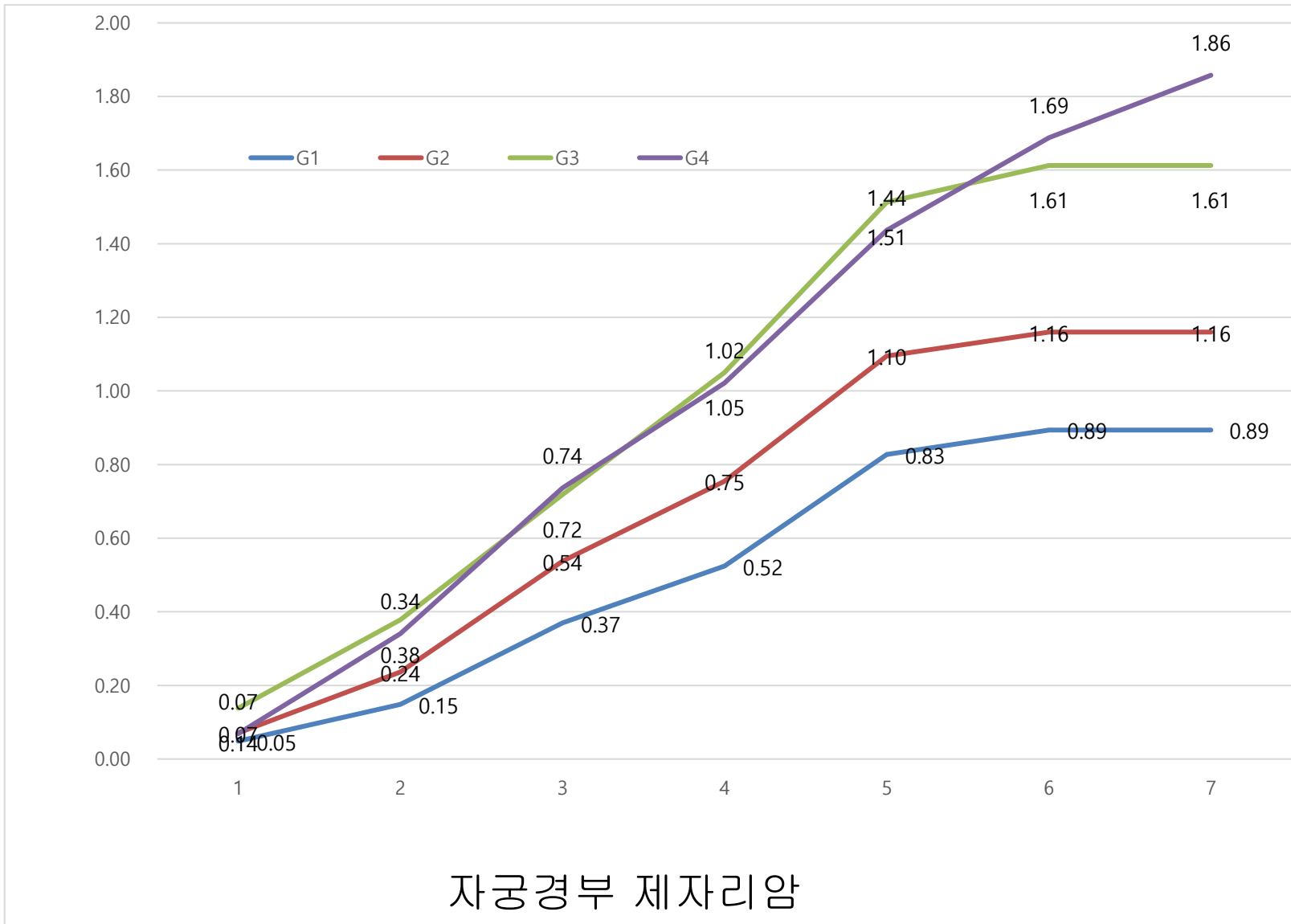
질문?

3회차 검진시점 이후 4군간 자궁암 발생율의 차이는?

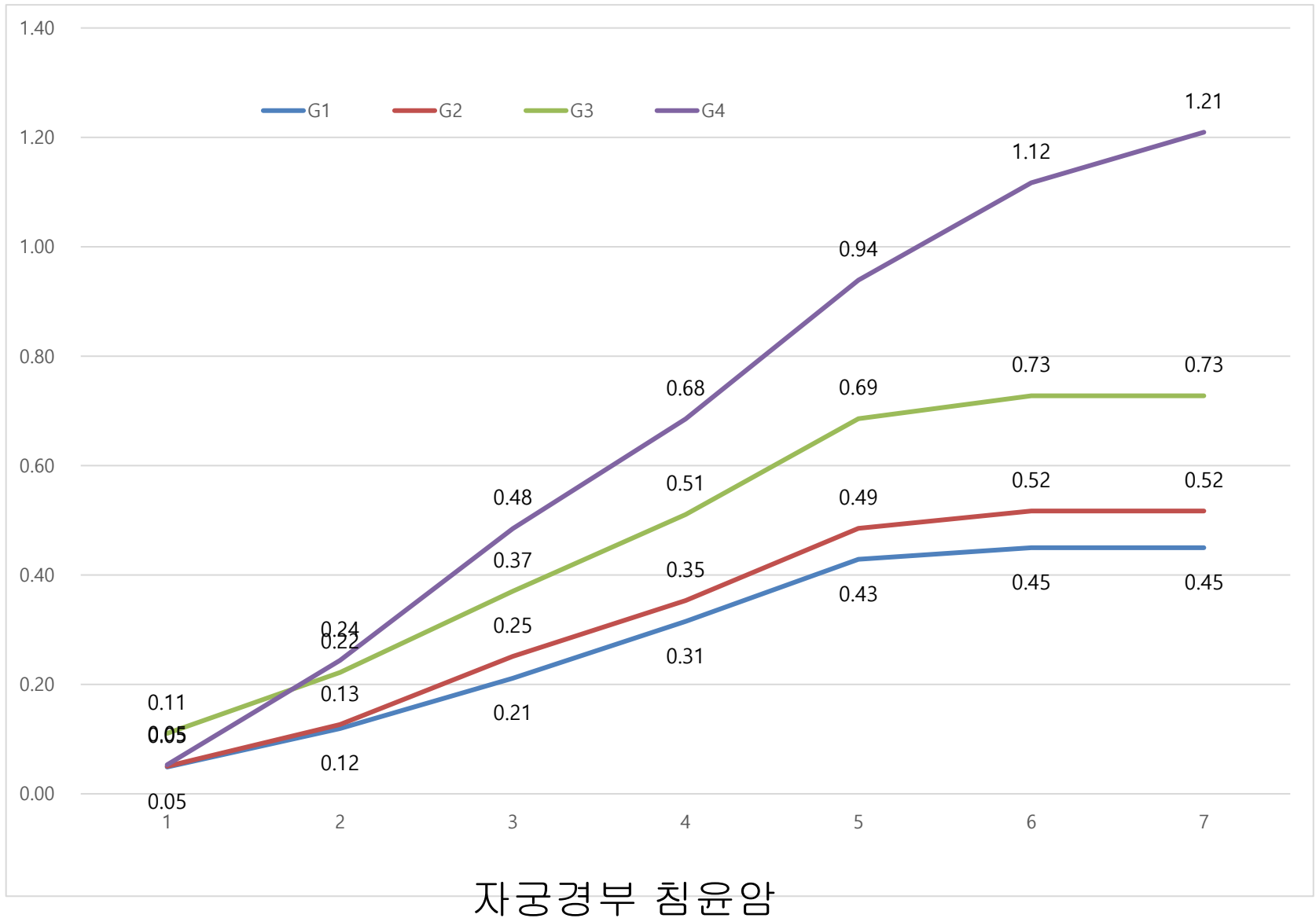
부인종양학 영역 활용사례 3_2

	O-O-O	X-O-O	X-X-O	X-X-X	계
30대	21,247	134,147	434,449	2,997,177	3,587,020
	0.59	3.74	12.11	83.56	32.52
40대	160,466	170,801	221,672	2,437,241	2,990,180
	5.37	5.71	7.41	81.51	27.11
50대	185,639	136,783	138,770	1,633,436	2,094,628
	8.86	6.53	6.63	77.98	18.99
60대	121,282	75,824	77,894	1,144,839	1,419,839
	8.54	5.34	5.49	80.63	12.87
70대	22,081	19,593	32,195	719,168	793,037
	2.78	2.47	4.06	90.69	7.19
80대	454	721	2,235	143,086	146,496
	0.31	0.49	1.53	97.67	1.33
계	511,169	537,869	907,215	9,074,947	11,031,200
	4.63	4.88	8.22	82.27	100.00

부인종양학 영역 활용사례 3_3



부인종양학 영역 활용사례 3_4



III. 결론

- **연구자 입장에서**
- **국민건강보험공단의 빅데이터를 이용하여**
- **부인종양학 분야의 연구를 수행할 수 있다.**

질의 & 응답