



Big Data Approaches for Conducting Clinical Research Using Claims Data

김희진
(서울대학교 암연구소)

DECLARATION OF INTERESTS

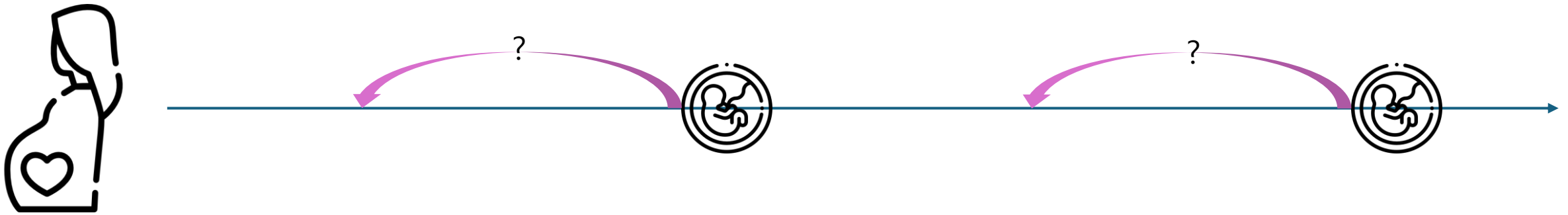
본 연구는 어떠한 이해상충관계도 없으며, 식품의약품안전처(21183MFDS433)의 지원을 받아 수행되었음을 밝힙니다.

청구자료 기반 임신부 에피소드 식별 알고리즘

청구자료:

- 임산부는 윤리적인 문제로 무작위 임상시험 수행이 어려워 약물 안전성 평가 시 관찰연구가 많이 활용됨.
- 그러나, 주요 소스인 건강보험 청구자료는 임신시작일(Last Menstrual Period, LMP)에 대한 정보가 없음.

<Gestational period>



청구자료 기반 임신부 에피소드 식별 알고리즘

Open Access

Volume: 46, Article ID: e2024012, 10 pages
<https://doi.org/10.4178/epih.e2024012>

ORIGINAL ARTICLE

epiH
Epidemiology
and Health

Identifying pregnancy episodes and estimating the last menstrual period using an administrative database in Korea: an application to patients with systemic lupus erythematosus

Yu-Seon Jung¹, Yeo-Jin Song^{2,3}, Jihyun Keum⁴, Ju Won Lee^{1,5}, Eun Jin Jang⁶, Soo-Kyung Cho^{2,3}, Yoon-Kyoung Jung^{2,3}, Sun-Young Jung^{1,5}

¹Chung-Ang University College of Pharmacy, Seoul, Korea; ²Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea; ³Hanyang University Institute for Rheumatology Research, Seoul, Korea; ⁴Department of Obstetrics and Gynecology, Hanyang University College of Medicine, Seoul, Korea; ⁵Department of Global Innovative Drugs, Graduate School of Chung-Ang University, Seoul, Korea; ⁶Department of Information Statistics, Andong National University, Andong, Korea

OBJECTIVES: This study developed an algorithm for identifying pregnancy episodes and estimating the last menstrual period (LMP) in an administrative claims database and applied it to investigate the use of pregnancy-incompatible immunosuppressants among pregnant women with systemic lupus erythematosus (SLE).

METHODS: An algorithm was developed and applied to a nationwide claims database in Korea. Pregnancy episodes were identified using a hierarchy of pregnancy outcomes and clinically plausible periods for subsequent episodes. The LMP was estimated using preterm delivery, sonography, and abortion procedure codes. Otherwise, outcome-specific estimates were applied, assigning a fixed gestational age to the corresponding pregnancy outcome. The algorithm was used to examine the prevalence of pregnancies and utilization of pregnancy-incompatible immunosuppressants (cyclophosphamide [CYC]/mycophenolate mofetil [MMF]/methotrexate [MTX]) and non-steroidal anti-inflammatory drugs (NSAIDs) during pregnancy in SLE patients.

RESULTS: The pregnancy outcomes identified in SLE patients included live births (67%), stillbirths (2%), and abortions (31%). The LMP was mostly estimated with outcome-specific estimates for full-term births (92.3%) and using sonography procedure codes (54.7%) and preterm delivery diagnosis codes (37.9%) for preterm births. The use of CYC/MMF/MTX decreased from 7.6% during preconception to 0.2% at the end of pregnancy. CYC/MMF/MTX use was observed in 3.6% of women within 3 months preconception and 2.5% during 0-7 weeks of pregnancy.

CONCLUSIONS: This study presents the first pregnancy algorithm using a Korean administrative claims database. Although further validation is necessary, this study provides a foundation for evaluating the safety of medications during pregnancy using secondary databases in Korea, especially for rare diseases.

KEY WORDS: Algorithms, Pregnancy, Pregnancy outcome, Administrative claims healthcare, Immunosuppressive agents, Lupus erythematosus systemic



청구자료 기반 임신부 에피소드 식별 알고리즘

임신 에피소드 식별 알고리즘:

- 기존에는 임신부 연구 시 임신결과발생일을 기준으로 고정된 임신기간(e.g., 280일, 245일 등)을 할당함.
- 그러나, 2016년 임신부 초음파 및 조산급여 확대로 청구자료 내 임신기간을 추정할 수 있는 정보가 많이 포함되게 되었음.
- 2023년, 2002-2018 국내 국민건강보험공단 청구자료를 활용하여 임신결과를 식별하고 초음파 기록, 진단 및 시술코드를 종합적으로 활용해 임신부의 마지막 월경일(LMP)을 추정하는 알고리즘이 개발되었음.

임신 에피소드 식별 1) 임신결과 식별

표. 청구자료 기반 임신결과 식별 조작적 정의

Pregnancy Outcomes	Definition
Stillbirth (KCD-7 code only)	Z371, Z373, Z374, Z376, Z377, O364
Delivery (Procedure code only)	R313, R3131, R3133, R3136, R3138, R314, R3141, R3143, R3146, R3148, R435, R4351, R4353, R4356, R4358, R436, R4361, R4362, R438, R4380, R450, R4507, R4508, R4509, R451, R4510, R4514, R4516, R4517, R4518, R4519, R4520, R4522, R500, R5001, R5002, RA31, RA311, RA312, RA313, RA314, RA315, RA316, RA317, RA318, RA36, RA361, RA362, RA38, RA380, RA43, RA431, RA432, RA433, RA434
Abortion (KCD-7code only)	O01, O02, O03, O04, O05, O06, O07, O08
Induced abortion (KCD-7 code or procedure code)	(KCD-7 code) O04 (Procedure code) R4452, R4456, R4457, R4458, R4459
Spontaneous abortion	All abortion episode without codes indicative of induced abortion

1. Identification of pregnancy outcomes

Ex)

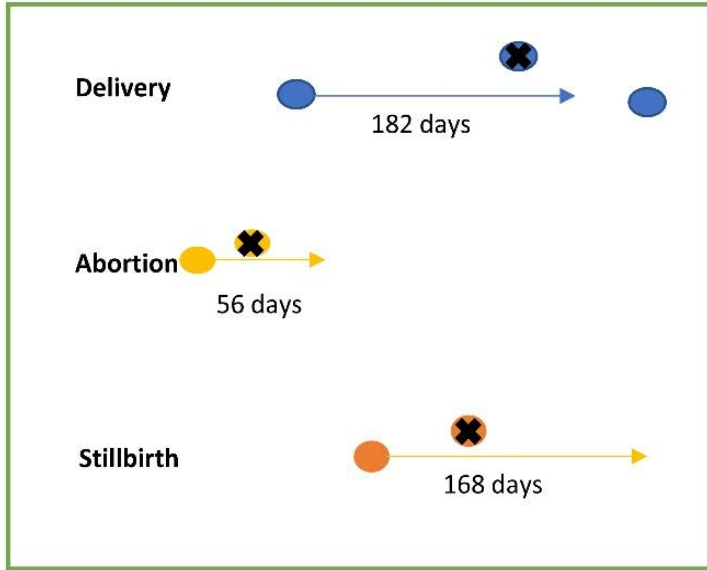


Person ID	Pregnancy outcome	Outcome date	Episode
10010	Abortion	2016-07-01	1
10010	Abortion	2016-07-30	2
10010	Delivery	2016-10-15	3
10010	Stillbirth	2016-12-09	4
10010	Stillbirth	2017-02-01	5
10010	Delivery	2017-02-15	6
10010	Delivery	2017-06-17	7

- 사산(Stillbirth): KCD-7 진단코드 활용
- 유산(Abortion): KCD-7 진단코드 활용
- 분만(Delivery): 시술코드 활용

임신 에피소드 식별 2) 동일 임신결과 간 에피소드 제거

2. Construction of pregnancy episodes in each pregnancy outcomes



Person ID	Pregnancy outcome	Outcome date	Episode	Duration within same outcomes
10010	Abortion	2016-07-01	1	.
10010	Abortion	2016-07-30	2	29
10010	Delivery	2016-10-15	2	.
10010	Stillbirth	2016-12-09	3	.
10010	Stillbirth	2017-02-01	5	54
10010	Delivery	2017-02-15	6	123
10010	Delivery	2017-06-17	4	245

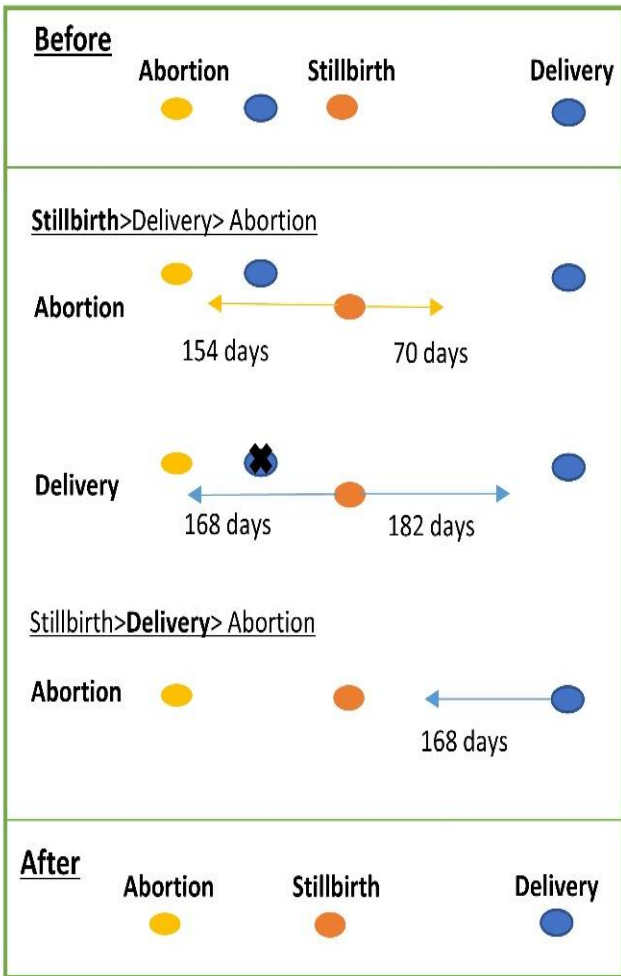
123일

245일

- 동일 임신결과가 여러 개 식별되면 첫 임신결과를 기준으로 아래 정의된 기간 내에 동일 임신결과가 발견될 경우 제외
 - 유산: 56일
 - 사산: 168일
 - 분만: 182일

임신 에피소드 식별 3) 서로 다른 임신결과 간 에피소드 제거

3. Hierarchical approach to resolve overlapping episodes of different pregnancy outcomes



Person ID	Pregnancy outcome	Outcome date	Episode	Priority
10010	Abortion	2016-07-01	1	3
10010	Delivery	2016-10-15	2	2
10010	Stillbirth	2016-12-09	2	1
10010	Delivery	2017-06-17	3	2

55일

서로 다른 임신결과의 중복을 제외

(사산 > 분만 > 유산 순서로 계층적 우선순위)

- 먼저 사산을 기준으로 아래 표에 정의된 기간 내 식별된 분만, 유산 결과 제외
- 그 다음 분만을 기준으로 유산 결과 제외

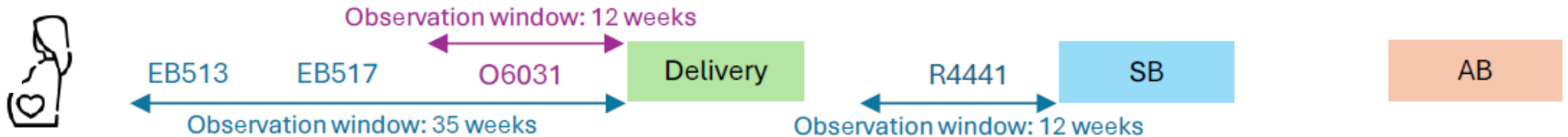
표. 후속 에피소드에 필요한 임상적으로 타당한 기간적 정의

선행 에피소드	후속 에피소드	최소기간
분만	사산	168일
분만	유산	70일
사산	분만	182일
사산	유산	70일
유산	분만	168일
유산	사산	154일

임신 에피소드 식별 4) LMP 및 임신기간(Gestational age, GA) 추정

■ 임신결과일 이전 기록된 코드 기반 LMP 추정

- ✓ 각 코드에 정의된 일수를 임신결과일 또는 초음파기록일로부터 역으로 계산하여 LMP 추정
- ✓ 초음파코드, 조산진단코드, 유산시술코드 활용

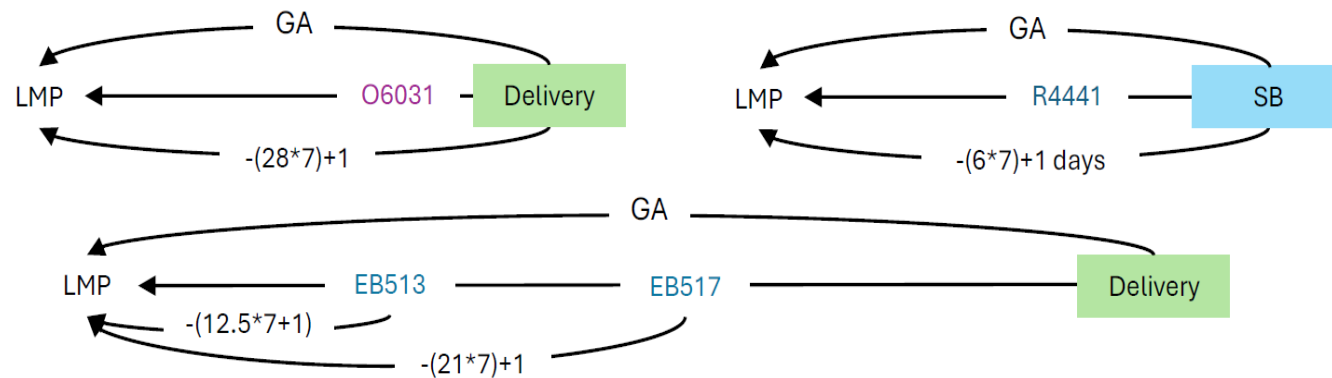


- ✓ **AB procedure code:** Subtract the midpoint of range of GA described in the code from the outcome date
- ✓ **PB code:** Subtract the midpoint of range of GA described in the code or 35 weeks from the outcome date
- ✓ **Ultrasound code:** Subtract the midpoint of trimesters described in the code from the service date

임신 에피소드 식별 4) LMP 및 임신기간(Gestational age, GA) 추정

① Hierarchy of LMP estimation method

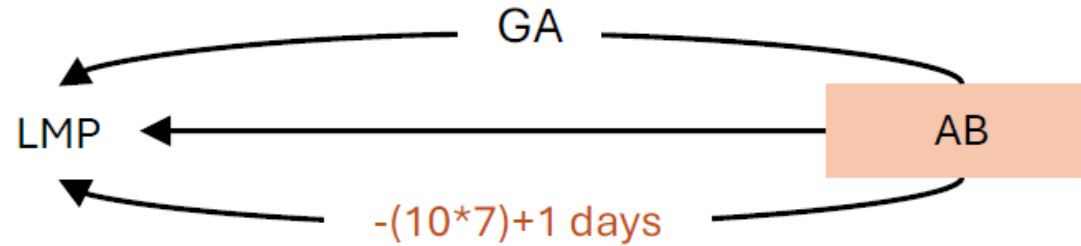
Priority	Ultrasound code	PB code	AB procedure code
First	2nd-3rd trimester TS	With specific GA	All related procedure code
Second	1st trimester TS	Without specific GA	
Third	1st trimester GS		
Fourth	2nd-3rd trimester GS		



분류	유형	계층적 우선순위	진단 및 시술코드	LMP 추정방법
초음파	1삼분기 일반초음파	3	E9471, EB511, EB512	초음파 일자-49일
	1삼분기 정밀초음파	2	E9473, EB513, EB514	초음파 일자-87.5일
	2-3삼분기 일반초음파	4	E9472, EB515, EB516	초음파 일자-196일
	2-3삼분기 정밀초음파	1	E9474, EB517, EB518	초음파 일자-147일
조산진단코드	특정GA가 있는	1	O6010, O6030	분만일자-147일
			O6011, O6031	분만일자-196일
			O6012, O6032	분만일자-248.5일
	특정GA가 없는	2	O601, O6019, O603, O6039	분만일자-245일
유산시술코드	유도유산 8주 미만	1	R4452	유사산일자-28일
	자연유산 12주 미만		R4441	유사산일자-42일
	유도유산 8-12주		R4456	유사산일자-70일
	유도유산 12-16주		R4457	유사산일자-98일
	자연유산 12주 이상		R4442	유사산일자-112일
	유도유산 16-20주		R4458	유사산일자-126일
	유도유산 20주 이상		R4459	유사산일자-141일

임신 에피소드 식별 4) LMP 및 임신기간(Gestational age, GA) 추정

- ✓ If none of the codes are available, an **outcome-specific estimation method** that assigns a fixed GA is applied.



■ 결과추정방법

- ① 유산: 70일
- ② 사산: 245일
- ③ 분만: 280일

■ 임신기간(Gestational Age, GA)

- LMP부터 임신결과일까지의 기간

임신 에피소드 식별 4) LMP 및 임신기간(Gestational age, GA) 추정

Selecting the high priority and smallest GA if multiple GA estimates are available

Episode	Code	Estimated GA	Original algorithm	
			Priority	Selected GA
Delivery	O6031	196 days	First	
	EB513	220 days	Second	
	EB517	160 days	First	✓
SB	R4441	42 days	First	✓
AB	None	70 days	None	✓

- 하나의 에피소드에 여러 개의 LMP가 추정되는 경우,
 - 더 상위계층의 코드 및 더 짧은 GA을 선택

임신 에피소드 식별 5) Retry period 정의

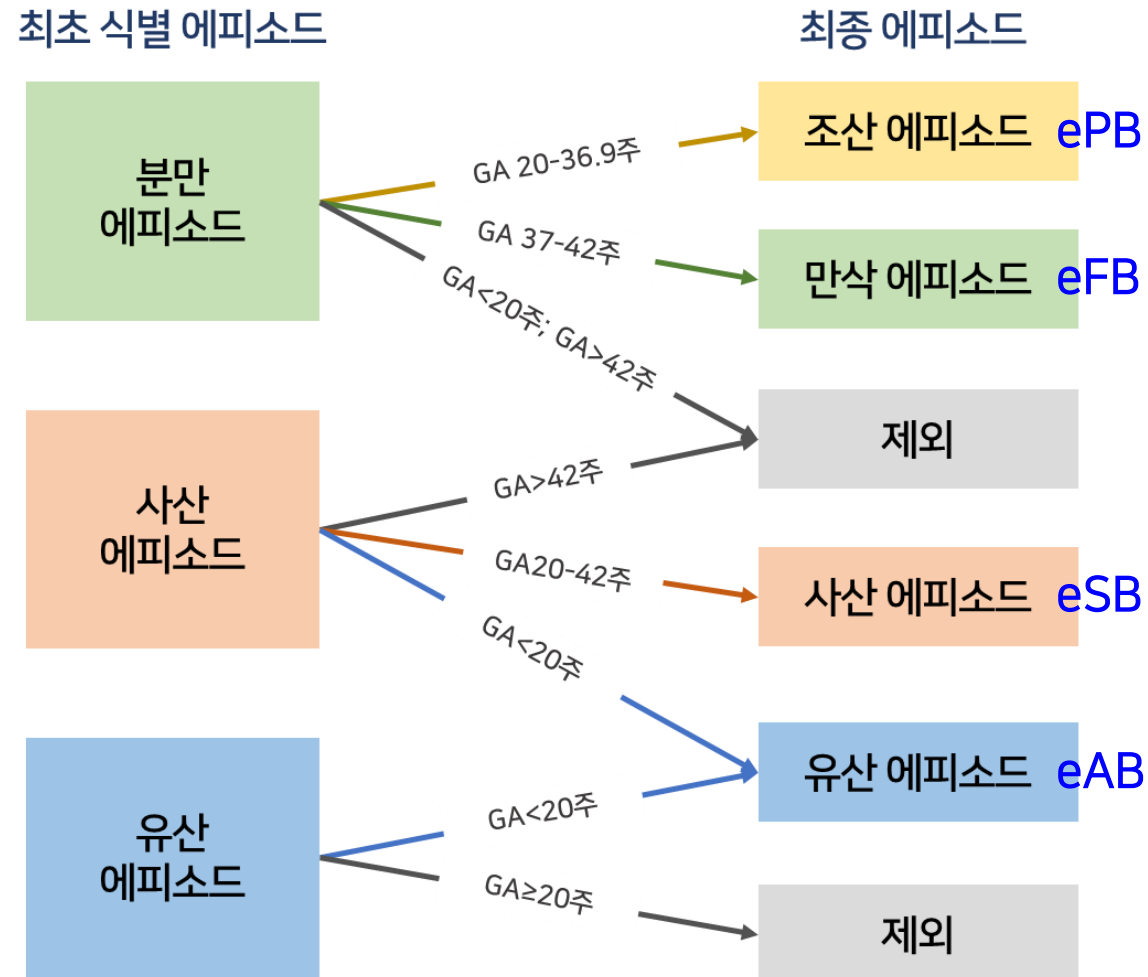
- ✓ **Retry period^{b)}**: Ensures the clinically required duration before a subsequent pregnancy episode can begin after termination



에피소드	Retry period: 후속 임신을 다시 시작하기까지 필요한 최소기간
분만	28일
사산	28일
유산	14일

임신 에피소드 식별 6) 추정된 임신기간 기반 최종 에피소드

❖ 추정된 GA 기반으로 최종 임신 에피소드의 임신결과를 재분류



알고리즘 타당도 연구: 1) 연구배경 및 필요성

- 선행 알고리즘은 조산 등의 나쁜 임신결과 유병률이 높은 것으로 알려진 전신성 홍반 루푸스(systemic lupus erythematosus, SLE) 환자 약 5,800명의 코호트를 기반으로 개발됨.
 - 일반 임산부 식별(일반화) 가능성이 추가로 평가되어야 함.
- 2016년 산전 초음파 및 조산급여 확대 전/후의 기간을 모두 포함하기 위해 2002-2018 건강보험 청구자료를 이용하여 개발됨.
 - 변화된 코딩체계 등 최신 데이터에 대한 업데이트가 필요할 수 있음.

알고리즘 타당도 연구: 1) 연구배경 및 필요성

- 초음파코드가 LMP 추정의 주요 출처임(전체 임신 에피소드의 60%-70%가 초음파기록 기반 LMP 산출).
 - 초음파 코드에 할당된 계층적 우선순위(hierarchy)가 임신기간의 추정에 상당한 영향을 미칠 수 있음.
 - 초음파 우선순위:
 - ① 임신 2-3삼분기 정밀
 - ② 1삼분기 정밀
 - ③ 1삼분기 일반
 - ④ 2-3삼분기 일반초음파

알고리즘 타당도 연구: 2) 연구목적 및 연구자료

■ 연구목적:

- ① 알고리즘의 초음파 계층적 우선순위 수정이 분만, 사산 및 유산의 LMP 및 GA 추정에 미치는 영향을 평가
- ② 청구자료에서 임신 에피소드를 식별하기 위한 최적의 알고리즘을 제시

■ 연구자료:

- 2020년 1월 1일 - 2022년 12월 31일 국민건강보험 청구자료

분류	유형	계층적 우선순위	진단 및 시술코드	LMP 추정방법
초음파	1삼분기 일반초음파	3	E9471, EB511, EB512	초음파 일자-49일
	1삼분기 정밀초음파	2	E9473, EB513, EB514	초음파 일자-87.5일
	2-3삼분기 일반초음파	4	E9472, EB515, EB516	초음파 일자-196일
	2-3삼분기 정밀초음파	1	E9474, EB517, EB518	초음파 일자-147일
조산진단코드	특정GA가 있는	1	O6010, O6030	분만일자-147일
			O6011, O6031	분만일자-196일
			O6012, O6032	분만일자-248.5일
	특정GA가 없는	2	O601, O6019, O603, O6039	분만일자-245일
유산시술코드	유도유산 8주 미만	1	R4452	유사산일자-28일
	자연유산 12주 미만		R4441	유사산일자-42일
	유도유산 8-12주		R4456	유사산일자-70일
	유도유산 12-16주		R4457	유사산일자-98일
	자연유산 12주 이상		R4442	유사산일자-112일
	유도유산 16-20주		R4458	유사산일자-126일
	유도유산 20주 이상		R4459	유사산일자-141일

알고리즘 타당도 연구: 3) 연구방법

❖ 임신결과 조작적정의 수정 및 수정된 알고리즘 기반 임신결과 식별

- 2021년 2월 27일 - 2022년 12월 31일 사이에 최소 한 번의 임신결과(분만, 사산, 유산)를 경험한 18-49세 여성을 식별
 - ✓ 임신결과 식별은 기존 알고리즘과 동일한 조작적정의를 활용(KCD-7를 KCD-8로 업데이트)
 - ✓ 유산의 경우 기존 연구보다 제한된 임신결과 적용 (기존: 001-008 -> 수정: 002-006)

표. 청구자료 기반 임신결과 식별 조작적 정의

Pregnancy Outcomes	Definition
Stillbirth (KCD-8 code only)	Z371, Z373, Z374, Z376, Z377, O364
Delivery (Procedure code only)	R313, R3131, R3133, R3136, R3138, R314, R3141, R3143, R3146, R3148, R435, R4351, R4353, R4356, R4358, R436, R4361, R4362, R438, R4380, R450, R4507, R4508, R4509, R451, R4510, R4514, R4516, R4517, R4518, R4519, R4520, R4522, R500, R5001, R5002, RA31, RA311, RA312, RA313, RA314, RA315, RA316, RA317, RA318, RA36, RA361, RA362, RA38, RA380, RA43, RA431, RA432, RA433, RA434
Abortion (KCD-8 code only)	002, 003, 004, 005, 006
Induced abortion (KCD-7 code or procedure code)	(KCD-8 code) 004 (Procedure code) R4452, R4456, R4457, R4458, R4459
Spontaneous abortion	All abortion episode without codes indicative of induced abortion

알고리즘 타당도 연구: 4) 연구방법

❖ 초음파 계층 수정

■ 초음파 계층적 우선순위 변경:

- 2016년 변경된 급여정책에 따르면 제1삼분기 정밀초음파는 급여가 임신기간 내 1번만 인정되며 인정 주수가 가장 줄음.

< 임신부 초음파 급여적용 횟수 >

행위명		인정 주수	인정 횟수
제1삼분기	일반	임신 13주 이하 - 임신여부 및 자궁 및 부속기의 종합적인 확인을 하는 경우 산정하고, 임신 여부만을 확인하는 경우 '주'항에 따라 산정	2회
	정밀	임신 11-13주	1회
제2,3삼분기	일반	임신 14-19주, 임신 20-35주, 임신 36주 이후	각1회
	정밀	임신 16주 이후	1회

- 기존 및 수정 알고리즘을 각각 적용하여 LMP 및 GA를 추정
- 추정된 GA를 기반으로 최종 에피소드(eFB, ePB, eSB, eAB)

도출

분류	유형	계층적 우선순위	진단 코드	수정된 계층적 우선순위	LMP 추정방법
초음파	1삼분기 일반초음파	3	E9471, E9473, E9474	3	초음파 일자-49일
	1삼분기 정밀초음파	2	E9473, E9474	1	초음파 일자-87.5일
	2-3삼분기 일반초음파	4	E9472, E9473, E9474	4	초음파 일자-196일
	2-3삼분기 정밀초음파	1	E9474, E9475	2	초음파 일자-147일
조산진단코드	특정GA가 있는	1	O6010, O6030		분만일자-147일
			O6011, O6031		분만일자-196일
			O6012, O6032		분만일자-248.5일
	특정GA가 없는	2	O601, O6019, O603, O6039		분만일자-245일
유산시술코드	유도유산 8주 미만	1	R4452		유사산일자-28일
	자연유산 12주 미만		R4441		유사산일자-42일
	유도유산 8-12주		R4456		유사산일자-70일
	유도유산 12-16주		R4457		유사산일자-98일
	자연유산 12주 이상		R4442		유사산일자-112일
	유도유산 16-20주		R4458		유사산일자-126일
	유도유산 20주 이상		R4459		유사산일자-141일

알고리즘 타당도 연구: 4) 연구방법

❖ 초음파 계층 수정



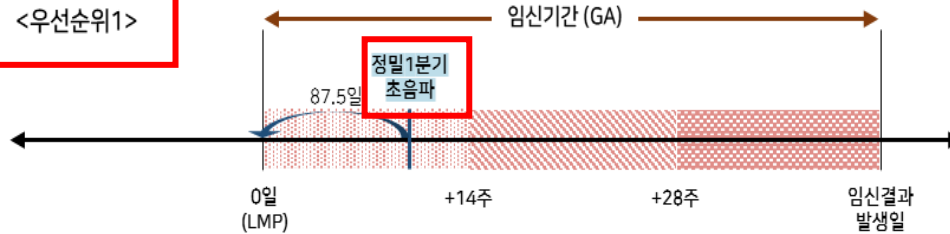
Original algorithm			
Priority	Ultrasound code	PB code	AB procedure code
First	2nd-3rd trimester TS	With specific GA	All related procedure code
Second	1st trimester TS	Without specific GA	
Third	1st trimester GS		
Fourth	2nd-3rd trimester GS		

Modified algorithm			
Priority	Ultrasound code	PB code	AB procedure code
First	1st trimester TS	With specific GA	All related procedure code
Second	2nd-3rd trimester TS	Without specific GA	
Third	1st trimester GS		
Fourth	2nd-3rd trimester GS		

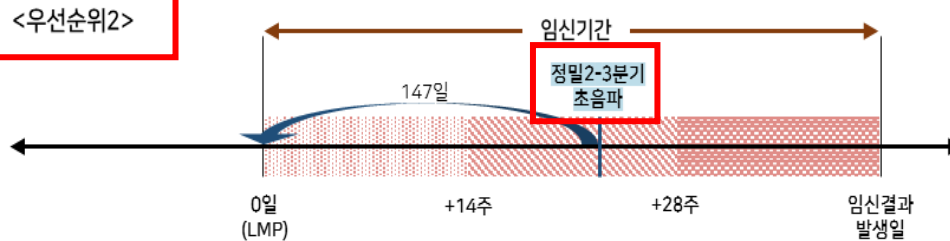
알고리즘 타당도 연구: 4) 연구방법

A) 초음파코드 기준 LMP 추정방법

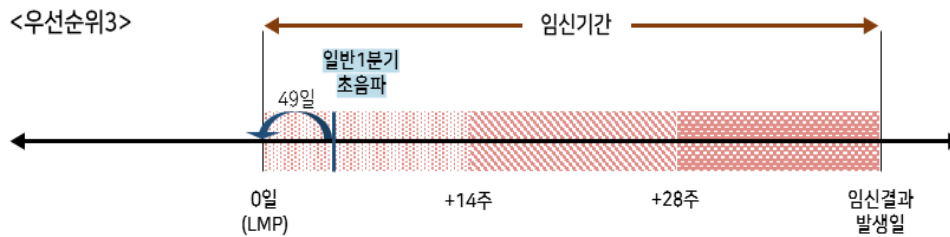
<우선순위1>



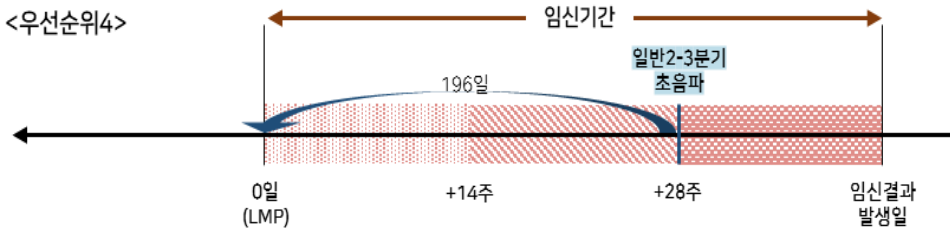
<우선순위2>



<우선순위3>

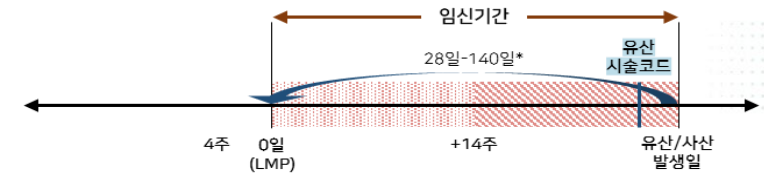


<우선순위4>



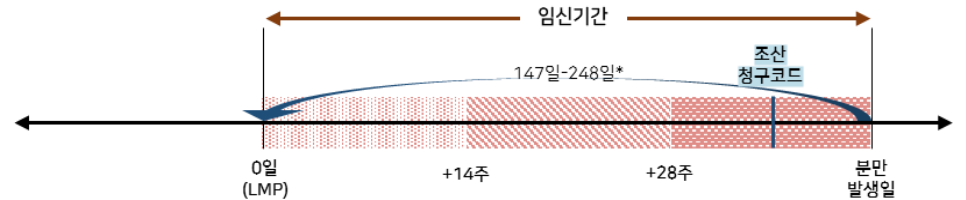
B) 유산/사산 시술코드 기준 LMP 추정방법

*코드에 따라 기간 상이

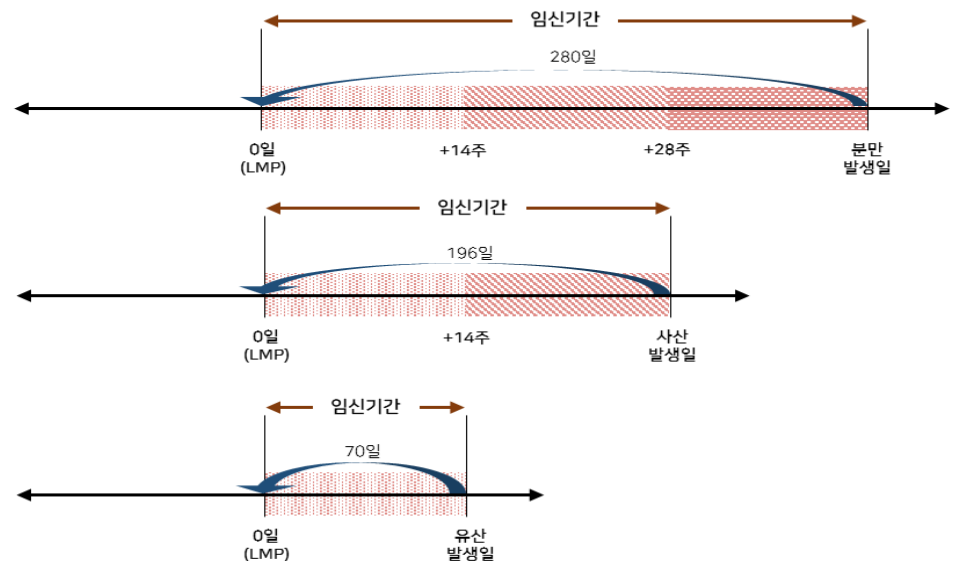


C) 조산 청구코드 기준 LMP 추정방법

*코드에 따라 기간 상이



D) 결과 기준 LMP 추정방법



알고리즘 타당도 연구: 4) 연구방법

전체 에피소드 평가

- ✓ 두 알고리즘에서 추정된 GA의 중앙값, Q1, Q3
- ✓ 추정된 GA를 기반으로 임신 에피소드를 분류했을 때, 다른 임신 결과 범주로 재분류 되거나 제외된 에피소드의 비율을 평가

조산 에피소드 평가

- **조산(ePB)**은 조산코드 유무로 결정된 것이 아닌 알고리즘에 의해 추정된 GA에 의해서만 분류되었기 때문에 [아래 3가지 접근 방식](#)을 사용
 - ① 특정 GA범위 조산코드를 포함하는 분만 에피소드에서 각 알고리즘으로 추정된 GA가 코드에 지정된 GA의 범위 내에 있는지 여부를 판단하여 일치도를 평가
 - 범위 내면 'agreement', 범위 ± 7 일이면 'near agreement', 그렇지 않으면 'disagreement'로 분류
 - ② 두 알고리즘에 의해 추정된 출생아 수 및 조산율을 통계청(2021-2022년)의 국가 출산통계 및 조산율과 비교
 - ③ 네가지 초음파코드의 계층적 우선순위의 24가지 조합(4!)에 대한 민감도 분석 추가 수행

알고리즘 타당도 연구: 5) 연구결과

❖ 타당도 평가 방법 (전체 에피소드)

- 두 알고리즘에서 추정된 임신 에피소드 GA의 중앙값, 제1사분위수(Q1), 제3사분위수(Q3)를 요약
- 추정된 GA를 기반으로 임신 에피소드를 분류했을 때, **다른 임신 결과 범주로 재분류되거나 제외된 에피소드의 비율을 평가**

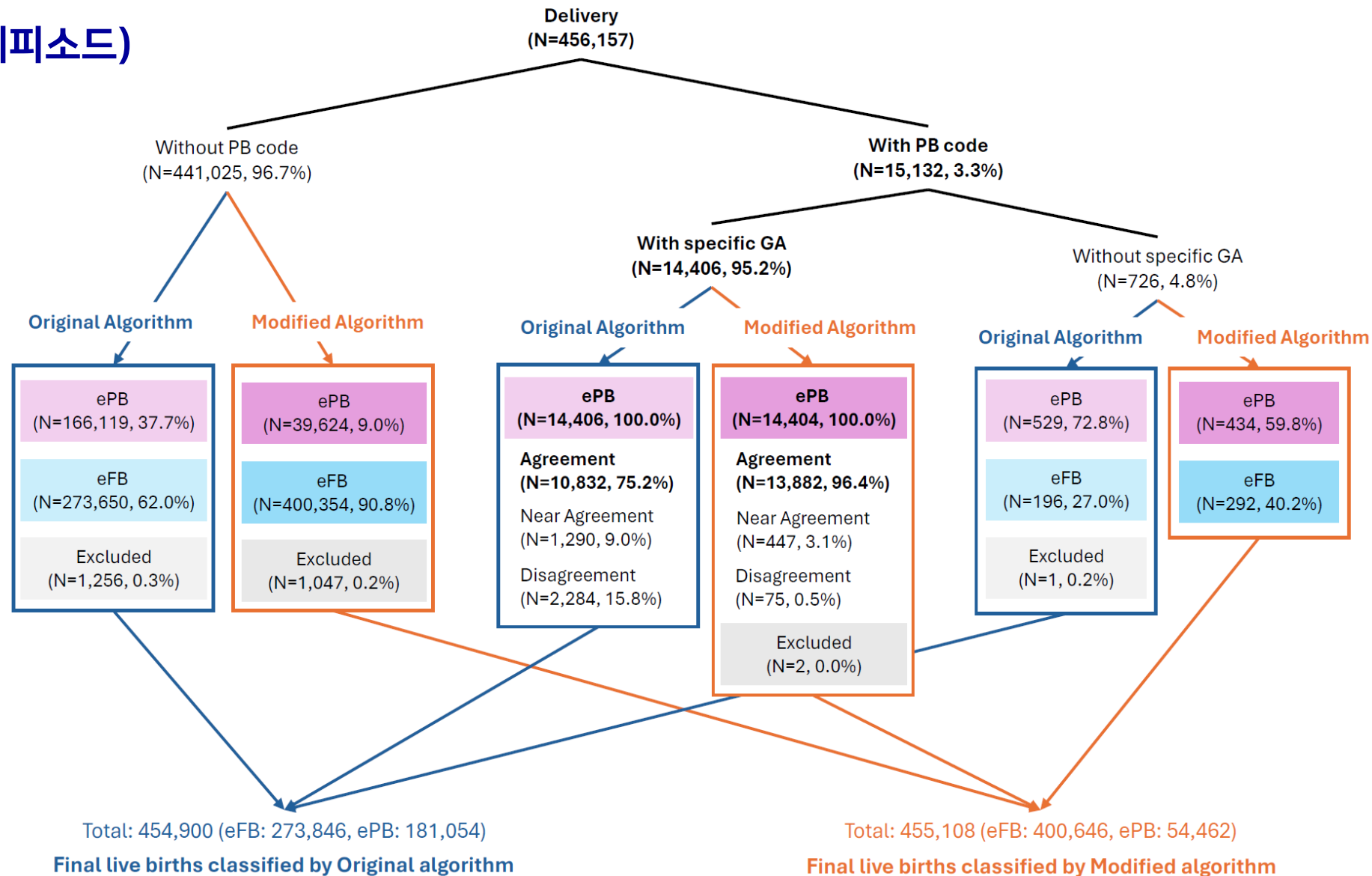
- a) 임신 에피소드에서 제외되어야 할 에피소드(회색).
- b) 조산(ePB)으로 추가 분류되어야 할 에피소드(노란색).
- c) 유산(eAB)으로 추가 분류되어야 할 에피소드(주황색).
- d) 잘 분류된 에피소드(파란색).

Pregnancy outcome	Number of identified pregnancy outcomes	Algorithm	Estimated GA				
			Median, week (Q1, Q3)	< 20week N (%)	20-37 week N (%)	37-42 week N (%)	> 42 week N (%)
Delivery	456,157	Original	37.6 (35.7, 38.9)	39 (0.0) ^{a)}	181,054 (39.7) ^{b)}	273,846 (60.0) ^{d)}	1,218 (0.3) ^{a)}
		Modified	38.8 (37.9, 39.6)	50 (0.0) ^{a)}	54,462 (11.9) ^{b)}	400,646 (87.8) ^{d)}	999 (0.2) ^{a)}
SB	3,050	Original	21.6 (17.9, 27.9)	1,175 (38.5) ^{c)}		1,867 (61.2) ^{d)}	8 (0.3) ^{a)}
		Modified	21.6 (17.1, 27.9)	1,241 (40.7) ^{c)}		1,803 (59.1) ^{d)}	6 (0.2) ^{a)}
AB	122,533	Original	6.9 (5.9, 7.7)	122,297 (99.8) ^{d)}		236 (0.2) ^{a)}	-
		Modified	6.9 (5.9, 7.7)	122,320 (99.8) ^{d)}		213 (0.2) ^{a)}	-

알고리즘 타당도 연구: 5) 연구결과

❖ 타당도 평가 방법 (조산 에피소드)

- ① 특정 GA 범위 조산코드를 포함하는 분만 에피소드의 해당 GA범위와 알고리즘에 의해 추정된 GA와의 일치도 평가



알고리즘 타당도 연구: 5) 연구결과

❖ 타당도 평가 방법 (조산 에피소드)

② 두 알고리즘에 의해 추정된 조산율을 통계청(KOSIS) 2021-2022년 국가 출산통계와 비교

Index	Birth identification period	Number of pregnancy episodes					ePB rate(%) ^{c)}	eSB rate(‰) ^{d)}	eAB rate(%) ^{e)}
		Total pregnancy ^{a)}	Live birth ^{b)}	ePB	eSB	eAB			
Original	Feb 27, 2021	580,239	454,900	181,054	1,867	123,472	39.8	4.1	21.3
Modified	- Dec 31, 2022	580,472	455,108	54,462	1,803	123,561	12.0	3.9	21.3
KOSIS	2021-2022	-	509,748	48,033	-	-	9.4	-	-

Abbreviation: KOSIS Korean Statistical Information Service. ePB estimated pre-term birth. eSB estimated stillbirth. eAB estimated abortive outcome.

a) The sum of live birth, stillbirth and abortion episodes.

b) The sum of preterm birth episodes and full-term birth episodes.

c) Preterm birth episodes per 100 live birth episode.

d) Stillbirth episodes per 1000 live birth and stillbirth.

e) Abortion episodes per 100 pregnancies.

알고리즘 타당도 연구: 5) 연구결과

❖ 타당도 평가 방법 (조산 에피소드)

③ 네가지 초음파코드의 계층적 우선순위 조합

24가지(4!)의 조산율에 대한 민감도 분석

Modified

Ultrasound code priority				ePB	eFB	ePB rate
1st GS	1st TS	2nd-3rd GS	2nd-3rd TS	N	N	
1st	2nd	3rd	4th	243,766	211,294	53.6
1st	2nd	4th	3rd	242,224	212,862	53.2
1st	3rd	2nd	4th	245,269	209,770	53.9
1st	3rd	4th	2nd	242,713	212,371	53.3
1st	4th	2nd	3rd	245,271	209,768	53.9
1st	4th	3rd	2nd	242,777	212,300	53.3
2nd	1st	3rd	4th	57,440	397,655	12.6
2nd	1st	4th	3rd	55,898	399,223	12.3
3rd	1st	2nd	4th	59,910	395,094	13.2
3rd	1st	4th	2nd	54,462	400,646	12.0
4th	1st	2nd	3rd	59,914	395,093	13.2
4th	1st	3rd	2nd	54,772	400,296	12.0
2nd	3rd	1st	4th	445,853	7,170	98.4
2nd	4th	1st	3rd	445,855	7,168	98.4
3rd	2nd	1st	4th	445,828	7,197	98.4
3rd	4th	1st	2nd	445,873	7,154	98.4
4th	2nd	1st	3rd	445,832	7,196	98.4
4th	3rd	1st	2nd	445,865	7,162	98.4
2nd	3rd	4th	1st	182,274	272,599	40.1
2nd	4th	3rd	1st	182,338	272,528	40.1
3rd	2nd	4th	1st	181,054	273,846	39.8
3rd	4th	2nd	1st	184,572	270,138	40.6
4th	2nd	3rd	1st	181,023	273,835	39.8
4th	3rd	2nd	1st	184,564	270,146	40.6

Original

논의

- 두 알고리즘은 2021년 2월 27일 - 2022년 12월 31일 454,900명(기존 알고리즘), 455,108명(수정 알고리즘)의 생존 출생아를 식별했으며 2개월의 기간 차이를 감안할 때, KOSIS에서 2021-2022년 보고된 509,748명과 매우 유사하였음.
- 그러나 기존 알고리즘으로 추정된 최종 조산율(ePB rate)은 39.8%로, KOSIS에서 2021-2022년 보고한 국가 평균 9.4%보다 다소 높아 GA가 과소 추정되었을 가능성을 시사함.
- 수정된 알고리즘은 12.0%로 비교적 낮은 조산율을 추정하였으며, KOSIS의 국가 통계보다 여전히 약간 높았지만, 미국(10.4%)과 전 세계(4-16%)에서 보고된 PB율과 비교했을 때 타당한 범위 내에 있었음.
- 민감도 분석 결과, 다른 초음파 유형보다 임신 1삼분기 정밀초음파를 우선시할 때 GA 과소추정이 최소화되었음.
- 두 알고리즘의 조산율의 차이는 GA 추정을 위한 초음파코드 계층 구조 차이에서 비롯된 것으로 보임.

논의

- 특정 GA범위 PB코드가 있는 분만 에피소드의 GA를 추정할 때, 수정 알고리즘으로 추정한 GA는 기존 알고리즘보다 특정 GA와 더 높은 일치도를 보임(agreement rate: 96.4% vs 75.2%).
- 이는 수정된 알고리즘의 GA 추정 정확도가 향상되었음을 시사함.
- 특정 GA범위가 없는 PB코드(예: KCD-8 코드: 0601, 0603, 06019, 06039)를 가진 에피소드는 59.8%~72.8%만 최종 ePB로 분류되었고, 나머지는 eFB로 분류되었음.
- 이는 특정 GA가 없는 PB코드는 조산을 항상 반영하는 것은 아니라는 점을 시사하며 임신부 연구 시 조산을 정의할 때 이러한 코드를 포함하는 것을 주의해야 함.

논의

❖ 강점:

- ① 한국 청구자료에서의 적용을 위해 특별히 개발된 임신 에피소드 식별 알고리즘을 일반 임신부를 대상으로 성능을 평가한 최초의 연구임.
- ② 제1삼분기 정밀초음파를 우선시하는 수정 알고리즘은 기존 보다 조산 에피소드 식별에서 향상된 성능을 보임.

❖ 제한점:

- ① 임상 기록이나 연계된 데이터 소스를 통한 검증이 아닌, KOSIS에서 보고한 집계 통계를 외부 벤치마크로 사용
 - 또한 KOSIS에서 사산율, 유산율이 공식 보고되지 않아 이에 대한 알고리즘 성능은 비교할 수 없었음.
 - 향후 대표병원의 전자의무기록과 같은 자료를 참조로 사용하여 GA 추정치를 검증하는 것을 목표로 해야 함.
- ② 임신결과 식별 시 자궁 외 임신 및 포상기태(KCD-8 코드: 000 및 001)와 같은 특정 유형의 유산을 제외
 - 유산율 및 일반인구의 모든 임신을 포착하는 것을 목표로 하는 연구에서 유산의 실제 부담이 과소평가될 수 있음.

논의

❖ 제한점:

③ 다태임신(multiple pregnancies)을 구별하지 않음.

- 두 임신결과가 동일하거나 근접한 날짜에 발생한 경우, 더 상위계층의 임신결과 및 동일계층인 경우 더 선행된 결과를 유지하였음.
- 예를 들어, 쌍둥이 중 한 태아가 사산 또는 조산으로, 다른 태아가 만삭분만으로 태어난 경우, 알고리즘에 의해 사산(정상분만보다 높은 계층적 우선순위를 가짐) 또는 조산(정상분만보다 더 선행된 결과)이 유지됨.
- 이러한 접근방식이 사산 및 조산에 비해 정상분만의 과소평가에 기여했을 가능성이 있음.

④ 유산 및 사산 에피소드의 오분류 가능성

- 여러 선행연구에 따르면 청구자료에서 사산은 동일/근접 날짜에 유산코드를 함께 가지고 있는 경우가 많음.
- 본 알고리즘은 사산이 유산보다 계층적 우선순위가 높기 때문에 사산날짜와 동일 또는 근접 날짜에 유산코드가 존재하는 경우에도 모두 사산으로 분류되도록 하여 오분류 발생 가능성이 있음.
- 본 알고리즘을 적용하여 사산 및 유산 에피소드를 식별할 때는 이러한 한계점을 신중하게 고려해야 함.

결론

- 그럼에도 불구하고 이러한 한계는 두 알고리즘 모두에 적용되므로, 수정 알고리즘이 조산 에피소드를 식별하는 데 더 효과적이라는 결론에 영향을 미치지 않는 것임.
- 청구자료에서 제 1삼분기 정밀초음파(초음파날짜 기준 -12.5주)는 임신부의 임신시작일을 가장 타당하게 추적할 수 있음.
- 본 연구의 결과는 한국 청구 데이터를 사용하는 향후 연구에 귀중한 방법론적 통찰력을 제공하지만, 신뢰도를 강화하기 위해 국내 대표적인 병원들의 임상기록을 통한 추가 검증이 필요함.

2026년 대한부인종양학회 제7회 동계학술대회 with Chemo-TIP Review

일자 2026년 1월 17일 (토)

장소 세종대학교 컨벤션센터

Thank you for your attention!



대한부인종양학회
Korean Society of Gynecologic Oncology

