Combination of Family History Disease and Person Concepts Enhances the Coverage of Observational Medical Outcome Partnership Common Data Model Mapping in Family History Information

Sunyoung Yoon B.S*1,2, Won Chul Cha M.D., Ph.D*1,2,3, Mira Kang M.D., Ph.D*1,4

*1) Department of Digital Health, Samsung Advanced Institute for Health Science & Technology (SAIHST), sungkyunkwan University, Seoul

*2) Smart Health Lab, Department of Digital Health, Samsung Advanced Institute for Health Science & Technology (SAIHST), sungkyunkwan University, Seoul

*3) Department of Emergency Medicine Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul

*4) Center for Health Promotion, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul

Smart Health Lab
Department of Digital Health

Smart Health Lab
Department of Digital Health

Introduction

- **Family History (FH)**

    Clinical care : early intervention or prevention of disease

    Precision medicine : gateway for decision for genetic test & interpretation

    minimal relative range for genetic test interpretation : first-degree & second-degree relative

- **OMOP-CDM**

    Observational Medical Outcome Partnership Common Data Model

    standardized clinical data + extension genetic data + FH

    great based for multicenter prediction or Machine learning research project

- **Conventional FH mapping method in OMOP CDM**

    CDM table   : Observation table

    vocabulary mapping : ATHENA (https://athena.ohdsi.org/)

Priority		e	Example
No	contents	Concept_id	NAME
1st	FH of disease in person	4215667	Family history of diabetes mellitus in first degree relative
2nd	FH of disease	4051114	Family history of diabetes mellitus

Table 1. conventional mapping process example

Introduction

- ATHENA's FH related code search result

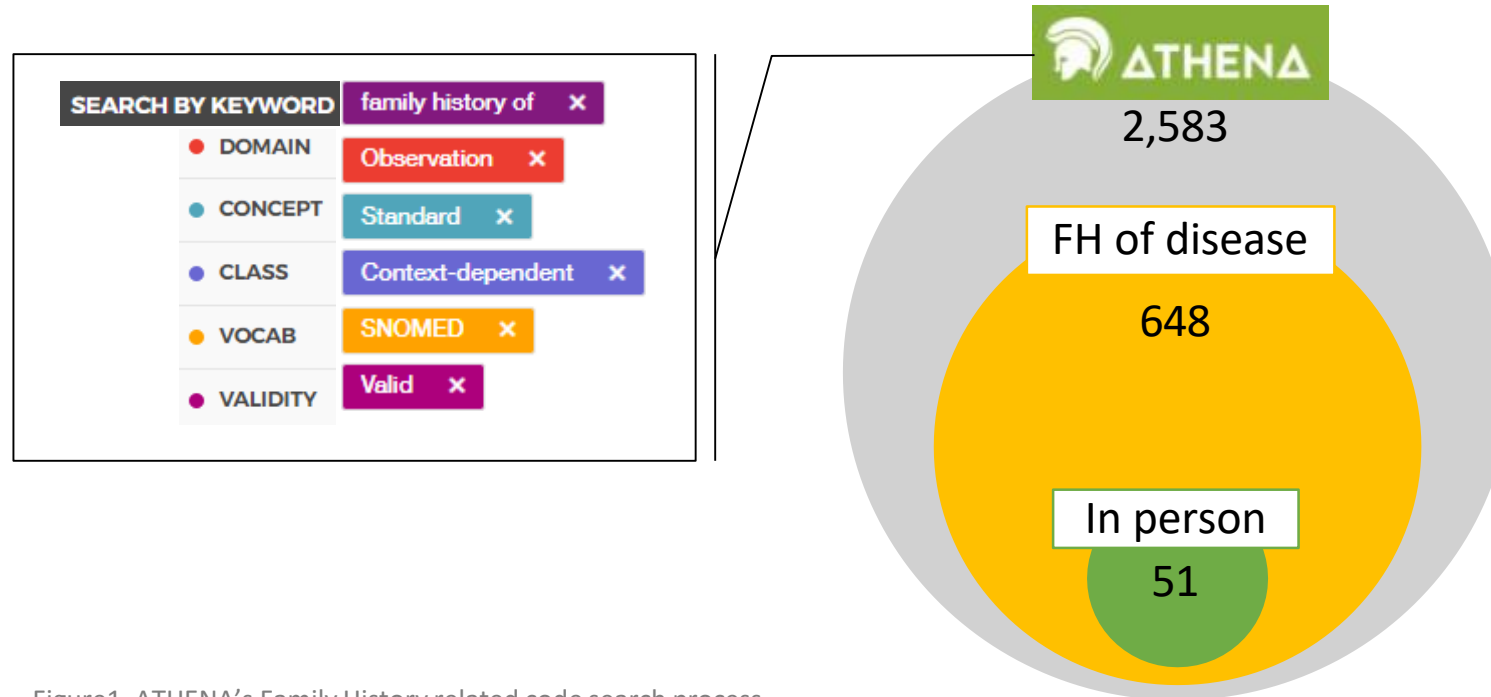 Only 51 code have both 'FH of disease' and 'in person'

Figure1. ATHENA's Family History related code search process

Introduction

- Previous research

  All of us data's Family History

  Convert FH of disease and person to observation_concept_id separately

  disadvantage : multiple FH of disease & person data ➜ difficult in accurately connecting disease and person

- Aim of this study

  Comparison of Conventional mapping method and New mapping method in single center medical check-up's
     family history survey data conversion.

  new method   : FH of disease ➜ observation_concept_id
                       in person   ➜ qualifier_concept_id

CDM Field	User Guide	ETL Conventions
observation_concept_id	The OBSERVATION_CONCEPT_ID field is recommended for primary use in analyses, and must be used for network studies.	The CONCEPT_ID that the OBSERVATION_SOURCE_CONCEPT_ID maps to. There is no specified domain that the Concepts in this table must adhere to. The only rule is that records with Concepts in the Condition, Procedure, Drug, Measurement, or Device domains MUST go to the corresponding table.
qualifier_concept_id	This field contains all attributes specifying the clinical fact further, such as as degrees, severities, drug-drug interaction alerts etc.	Use your best judgement as to what Concepts to use here and if they are necessary to accurately represent the clinical record. There is no restriction on the domain of these Concepts, they just need to be Standard.

Figure2. OMOP CDM observation table specification

Methods

- Single center Family history survey data

  Tertiary general hospital's medical check-up Family history survey data (2001.01.02 ~ 2016.07.01)

  191,619 patient – 414,616 visit – 65,538,457 data

- Total 436 category

Family history category		
Disease	Cancer	Stomach, Breast, Colorectal, Lung, Uterine, Liver, Thyroid, Ovarian, Cervical, Gallbladder / Biliary tract, Bladder, Esophageal, prostate, pancreas, others
	Chronic	Hypertension, Diabetes mellitus, Chronic hepatitis/Liver cirrhosis, Asthma / Chronic obstructive pulmonary disease (COPD)
	Congenital	Congenial Heart disease, Cleft lip or palate, other congenital malformation
	Cardiovascular	Stroke, Myocardial Infarction (MI), Angina, MI / Angina
	Other	Dementia, Tuberculosis
Answer	Yes or No	
Person	First degree	Father, Mother, Sibling, Child
	Second degree	Paternal grandfather, Paternal grandmother, Maternal grandfather, Maternal grandmother

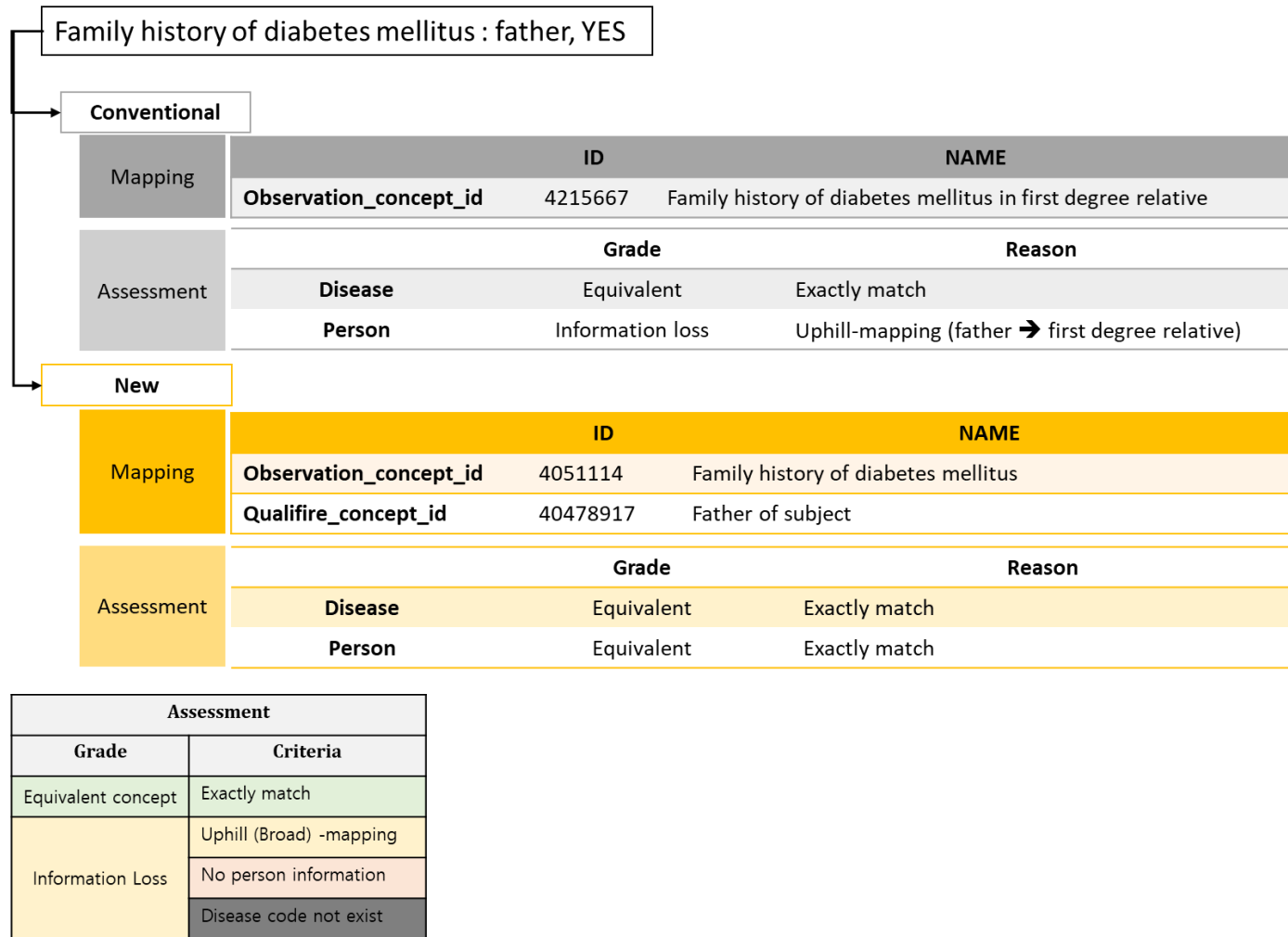Figure3. Family history survey category

- Mapping

  vocabulary : SNOMED CT

- New methods in person mapping

ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
40478917	444295003	Father of subject	Social Context	Standard	Valid	Relationship	SNOMED
40478925	444301002	Mother of subject	Social Context	Standard	Valid	Relationship	SNOMED
40478926	444302009	Sibling of subject	Social Context	Standard	Valid	Relationship	SNOMED
40485452	444192005	Child of subject	Social Context	Standard	Valid	Relationship	SNOMED
40478914	444292000	Paternal grandfather of subject	Social Context	Standard	Valid	Relationship	SNOMED
40478915	444293005	Paternal grandmother of subject	Social Context	Standard	Valid	Relationship	SNOMED
40485508	444243006	Maternal grandfather of subject	Social Context	Standard	Valid	Relationship	SNOMED
40485509	444244000	Maternal grandmother of subject	Social Context	Standard	Valid	Relationship	SNOMED

Figure4. List of 8 codes used by the new method to map 'in person'

Smart Health Lab
Department of Digital Health

Methods

Family history of diabetes mellitus : father, YES

Conventional

Mapping	ID	NAME
Observation_concept_id	4215667	Family history of diabetes mellitus in first degree relative

Assessment		Grade	Reason
	Disease	Equivalent	Exactly match
	Person	Information loss	Uphill-mapping (father ➜ first degree relative)

New

Mapping	ID	NAME
Observation_concept_id	4051114	Family history of diabetes mellitus
Qualifire_concept_id	40478917	Father of subject

Assessment		Grade	Reason
	Disease	Equivalent	Exactly match
	Person	Equivalent	Exactly match

Assessment	
Grade	**Criteria**
Equivalent concept	Exactly match
Information Loss	Uphill (Broad) -mapping
	No person information
	Disease code not exist

Figure 5. Two method's mapping & assessment process example

Compare the 2 method
Assessment grade count (disease & person)
the number of observation_concept_id used
the number of expressible data count

Smart Health Lab
Department of Digital Health

Results

- **310** category is mapped from total 436 category

  126 category that could not mapped (28.9% of total category, 51.3% of data count)

  Reason : Not exist the 'No family history of disease' code

- The number of observation_concept_id used

  conventional mapping : 49 concept_id

  new mapping : 41 concept_id (49 concept_id : when 8 'in person' concept_id included)

- Disease information assessment grade comparisons

  The two methods showed the **same** results.

  Information Loss was 18.7% of total category

  - 1 survey question asked 2 related disease

  - 1 survey question asked 1 disease

    but the exactly match disease concept was not found

Disease Assessment category count			
Assessment		Method	
Grade	Criteria	Conventional	New
Equivalent concept	Exactly match	238	238
Information Loss	Uphill (Broad) -mapping	72	72
	No person information	0	0
	Disease code not exist	126	126
Sum		436	436

Figure 6. Disease information assessment grade result of two methods

Smart Health Lab
Department of Digital Health

Results

- **Person** information assessment grade comparisons

Person Assessment category count					Person Assessment data coverage count			
Assessment		Method			Assessment		Method	
Grade	Criteria	Conventional	New		Grade	Criteria	Conventional	New
Equivalent concept	Exactly match	0	310		Equivalent concept	Exactly match	0 (0 %)	31,902,648 (100%)
Information Loss	Uphill (Broad) -mapping	21	0		Information Loss	Uphill (Broad) -mapping	140,923 (0.4%)	0 (0 %)
	No person information	289	0			No person information	31,761,725 (99.6%)	0 (0 %)
Sum		310	310		Sum		31,902,648 (100 %)	31,902,648 (100 %)

Figure 7. person information assessment grade result of two methods

- Code not exist was excluded to compare
- New method can make it possible to 100% equivalent mapping
- But conventional method resulted in all 'in person' information loss
  only 0.4% data was possible to uphill / Broad mapping
  99.6% information loss in conventional method

Smart Health Lab
Department of Digital Health

Conclusions

- New FH mapping method is possible to **minimize** 'in person' information loss ➜ **more accurate**
  - New method showed 0 % of 'in person' information loss
  - Conventional method result in 100% information loss and show only 0.4% of uphill(broad) mapping is possible

- Unbalance of 'in person' and disease in single code expression

survey example	Father	mother	children	Sibling	paternal grandfather	paternal grandmother	maternal grandfather	maternal grandmother
family history of breast cancer; YES								

ID	NAME
4210263	Family history of neoplasm of breast
46270155	Family history of malignant neoplasm of breast diagnosed before 45 years of age
46270130	Family history of malignant neoplasm of breast in first degree relative less than 50 years of age
42535025	Family history of malignant neoplasm of breast at under age 50 in second degree relative
35624517	Family history of malignant neoplasm of breast at under age 50 in second degree female relative
4051257	FH: Father alive with problem
4051704	FH: Mother alive with problem
4051259	FH: Sister alive with problem
4052797	FH: Son alive with problem
4054435	FH: Daughter alive with problem

Exact disease
But
Too specific 'in person' criteria

Exact 'in person'
But not specific disease

Figure 8. Example of information unbalance code

Smart Health Lab
Department of Digital Health

10

Conclusions

- Conventional method is complex & Labor intensive
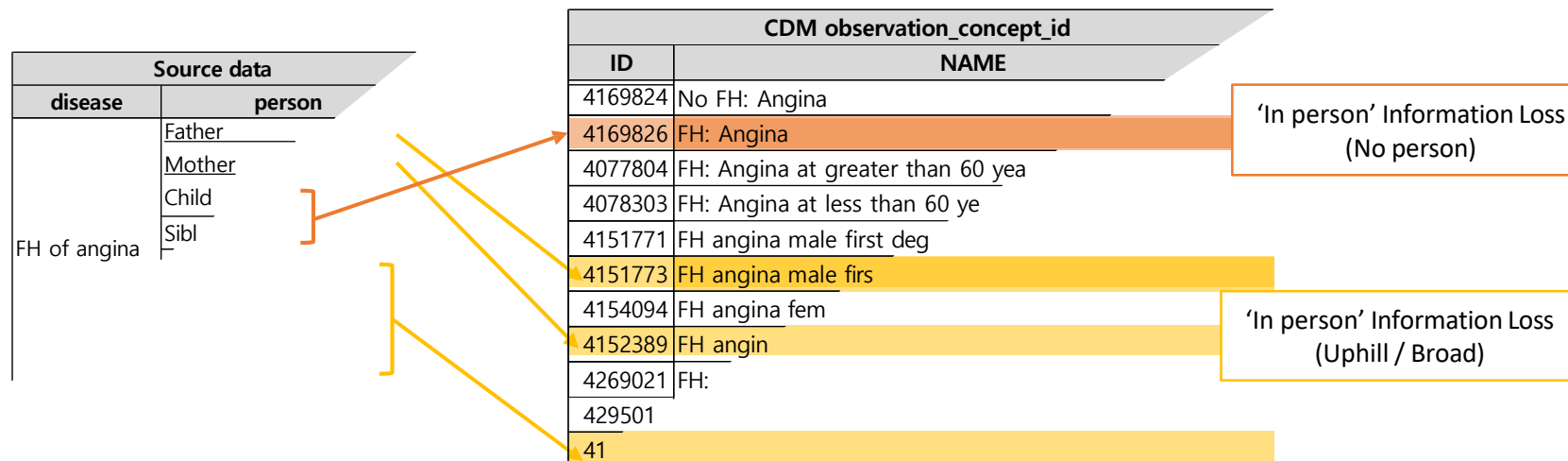 - Conventional method is need to multiple search process for proper concept_id

Figure 9. Example of conventional mapping process

Conclusions

- New FH mapping method is **simpler** process ➔ **more efficient**
  - The new method once mapped ' in person' concept_id from the source data,
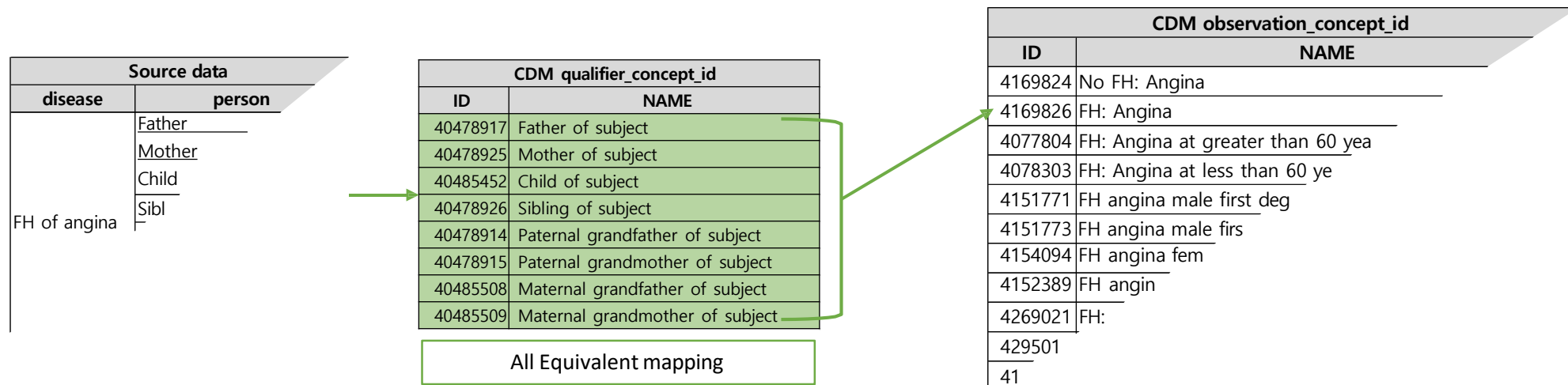   then only a single search process for the proper disease concept_id

Source data	
disease	**person**
	Father
	Mother
	Child
	Sibl
FH of angina	

CDM qualifier_concept_id	
ID	**NAME**
40478917	Father of subject
40478925	Mother of subject
40485452	Child of subject
40478926	Sibling of subject
40478914	Paternal grandfather of subject
40478915	Paternal grandmother of subject
40485508	Maternal grandfather of subject
40485509	Maternal grandmother of subject

All Equivalent mapping

CDM observation_concept_id	
ID	**NAME**
4169824	No FH: Angina
4169826	FH: Angina
4077804	FH: Angina at greater than 60 yea
4078303	FH: Angina at less than 60 ye
4151771	FH angina male first deg
4151773	FH angina male firs
4154094	FH angina fem
4152389	FH angin
4269021	FH:
429501	
41	

Figure 10. Example of new mapping process

Smart Health Lab
Department of Digital Health

Conclusions

- New method is possible to **more detailed** relative information expression
  - if source data is possible

  - 194 concept_id family relative information in ATHENA's SNOMED CT vocabulary
  - 34 concept_id representing first & seoncd-degree relative information

- Source data's broad questions are also expressible

ID	NAME
40571962	Adopted child
40571964	Adopted daughter
40567976	Adopted son
4077009	Adoptive brother
4264648	Adoptive father
4225414	Adoptive grandfather
4136361	Adoptive grandmother
4331173	Adoptive grandparent
4068928	Adoptive mother
4224702	Adoptive parent
4196190	Adoptive sibling
4294846	Adoptive sister

ID	NAME
44783070	Second degree blood relative
4050684	Half-sibling
4187654	Half-brother
36714504	Maternal half brother
36714501	Paternal half brother
4331453	Half-sister
36714503	Maternal half sister
36717756	Paternal half sister
4301632	Natural grandparent
4265919	Natural grandfather
4077002	Natural grandmother

ID	NAME
4053607	First degree blood relative
4326600	Natural child
308126	Natural daughter
4014096	Natural son
4029630	Natural parent
4321888	Natural father
4277283	Natural mother
4218412	Natural sibling
4263682	Natural brother
4217930	Twin brother
4218588	Fraternal twin brother
4299961	Identical twin brother
4251326	Natural sister
4000174	Twin sister
4106384	Fraternal twin sister
4173558	Identical twin sister
4013484	Twin sibling
4196733	Fraternal twin
4010423	Fraternal twin sibling
4193698	Identical twin
4031588	Identical twin sibling
4217930	Twin brother
4000174	Twin sister

Figure11 . Example code of family relative information

Conclusions

- The parent-children relationship of SNOMED CT's relative code is made free to study in **unconstraint** on the diversity of family relations information investigated by various organizations.
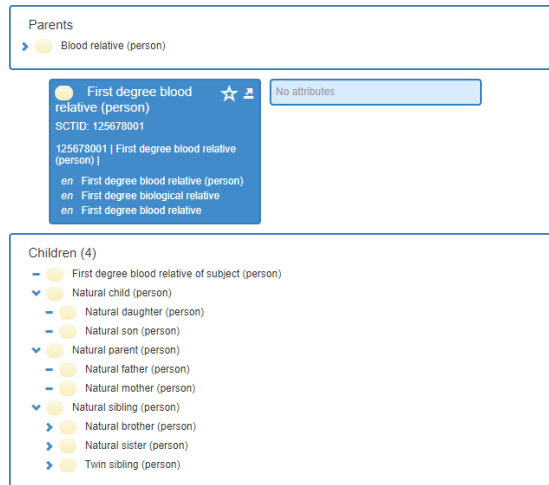

Figure12 . person information assessment grade result of two methods

- This study is based on medical check-up family history survey data from a single institution, So actual CDM data conversion and multicenter study was required.

Reference

1. Bennett R. L. (2019). Family Health History: The First Genetic Test in Precision Medicine. The Medical clinics of North America, 103(6), 957–966.

2. Cronin, R. M., Halvorson, A. E., Springer, C., Feng, X., Sulieman, L., Loperena-Cortes, R., Mayo, K., Carroll, R. J., Chen, Q., Ahmedani, B. K., Karnes, J., Korf, B., O'Donnell, C. J., Qian, J., & Ramirez, A. H. (2021). Comparison of family health history in surveys vs electronic health record data mapped to the observational medical outcomes partnership data model in the All of Us Research Program. Journal of the American Medical Informatics Association : JAMIA, 28(4), 695–703.

Acknowledgement

Smart Health Lab
Department of Digital Health

Thank you

Smart Health Lab
Department of Digital Health