Using FeatureExtraction (í•œêµì–´)

Jeon Ga Bin & Martijn J. Schuemie

2019-07-19

Contents

1 ì„œë¡ 1

2 ê³µë³€ëŸ‰ ì„¤ì • 1

2.1 ê¸°ë³¸ ê³µë³€ëŸ‰ ì§‘í•© ì‚¬ìš©í•˜ê¸° . . . 2
2.2 ì‚¬ì „ ì§€ì •ë œ ë¶„ì„ ì‚¬ìš© . . . 2
2.3 ì‚¬ìš©ìž ê³µë³€ëŸ‰ ì§‘í•©ì„ ìƒì„± . . . 3
2.4 ì‹œê°„ ê³µë³€ëŸ‰ . . . 4

3 ê´€ì‹¬ ì½”í˜¸íŠ¸ì— ëŒ€í•œ ê³µë³€ëŸ‰ êµ¬ì„± 5

3.1 ì„œë²„ ì—°ê²° ì„¤ì • . . . 5
3.2 ê´€ì‹¬ ì½”í˜¸íŠ¸ ìƒì„± . . . 6
3.3 ê´€ì‹¬ ì½”í˜¸íŠ¸ì— ëŒ€í•œ 1ì¸ë‹¹ ê³µë³€ëŸ‰ ìƒì„± . . . 7
3.4 ì½”í˜¸íŠ¸ì— ëŒ€í•œ ì§‘ê³„ ê³µë³€ëŸ‰ ìƒì„± . . . 9
3.5 í…Œì´ë¸” ë§Œë“¤ê¸° 1 . . . 11

1 ì„œë¡

íŠ¹ì§•ì¶”ì¶œ íŒ¨í‚¤ì§€ëŠ” ê³µí†µ ë°ì´í„°ëª¨ë¸ì— ì €ìž¥ëœ ì •ë³´ë¥¼ ì‚¬ìš©í•˜ì—¬ ì§‘ë‹¨(cohort)ì„ ìœ„í•œ íŠ¹ì§•(feature)ì„ ë§Œë“¤ ìˆ˜ ìžˆë‹¤. ì§‘ë‹¨ì´ëž€ ì¼ì • ê¸°ê°„ë™ì•ˆ í•˜ë‚˜ ì´ìƒì˜ í¬í•¨ ê¸°ì¤€ì„ ë§Œì¡±í•˜ëŠ” ì‚¬ëžŒë“¤ì˜ ì§‘í•©ìœ¼ë¡œ ì •ì˜ëœë‹¤. ìˆ˜ë¥¼ ë“¤ì–´ ì§‘ë‹¨ì— ë“¤ì–´ê°€ê¸° ì „ì— ê´€ì°°ëœ ì§„ë‹¨ì´ íŠ¹ì§•ì¼ ìˆ˜ ìžˆë‹¤. ì–´ë–¤ ì‚¬ëžŒë“¤ì€ ë˜ í•œ 'ê¸°ì¤€ íŠ¹ì„±'ê³¼ ê°™ì€ íŠ¹ì§•ì´ë‚˜ ì¼ë°˜ì ìœ¼ë¡œ 'ê³µë³€ëŸ‰'ìœ¼ë¡œ ë¶ˆë¦¬ëŠ” íŠ¹ì§•ë“¤ì„ ì–¸ê¸‰í• ìˆ˜ ìžˆìœ¼ë©°, ìš°ë¦¬ëŠ” ì´ëŸ¬í•œ ìš©ì–´ë“¤ì„ ì´ ê¸€ ì „ì²´ì— ê±¸ì³ ì„œë¡œ í˜¼ìš©í•˜ì—¬ ì‚¬ìš©í• ê²ƒì´ë‹¤.

ì´ ê¸€ì€ íŒ¨í‚¤ì§€ì— í¬í•¨ëœ ê¸°ë³¸ ê³µë³€ëŸ‰ ì •ì˜ë¥¼ ì‚¬ìš©í•˜ì—¬ íŠ¹ì§•ë“¤ì„ êµ¬ì„±í•˜ëŠ” ë°©ë²•ì„ ì„¤ëª…í•œë‹¤. ì´ëŸ¬í•œ ì •ì˜ë“¤ì€ ë¯¸ë¦¬ ì •ì˜ëœ ë§¤ê°œ ë³€ìˆ˜ë“¤ì„ í†µí•´ ìƒë‹¹ížˆ ë§Žì€ ì‚¬ìš©ìž ì„¤ì •ì„ í—ˆìš©í•˜ì§€ë§Œ, ë•Œë•Œë¡œ ë” ë§Žì€ ì‚¬ìš©ìž ì„¤ì •ì´ í•„ìš”í• ìˆ˜ ìžˆë‹¤. ì´ ê²½ìš° ë…ìž ëŠ” ì œëŒ€ë¡œ ì‚¬ìš©ìž ê³µë³€ëŸ‰ì„ ë‹¤ë£¨ëŠ” ì´ íŒ¨í‚¤ì§€ì— í¬í•¨ëœ ë‹¤ë¥¸ ê¸€ì„ ì°¸ì¡°í•˜ë©´ ëœë‹¤.

ì´ ê¸€ì€ ë‹¨ì§€ ì–´ë–¤ íŠ¹ì§•ë“¤ì„ êµ¬ì„±í• ì§€ ì§€ì •í•˜ëŠ” ë°©ë²•ì„ ì„¤ëª…í•œë‹¤. ë§Žì€ ìƒí™©ì—ì„œ ìˆ˜ë¥¼ë“¤ì–´ FeatureExtractionì„ CohortMethod ë˜ëŠ” PatientLevelPredictionê³¼ ê°™ì€ ë‹¤ë¥¸ íŒ¨í‚¤ì§€ì˜ ì¼ë¶€ë¡œ ì‚¬ìš©í•˜ëŠ” ê²½ìš° ì‹¤ì œ íŒ¨í‚¤ì§€ í˜¸ì¶œì´ ë‹¤ë¥¸ íŒ¨í‚¤ì§€ì— ì˜í•´ ì¼ì–´ë‚˜ë¯€ë¡œ FeatureExtraction íŒ¨í‚¤ì§€ì— ëŒ€í•´ ì•Œì•„ì•¼ í• ëª¨ë“ ê²ƒì´ ìžˆìŠµë‹ˆë‹¤. ê·¸ëŸ¬ë‚˜ ì§‘ë‹¨ì˜ ì„¤ëª…ì„ ì¢…ì´ì— ìž‘ì„±í•˜ëŠ” ë°©ë²•ì´ íŒ¨í‚¤ì§€ë¥¼ ì§ì ‘ ì´ìš©ìœ¼ë¡œ ì‚¬ìš©í•˜ëŠ” ê²ƒë„ ê°€ëŠ¥í•©ë‹ˆë‹¤.

2 ê³µë³€ëŸ‰ ì„¤ì •

ì‚¬ìš©ìžëŠ” ë‹¤ìŒ ì„¸ê°€ì§€ ë°©ë²•ìœ¼ë¡œ êµ¬ì„±í• ê³µë³€ëŸ‰ì„ ì§€ì •í• ìˆ˜ ìžˆë‹¤.

1. ê¸°ë³¸ ê³µë³€ëŸ‰ ì§‘í•© ì„ íƒ
2. ì‚¬ì „ì— ì§€ì •ëœ ë¶„ì„ ì§‘í•©ì—ì„œ ì„ íƒ
3. ì‚¬ìš©ìž ì •ì˜ ë¶„ì„ ì§‘í•© ìƒì„±

ë¶„ì„ ì € í• ˜ë‚˜ ì ´ìf ì ˜ ìœ ì‚¬í• œ ê³µë³€ëŸ‰ì „ „ ìf ì„±í• ˜ëŠ" ê³¼ì • •ì ´ë‹¤. ì˜ˆë¥¼ ë"¤ì–´ í• ˜ë‚˜ì ˜
ë¶„ì„ ì € ì€ ½ì"í~ˆíŠ˜ ì‹œ(cohort start)ì „ „ í• ˜ì— ì¡°ê±´´ë°œif í'œì—ì„œ ê´€ì¡ê ê° ê° ì¡°ê±´ì—€ ëŒ€íˆ´
ì ´ì§„ ê³µë³€ëŸ‰ì „ „ ìf ì„±í• ì̂ ̂ ìž̂ìœ¼ë©°, ë‹¤ë¥¸ ë¶„ì„ ì—ì„œëŠ" ì°° ìŠ¨ ê³µë³€ì̂± ì§€ì̂˜ë¥¼
ë˜̃ífë¸Ì'ëŠ" ë‹̃ì ¼ ê³µë³€ëŸ‰ì „ „ ìf ì„±í• ì̂ ̂ ìž̂ë‹¤.

íŠ¹ì§ • ë"¤ì „ ê¸¬ì„±í̂± ˜ëŠ"ë°€ ì̃ ½ë̃¸í• ì̂ ̂ ìž̂ê±±°ë̃ ì̃ ì̃±ìœ í• ì̂ ̂ ì—†ëŠ" ê°œë̋… ID ì§'í• ©ì "" ì§̂ì • •í• ˜ëŠ"
ê²fì ´ìf ê²€€ëŠÌ¥í̃ ˜ë‹¤. ê̄°ê³ ì§'í̂±©(ì̃ µì…̃ 1) ë̃ ̃ ëŠ" ì̃̃ "" ì§̂ì •ë • ễë¶„ì̋„ (ì̃ µì̃̃ 2)ì ̃ ˜̃, ífí• ë•ŒëŠ"
ë̂ ̃ë̃̃ ë¶„ì̋„ ì̃ ì̃, œ ì ´ ìž̃ì̋̃—…ì "" ì̂ ̂í̃—‰ í• ì̂ ̂ ìž̂ë̃¤. ì,¬ìš©ìž ì "ì̃ ˜ ë¶„ì̋„ (ì̃ µì̃̃ 3)ì "" ífí• ± í• ì̂ ̃ ë̃ Œ
ë¶„ì̋„ ë̃³„ìœ œ ì§̂ì •í• ì̂ ̂ ìž̂ë̃¤.

ê³ ê̄,‰ ì,¬ìš©ìž ë¥¼ ìœ œ,í̃ ´: ífí• ì̂í• ì̃ ¼ í• ˜ëŠ" ê³µë³€ëŸ‰ ID ì§'í• ©ì "" ì§̂ì • •í• ì̂ ̂ë̃ „ ìž̂ë̃¤.
í• œ ê°œì̃ ˜ ê³µë³€ëŸ‰ IDëŠ" íŠ¹ì • • ê³µë³€ëŸ‰ì "" ì̃ ë̃³„í• œ ì̃ œì̂ (ì̃̃: Charlson Comorbidity ì§̂ì̃̂ ë̃ „ëŠ"
íŠ¹ì • • ì̃ œì̂„ì̃, œ íŠ¹ì • • ì̂ í̂±°í̃ ê°œë̃̃… ì̃ ˜ ê²°ëˆf). ê̄· ëŸ¬ë̃ €ê¡œ ê³µë³€ëŸ‰ IDëŠ" ê°œë̃̃… IDì̃™€
í̃ ¼ë ™í• ˜ễì̃´ ì ì̂•ë œë̃¤. ífí• ì̂ ê³µë³€ëŸ‰ IDë¥¼ ì§̂ì • •í• ì̂ ̂ ìž̂ëŠ" ì ¼ë̃°ì̃ ì̂ ì̃œì̃, ¦í̃ì̃̃ ¤ëŠ"
ë̃„ê̯°ê̂±°ê̂€ ì ´ë̄̃, í• ì̃ œ ë̃̃̃ ì̂§'ë̃˜ì̃—ì̃, œ ê³µë³€ëŸ‰ì "" ë̃§Œë̃¤̃—ì̃—í̃ì̃̄¸ë̄̃, ê³µë³€ëŸ‰ì ̃ ̃̃ í• ˜ìœ œ„ ì§̂ì • •í• ̃ "
ì̃³¾ì̃•„„ë̃f̃±°í̃ì̃̃ ̃, ë̃¤ë̃¥̃ ễì̃§̂ë̃̃—ì̃, œ ê̄°ê̄,±ë̃ œ ê³µë³€ëŸ‰ì̃ "" ễ̂ì§̂ê³ ì̃¶̃ "" ê̄½̃ì̃°̃ ´ë̃¤

2.1 ê̄°ë³¸ ê³µë³€ëŸ‰ ì§'í• © ì,¬ìš©í• ˜ê̄°

ê̄°ë³¸ ê³µë³€ëŸ‰ ì§'í• ©ì "" ì,¬ìš©í• ˜ëŠ" ê²fì € ê°„ë̃„í• ˜ë̃¤:

```
settings <-createDefaultCovariateSettings()
```

ì ´ê²fì € ì ̄°ë̃̃¬í̃µê³„„í̃ ̃™ì̃— ì̃,œë¶€í̃̃,° ì̃̃ê±±ễ¾ ì ½ë̃̃¼ì "" í̃̃µí̃ì̃œ ì̃—ëŸ¬ ìœ,í̃̃ ì ì̂˜ễÌ̂ €̂ ë̃¤ì̃—
í• ì̃ œ íŠ̃ì§̂̂ë̃̃̃ "" ë̃§Œë̃¤̃—ë̃¼ ễ̂ë̃¤.

ê³µë³€ëŸ‰ì "" ífí• ì̃ ˜ëŠ"ë̃ ° ì̃,ì̂ì̃ ì̃,œễ ì̃̃ ̂ë̃ ˜̃ ê°œë̃̃… ì§'í• ©ì "" ì§̂ì • •í• ̃̃ë̃ „ ìž̂ë̃¤

ì̃̂ë¥¼ë̃̃¤ì̃-:

```
settings <-createDefaultCovariateSettings(excludedCovariateConceptIds = 1124300,
                                          addDescendantsToExcl
```

ì ´ê²fì € ê°œë̃̃… 1124300(ì,±ë̃¶„ ë̃̃"í ´ë̃jœìŽ̃˜ë̃™)ê³¼ ê̄· í̃,ì̃†í• ̃"¤(ì̃¦̃‰ ë̃̃̃"í ´ë̃jœìŽ̃˜ë̃™ ì,±ë̃¶„ì̃̃ „
í•¨ìœ̃ëœí• ì̃œ ë̃ª̃"̃̃ë̃ª̃°ì̃̃¦½ë̃̃¼)ì̃—ì̃̃,œ íŒ̃Œìf ë̃ œ ê²fì̃ €̃ễ ì̃™í• ̃̃³ ê̄°ë³¸ ê³µë³€ëŸ‰ ì§'í• ©ì "" ífí• ì̃ œë̃̃¤.

2.2 ì,¬ì „ ì§̂ì • ë̃ œ ë̃¶„ì̋„ ì,¬ìš©

createCovariateSettings ê̄°ëŠ¥í̃̃ ̃ ̃ ì,¬ìš©ìž ê°̃̃ ̃ ë̃̄̃ë̃¦̃ ì ̃ •ì̃ ˜ ễ œ ê³µë³€ëŸ‰ ì§'í• ©ì̃—ì̃,œ ì̃„ ífí• ì̃ ̃̃
ìž̂ë̃ ë̃¡̃ í• ̃̃œë̃jœë̃¤.

ì,¬ìš© ễ̂ëŠ¥í̃ í• œ ì̃ µì̃̃̃ì̃— ë̃ Œ̃̃í ì̃ œ ê°œë̃ì̃̃™ë¥¼ ë̃³̃ ¤ì̃ œ́ createCovariateSettingsë¥¼ ì,¬ìš©í• ̃ œë̃¤

ì̃̂ë¥¼ë̃̃¤ì̃-:

```
settings <-createCovariateSettings(useDemographicsGender = TRUE,
                                   useDemographicsAgeGroup = TRUE,
                                   useConditionOccurrenceAnyTimePrior
```

ì ´ë ‡ê²Œ í• ˜ë̃̃´ ì,±ë³„„, ì—°ë ¹(5ì̃„ „ ê̄· ê£¹) ë̃° ì½"í̃̃̃íŠ̃ ì̃œì̃ž̃ ë̃, ì§̂œ ì ̂̃, ̃(ê̄· ̃œ¦̃ê³ í́í• ̃ ̃ ̃̃̃íŠ")
ì̃°ê±±ễ€̃ ̂ë̃œễjœíf ê̄°ëŠ¥í̃̃ ̃̃— ì̃, œ ê̄°ì̃̂°̃ë̃̃ œ ễ ê°œë̃̃… ì̃— íŒ̃̂ì̃ œ ì ̃§', ê³µë³€ëŸ‰ì ̃ ̃ ífí• ì̃ œë̃¤.

ì,¬ì „ ì§̂ì • ë̃ œ ë̃¶„ì̋„, ìf̃ ̃ë̃§̃Ž̃̃ ̃ ễ̃ë̃¶̃ë̃¶̃ì̃ ̃ ́ë̃̃̃̃,̃, ìf̃ ̃ễ, ̃ë̃̃ ̃ëŠ" ìž̃¥̃ễ, ̃ễ° ̃̂°̃„ „ ë̃, ̃̃ ̃ífí̃ë̃̃,ë̃̃¤. ê̄°ë³¸ì̃ ìœ̃¼ë̃ jœ
ë̃̃̃ì̃— ì̃™̃̃ €̃ ễ̃™ì̃ ́ì̃ ̃•ì̃ ̃ ̃ ̃̃̃ ë̃œë̃̃̃¤

- ìž̂¥̃ễ, ̃ : ì½"í̃̃̃íŠ̃ ì̃œ̃ìž̃ ̃ë̃̃, ì§̂œễ ̃ Œ̃ì§̂̂ 365ì̃ ¼ ì̃̃̃ ̃ „ễ° í̃¬ì̃ ̃ ̃̃̃ ¨.
- ì¤̃ễ, ̃ : ìµ̃œ̃ễŒ̃̃ 180ì̃ ¼ ì̃̃ „ ễ° ì̃½"í̃̃̃íŠ̃ ì̃œ̃ìž̃ ̃ë̃̃, ì§̂œ í̃¬ì̃ ̃ ̃̃̃ ¨.
- ë̃̃̃ễ, ̃ : ì½"í̃̃̃íŠ̃ ì̃œ̃ìž̃ ̃ì̃̃¼̃ì̃̃ ̃ë̃¼ễjœ̃¶ễ¶í̃¬ì̃̃,° ̃ìµ̃œ̃̃ Œ̃̃ 30ì̃ ¼ ì̃ „ễ̂œ̃ì§̂̂.

ê̄· ̃ë̃Ÿ¬̃ë̃̃̃, ì,¬ìš©ìž ëŠ" ì̃ ́ë̃̃̃̃ ̃ë̃ì̃ ̃ ́ ễ̃°̃ì̃ "" ê³ ễ̃̂½̃ í• ì̃ ̃̃̃ ̃ ̃ìž̃̃ë̃̃¤.

ì̃̂ë¥¼ë̃̃¤ì̃-:

2

```r
settings <-createCovariateSettings(useConditionEraLongTerm = TRUE,
                                   useConditionEraShortTerm = TRUE,
                                   useDrugEraLongTerm = TRUE,
                                   useDrugEraShortTerm = TRUE,
                                   longTermStartDays = -180,
                                   shortTermStartDays = -14,
                                   endDays = -1)
```

ì ´ê²fi €ì½"í˜ ̡Š ì‹œìž'ì ¼ë¡œë¶í„° 180ì ¼ ì„(í ¬í• ¨í• ˜ì§€ì•Ší Œ)ê¹Œì§€ ìž¥ê ̃°ë¥¼ ìž¬ì •ì ˜ê³ ì½"í˜ ̡Š ì‹œìž'ì ¼ë¡œë¶í„° 14ì ¼ ì„(í ¬í• ¨í• ˜ì§€ ì•Š Œ)ê¹Œì§€ ë‹ễ°ë¥¼ ìž¬ì •ì ˜ í•œë‹¤.

ë˜ í•œœ ê³µë³€ëŸ‰ ê¹ ¬ì„¸± — ì,¬ì©í• ˜ì• ¼ í˜Š" ê°ë… IDë¥¼ ì§€ •í• ì^ë„•ì•í ì^ë̃ž^ë‹¤.

```r
settings <-createCovariateSettings(useConditionEraLongTerm = TRUE,
                                   useConditionEraShortTerm = TRUE,
                                   useDrugEraLongTerm = TRUE,
                                   useDrugEraShortTerm = TRUE,
                                   longTermStartDays = -180,
                                   shortTermStartDays = -14,
                                   endDays = -1,
                                   excludedCovariateConceptIds = 11243
                                   addDescendantsToExclude = TRUE)
```

2.3   ì,¬ìš©ìž ê³µë³€ëŸ‰ ì§'í•©ì „ ìf ì„±

ì ´ ì˜µì…˜ì € ê³ ê‰‰ ì,¬ìš©ìž ì§Œ ì,¬ìš©í• ´ì•¼ í•œë‹¤. ê¹¬í˜„ ë ˆë²ì—ì„œ ë¶„ì„ ì € ê³ ë ˜ë¡œœ ë§¤ê°œ ë³€ì ^í™" ë œ SQLê³¼ ë§¤ê°œ ë³€ì ^ ê°¹ì ˜ ì Š¤íŽ™ì „ ê²²ì •í• œ ê²²fiž„ì „ ì ´í• ´í• ˜ë¼¼í• œ œë̃¤. ì,¬ìš© ê°°ëŠ¥¥í• œ ì˜µì…̃ì „ ì ´í• ´í• ˜ëŠ"œê°°ìž¥ ì¢ì €©ë² ì € ì,ì „ ì§Œì •ë œ ë¶„ì„ ì „ ì§'í•©ì „ ̃°ë¡œ ë³€í• œ í•˜ì`ì ´í• œ ê²°ê³¼ë¥¼ œ ê²°fi ´ë‹¤

```r
settings <-createCovariateSettings(useConditionEraLongTerm = TRUE)
settings2 <-convertPrespecSettingsToDetailedSettings(settings)
settings2$analyses[[1]]
```

```
## ê²°ê³¼
## $analysisId
## [1] 202
##
## $sqlFileName
## [1] "DomainConcept.sql"
##
## $parameters
## $parameters$analysisId
## [1] 202
##
## $parameters$analysisName
## [1] "ConditionEraLongTerm"
##
## $parameters$startDay
## [1] -365
##
## $parameters$endDay
## [1] 0
##
## $parameters$subType
## [1] "all"
```

```
## 
## $parameters$domainId
## [1] "Condition"
## 
## $parameters$domainTable
## [1] "condition_era"
## 
## $parameters$domainConceptId
## [1] "condition_concept_id"3
## 
## $parameters$domainStartDate
## [1] "condition_era_start_date"
## 
## $parameters$domainEndDate
## [1] "condition_era_end_date"
## 
## $parameters$description
## [1] "One covariate per condition in the condition_era table overlapping with any part of the long te
## 
## 
## $includedCovariateConceptIds
## list()
## 
## $includedCovariateIds
## list()
## 
## $addDescendantsToInclude
## [1] FALSE
## 
## $excludedCovariateConceptIds
## list()
## 
## $addDescendantsToExclude
## [1] FALSE
```

ìf ì„„‚ ë¶‚ì‚ ì‚¤ì • ê°° ì²´ë¥¼ ì²˜ì Œë¶€í‚°° ë§Œë“ë“¤ ì^˜ ìž^ê³ ì ´ê²fì „ ì‚¬š©í• ˜ì—¬ ì‚„ë¶€ ì‚¤ • ê°° ì²´ë¥¼
ìf ì„„±í•œ ë‹¤

```
analysisDetails <-createAnalysisDetails(analysisId = 1,
                                                              sqlFileName = "Demograph
                                                              parameters =list(analysi

                                                              includedCovariateConcept
                                                              addDescendantsToInclude
                                                              excludedCovariateConcept
                                                              addDescendantsToExclude
                                                              includedCovariateIds =c
settings <-createDetailedCovariateSettings(list(analysisDetails))
```

2.4 ì‹œê°„ ê³µê³€ë³€ìˆ˜

ì¼4ë°˜ì ìœ¼ë¡œ ê³µê³€ë³€ìˆ˜ê°€ ê°´€ì‹¬ ì‹œê°ŽëŠ" ë‹¨ ë§‡ ê°°œì ˜ ì‹œê°í•„(ì˜~^: ì •žì—ì„œ ì‚œ ì‚¤ë³…í•¬œ ë‹"ê³¸ë°°°,ì¤'ê³¸ë°°,ìž¥ê°¸ë°°)ì— ëŒ€í•´ ´ì„œé§Œ ì‹fì‚±ë œë‹¤.

ê· ‚ëŸ¬ë‚˜ ë•Œ¡œë§Š" ë "ì„ì‚ ê²Œ ì‹œê°ñ,„ ë¶„í•´ ´í•œ ~ëŠ" ê²fî ´ í•‚,ìš"í• ë•Œë „ ìž^ë‹¤. ì~^ë¥¼

ë"¤ì–´, ì½"í˜¸íŠ¸ ì‹œìž' 365ì ¼ ì „ì— ê° ë, ì§œì— ëŒ€í•œ ê³µë³€ëŸ‰ì „ ë³€ë¡œ ìž'ì"±í•œëë‹¤. ìš°ë¦¬ëŠ" ì ´ëŸ° ì¢…ë¥˜'ì ˜ ê³µë³€ëŸ‰ì „ ì½œ¿" ê³µë³€ëŸ‰ì ´ë ¼ê³ í•œëë‹¤.

ì½œ¿" ê³µë³€ëŸ‰ì € ì½ê°"ëŒ€ì—ì"œ ë™ì ¼í • ê³µë³€ëŸ‰ ID ê³µìœ í•˜ê©© ë³"ë "ì ˜ ì½œ¿" ID능¼ ì¬ìŠ©í•'ì—¬ ì½œ¿"ëŒ€ëŸ¼ êµ¬ë³"í•©ì ˆë‹¤. í˜ˆìž¬ëŠ" ì½œ¿" ê³µë³€ëŸ‰ì € ì²˜ë¦¬í• ì ˆ ìž^ëŠ" ë§Žì € ì¬'í"Œ리ì¼œ ì…¿ì ´ ì—†ëŠ¤. ì^˜ë¥¼ë"¤ì–´ CohortMethod íŒ"í,¤ì§€ëŠ" ì½œ¿" ê³µë³€ëŸ‰ì " ì ¨ë³µí• ë•Œ ê°^¹§§¸ë‹¤. ê·¸ëŸ¬ë,˜ PatientLevelPrediction íŒ"í,¤ì§€ì— ëŠ" ì½œ¿" ê³µë³€ëŸ‰ì ´ í•"ìš"í•œ ê,ºê³,í•™ìŠµ ì•Œê³ ë¦¬ì¦ì ´ ìžˆë‹¤.

ë‹¤ì½œ ê,ºê³³ì"¤ì " ì"ë¡œ í•œ ê¸ ì" íƒí• ì^˜ ìž^ì Œ

settings <-createDefaultTemporalCovariateSettings()

ë˜ ëŠ", ë¯"ë¦¬ ì •ì˜ëœ ì½œ¿" ê³µë³€ëŸ‰ ì¤'ì— ì"œ ì" íƒí• ì^˜ ìž^ë‹¤.

settings <-createTemporalCovariateSettings(useConditionOccurrence = TRUE,

useMeasurementValue

ì ´ ê²½ìš°ìŠ° ìŠ°ë¦¬ëŠ" condition_occurrence í...Œì "ë¸"ì ˜ ê° ê°œë...ì— ëŒ€í•´ ì ì§", ê³µë³€ëŸ‰ì " if ì,±í• ˜ê¸œìœ ì, if í• ˜ì^€ê³, CDMì˜ measurement í...Œì "ë"ì— ìž^ëŠ" ê° ì¸ì • ì¡°í•©ì— ëŒ€í•´ ì — °† ê³µë³€ëŸ‰ì " if ì,±í• ˜ê¸œìœ í–ê‹¤. ê¸ê³³ì ìœ¼½ì ì½œ¿" ê³µë³€ëŸ‰ì € ì½œ¿ìž'ì ¼ 365ì ¼ ì „ì (í ¬í• ¨í• ˜ì§€ ì•Šì Œ)ì— ê°œìš"³ì ìœ¼½ì ê° ë, ì— ëŒ€í•´ ´ if ì,±ë œë‹¤. ë‹¤ë¥, ì½œ¿"ëŒ€ë¥¼ ì§€€ ì• í• ì^˜ë " ìž^ì^˜ë " ìž^ëŠ"ë° ì^˜ë¥¼ ë"ë©´ ë°ì— ì½œ¿"ìœ ì²˜ëŸ¼ ë§Œì œ"¤ 7ì ¼ê°œ"ê²©ìœ¼ìœ¼ë¡œ ë§ŒëŠ¤ ì^˜ ìž^ë‹¤

settings <-createTemporalCovariateSettings(useConditionOccurrence = TRUE,

useMeasurementValue
temporalStartDays =
temporalEndDays =se

ê° ì½œ¿"ëŒ€ëŠ" ì§€€ ë˜ì½ ì½œ¿ìž'ì ¼ê³¼ ì¢...ê£Œì ¼ì ¼ì „ í¬í• ˜í• ì—†ë‹¤. ì ¼ë˜˜ ê³µë³€ëŸ‰ê³¾ ìœ ì,¬í• ˜ê²Œ ê³ ê¸‰ ì,¬ìš©ì ë,¬ìš©ì ì•ì ˜ ë¶,ì,ì,ì •ì ˜ ì^˜ ìž^ë‹¤.

analysisDetails <-createAnalysisDetails(analysisId = 1,

sqlFileName = "Measureme
parameters =list(analysi

includedCovariateConcep
addDescendantsToInclude
excludedCovariateConcep
addDescendantsToExclude
includedCovariateIds =c

settings <-createDetailedTemporalCovariateSettings(list(analysisDetails))

3 ê´€ì‹¬ ì½œ¿"íŠ¸ì— ëŒ€í•œ ê³µë³€ëŸ‰ êµ¬ì„±

ì—¬ê,ºì"œëŠ" ë´ ê°ˆì‹¬ ê´€ì‹¬ ì½œ¿"íŠ¸ì— ëŒ€í•œ ê³µë³€ëŸ‰ if ì,±, ì¦ diclofenacaì˜ ìˆˆ†ì¡œìŠ´ ì,¬ìŠ©ìž ì™€ cellecoxiì˜ ìˆˆ†ì¡œìŠ´ ì,¬ìŠ©ìž ì^˜ë¥¼ ì,'íŽ ë³'ê,ºì ìŠ©í• ì•ëŠ¤.

3.1 ì,œë²", ì—°ê²° ì,¤ì •

ìŠ°ë¦¬ëŠ" Rì–,ì–´ë©œ ë °ì ´í,°ê²€ ìž^ëŠ" ì,œë²²,ì™€ ì—°ê²°í• ˜ëŠ" ë³©ë²•ì•ì ˜ ì•ŒëŠ•ì ¼ í• ìŠ©ë‹¤. CohortMethodëŠ" createConnectionDetails í• ì̂˜ë¥¼ ì œê³µí• ˜ëŠ" DatabaseConnector íŒ"í,¤ì§§€¥¼ ì,¬ìŠ©í• ˜ë‹¤

ë‹¤ì–'í• ˜ì œ ë°ì ´í,°ê²²ë ¼ ì˜Š¤ ê´€ë¦¬ ì½œìŠ¤í…œ (DBMS)ì— í• ,ìŠ"í• œ íŠ¹ì ¼ì • ì,¤ì •ì •¤ ,ìœ,íœ œ createConnectionDetailsì " ì,¬ìŠ©í• ˜ëŠ¤. ìˆˆ†ë¥¼ë"¤ì–´ ì ´ ì½'ë"œìŠ¤¥¼ ì,¬ìŠ©í• ˜ëŠ" PostgreSQL ë°ì´í,°ë²² ì˜Šê¤ì—ì—ê²²'í• ì^˜ ìž^ë‹¤.

```
connectionDetails <-createConnectionDetails(dbms = "postgresql",
                                             server = "local
                                             user = "joe",
                                             password = "sup

cdmDatabaseSchema <- "my_cdm_data"
resultsDatabaseSchema <- "my_results"
```

ë§ˆì§€ë§‰ ë' ë ¼ì¸ì € cdmDatabaseSchema ë³€ì^^ì™€ resultsDatabaseSchema ë³€ì^^ë¥¼ ì •ì ˜í—^ë‹¤. ë¸ì¤‹— CDM í^•ì‹ì ˜ ë°ì í‚°ê°€ ì–'ë""ì— ìž^ëŠ"ì§€, ì¤'ê°, ë° ê²°ê³¼ í'œë¥¼ ìž'ì„±í• ˜ë ¤ëŠ" ìœ"ì¹¹^ë¥¼ Rì–'ì–'ëjœ ì 'ê·¼í• ˜ê¸° ìœ"í•´ 'ì 'ëŸ¬í• œ ë³€ì^^ë¥¼ ì,¬ìš©í• œœë‹¤. Microsoft SQL Serverì ˜ ê²½ìš° ë °ì í""ë²² ì 'ìŠ¤ ìŠ¤í,¤ë§^ëŠ" ë °ì í""ë²² ì 'ìŠ¤ì™€ ìŠ¤í,¤ë§^ë¥¼ ëª"ë' ì§€ì •í• ˜ì• 'ì•¼í•˜ë¯€ëjœ ì^ˆë¥¼ ë"¤ì–' cdmDatabaseSchema <-"my_cdm_data.dbo" ì 'ëŸ°ì‹ ì 'ë‹¤.

3.2 ê´€ì‹¬ ì½"í̃¸íŠ¸ ì íˆ,±

FeatureExtractionì „ ì,¬ìš©í•´ë ¤ë©´ ê³µí†µ ë°ì í‚° ëª"ë¸ì¸ ì½"í̃¸íŠ¸ í…Œì´ë„" ë̃ ëŠ" ì½"í̃¸íŠ¸ í…Œì´ë,"ê³¼ ë ™ì ¼í•´ ê¶ì—¡ìj°ë¥¼ ê°–ëŠ" í…Œì´ë,"ì— ì„,œì½"í̃¸íŠ¸ë¥¼ ì ìŠ¤í,¸ìŠ¤í™"í• ì•ì•¼í•©ë‹¤. ì½"í̃¸íŠ¸ ë' ë í´'ëjœ ì½"í̃¸íŠ¸ë¥¼ ë§Œë"¤ ì^ 'ìž^ì§€§Œ ì—¬ê¸°ì„,œëŠ" ê°"ë„'ë̃ ì̃— ˜ì— — 1ì ˆë‹¹ ì²«ë²²^ì§€ ì•½½ë¬¼ì‹œœ^ë°¥¼ ì³¾ëŠ"ë‹¤. ì½"í̃¸íŠ¸ë¥¼ ì‹œìž'í•´˜ê, ì ,ì— — ê³µê²³€ëŸ‰‹ „ ìf ì,±í• ˜ë• Œë¬¬ì— — ì²«ë²²^ì§§ ê…ì¶¶€ë¥¼ ì ,ì— — 365ì ¼ ê°",ì ˜ ê´€ì°°ì ´ í• "ìš"í• ˜ë‹¤. FeatureExtractionê°€í•§ễ •ë œ ê¸°ê°",„ ë ™ì•^ í'¼ì,¬ì²´ê€ ê´€ì°°ë ˜ŠŠ"ì§€ í™•ì ,í• ˜ì§€§€ì•• ŠˆŠë¤.

/*************************************/

File cohortsOfInterest.sql

*************************************/

```
IF OBJECT_ID('@resultsDatabaseSchema.cohorts_of_interest','U') IS NOT NULL
    DROP TABLE@resultsDatabaseSchema.cohorts_of_interest;
SELECT first_use.*
INTO @resultsDatabaseSchema.cohorts_of_interest
FROM(
    SELECT drug_concept_id AS cohort_definition_id,
    MIN(drug_era_start_date)AS cohort_start_date,
    MIN(drug_era_end_date) AS cohort_end_date,
    person_id
FROM @cdmDatabaseSchema.drug_era
WHERE drug_concept_id=1118084 -- celecoxib
    OR drug_concept_id=1124300 --diclofenac
GROUP BYdrug_concept_id,
    person_id
) first_use
INNER JOIN @cdmDatabaseSchema.observation_period
    ON first_use.person_id=observation_period.person_id
    AND cohort_start_date>=observation_period_start_date
    AND cohort_end_date<=observation_period_end_date
WHERE DATEDIFF(DAY, observation_period_start_date, cohort_start_date)>=365;
```

ì ´ Sqlì € SqlRender íŒ¨í‚¤ì§€ì—ì,œ ì,¬ìš©í• ì^^ ìž"ëŠ" ë§¤ê°œœë³€ì^^í"™"ë̃œ SQLì 'ë‹¤. ë§¤ê°œœ ë³€ì^^ê°€ í¬í•¨ë̃ SQLì „ ì,¬ìš©í•´ë €ëjœ CDM ë° ê²°ê³¼ ìŠ¤í,¤ë§^ì ˜ ì 'ë¦„ì „ ë̃ ¸ë¦¬ ì§€ ì•í• í•„ìŠ"ê°€ ì—†ë‹¤. ê· ´ëŸ° ì‹ ìœ¼½ëjœ ë°¤ë¥ ìŠ¤í,¤ë§^ì— ì,œ SQLì „ ì‹í—‰‰í• ˜ë ¤ê³ ì§€ê°œœ ë³€ì^^ ê°'ë§Œ ë³€ê²½í•˜´ë©´ ë̃œ ë‹¤. ìŠ"ë¦¬ëŠ" SQL ì½"ë"œ뜼 ë³€ê²'½í•˜´ í•„ìŠ"ê°€ ì—†ë‹¤. ë̃ í• œ SqlRenderì ˜ ë²'í• ì€ ê° ˜ëŠŠ¥ì „ ì,¬ìš©í•´í•ì—¬ ë̃ë¥ì— — ë‹¤í•ì ˜ í•ì ˜ 츤½ì— — ì,œ SQL ì½"ë"œ뜼 ì‹í—‰%í• ì^^ ìž^ë‹¤.

```
library(SqlRender)
sql <-readSql("cohortsOfInterest.sql")
sql <-render(sql,
```

```
                         cdmDatabaseSchema = cdmDatabaseSchema,
                         resultsDatabaseSchema = resultsDatabaseSchema)
sql <-translate(sql, targetDialect = connectionDetails$dbms)
connection <-connect(connectionDetails)
executeSql(connection, sql)
```

ì ´ ì½”ë“œì— ì„œëŠ” ë¨¼ì € íŒŒë¼ë¯¸í„°ë¥¼ ì„œ ë©©ë°¨ë¦¬ë¡œ SQLì „ ì ½ëŠ”ë‹¤. ë‹¤ì Œ ì¤„ì—ì„œ ë‹œ ê°œì ˜ ë§¤ê°œ ë³€í•˜ì ´ë¦¬ì „„ ì‹¤í–‰ë‹¤. ê· ¸ëŸ° ë‹¤ì Œ SQLì „ ì—°ê²° ì‚¸ë¶€ ì •ë³´ì—ì„œ ì •ì˜í•œ ì‹œ ê§€ì •ë« í•œ ì¤‘ì—ì„œ DBMSì— ì í •ì•œ ë³€í™˜í•´ ì¤€ë‹¤. ë‹¤ì Œì½”ì˜œ, ì‹œ°ë¦¬ëŠ" ì"œë²²"ì— ì—°ê²ºí• ~ê³ ë Œë "ë§ ë "ê³ ë²^ì— ë œ SQLì „ ì œ¶œí•œ ì œ¶©ë‹¤.

ì§€ê´ˆ¹êµ¹ì§€ ìž˜í-ê'¤ëˆ©´, ìš°ë¦¬ëŠ" ì ´ì œ ê´€ì‹¬ ë°½"í~íŠ‚ì ~ í…Œì ´ë"ì " ë§§Œë"^ë‹¤. ì‹œ°ë¦¬ëŠ" í• ¸ ì¢…ë¥˜ë"¹ì ì—¼ë§§ë§¤ë²ˆì ì œ~Žì € ì ´²¤íŠ‚ë¥¼ ë³¼ ì^ ìž^ëŠ" ì§€ ì• ˜ ìˆ^ì ìž^ë‹¤.

```
sql <-paste("SELECT cohort_definition_id, COUNT(*) AS count",
                      "FROM @resultsDatabaseSchema.cohorts_of_interest",
                      "GROUP BY cohort_definition_id")
sql <-render(sql, resultsDatabaseSchema = resultsDatabaseSchema)
sql <-translate(sql, targetDialect = connectionDetails$dbms)
querySql(connection, sql)

## ê²°ê³¼
##    cohort_concept_id  count
## 1           1124300 240761
## 2           1118084  47293
```

3.3   ê´€ì‹¬ ì½"íŠ¸íŠ¸ì— ëŒ€í• œ 1ì ¸ì‹ ê³µë³€ëŸ‰ ìf ì„±

ì˜ˆ¥¼ ë"¤ì–´ ê¸°ë³¸ ì„¤ì •ì •ì „ì„,ì"¬ì˜í• ~ì—¬ ê´€ì‹¬ ë°½"í~íŠ‚ ì¤'í ˜,ì—¼ ëŒ€í• œ 1ì ¸ì‹ ê³µë³€ëŸ‰ì „ ìf ì„± í• ì^ ìž^ë‹¤.

```
covariateSettings <-createDefaultCovariateSettings()
covariateData <-getDbCovariateData(connectionDetails = connectionDetails,
                                            cdmDatabaseSchema = cdmDatabaseSch
                                            cohortDatabaseSchema = resultsDatal
                                            cohortTable = "cohorts_of_interest"
                                            cohortId = 1118084,
                                            rowIdField = "subject_id",
                                            covariateSettings = covariateSettin
summary(covariateData)

## ê²°ê³¼
## CovariateData object summary
##
## Number of covariates: 41330
## Number of non-zero covariate values: 25892630
```

3.3.1   ì,¬ìš©ìž ê³µë³€ëŸ‰ ì¶œë ¥ í˜ •ì‹

covariateData ê° ì²´ì " ì£¼ìš" ê¶¬ì„,ë± ìš"ì††ê²Š" ê³µë³€ëŸ‰ì ´ë‹¤.

```
covariateData$covariates

## ffdf (all open) dim=c(25892630,3), dimorder=c(1,2) row.names=NULL
## ffdf virtual mapping
##              PhysicalName VirtualVmode PhysicalVmode  AsIs VirtualIsMatrix PhysicalIsMatrix Physica
## rowId              rowId      integer       integer FALSE           FALSE            FALSE
```

```
## covariateId       list...1     double    double FALSE         FALSE            FALSE
## covariateValue     list...2     double    double FALSE         FALSE            FALSE
## ffdf data
##              rowId covariateId covariateValue
## 1                1  4185976212              1
## 2                2   316866212              1
## 3                3  4028876212              1
## 4                4  4132926212              1
## 5                5   318736212              1
## 6                6  4208390212              1
## 7                7  4154739212              1
## 8                8  4197065212              1
## 25892623  25892623        1903              3
## 25892624  25892624        1903              1
## 25892625  25892625        1903              2
## 25892626  25892626        1903              1
## 25892627  25892627        1903              2
## 25892628  25892628        1903              3
## 25892629  25892629        1903              1
## 25892630  25892630        1903              2
```

ì»¬ëŸ¼ë“¤ì € ë‹¤ì Œê³¼ ê°™ì ´ ì •ë ˜ë œë‹¤

- rowId ê³ ìœ í• ê²Œ ì½”ìŠ¸ í• ëª©ì „ „ì‹ ê³ í•œë‹¤. getDbCovariateDataë¥¼ í˜¸ì¶œ í• ë•Œ rowIdField = "subject_id"ë¥¼ ì •ì í–^ìœ¼ë¯€ë¡œ ì ´ ê²½ìš° rowIdëŠ" ì½”ìŠ¸ í…Œì ´ë¸ˆ ì— subject_idì™€ ë™ ì¼í•˜ë‹¤. ì ´ ë‹ì ~ ì£¼ì˜ œ€ í• ì •ì•Šì „— ë'íŒ€, ì‹œì¦ˆ˜"ê²½ì ˆ ìŠ¤í— í…Œì ´ë¸ˆ í• „"œ"œìŠ¤í— í…Œì ´ê²½ì í˜„ê³ ì½”ìŠ¤ ì½”ëž¬ í• „"œ"— ìž'ì„±í• ê³ í•¥ rowId í• „"œ"œëŠ" ê²Œ ë¬ ¤ìž^ë‹¤.
- covariateëŠ" ê³ ìœ í•ŽÂ°ì½”ìŠ¤ì "ê³ covariatesì~ ì • ì ~ëŠ" cohortData$covariateRef ê° ì²˜ì— ì,œ ì ³¾ì „—ìž^ë‹¤.
- covariateValue í• „"œ"œëŠ" ê°°ì" ì œ³ í• œ ë‹¤.

3.3.2 ë°°ì ´í„°ë¥¼ íŒŒì‹¼ë¡œ ì €ìž¥

ê³ ìœ í•ŽÂ°ì € í•í•±ì „"œ ê²fí € ìf ë‹í• œ ì»í"íŒ… ì«œê°°ì ´ ë°"¤ê³ , ë°¸ëž˜ë¥¼ ëŒ€ë¹ í• ì—¬ í–¥í•„ ì‹í… œ ì«œ ì € ê³ ìœ í•ŽÂ°ì € ì €ìž¥í• ì™€ ê²fíŒ ì • ê²fí ë‹¤. covariateDataobjectsëŠ" ff벼 ì ¬ìŠ©í• ë¯€ì€ Rì~ ì ¼ë¡œ ì € ì €ìž¥ í•íŒ í•"œ ì ¬ìŠ©í• ì ^ ì ³"íˆë‹¤. ëŒ€ì‹ saveCovariateData () í• íŒ í•"œ ì ¬ìŠ©í• ì €í•¼í• œ ë‹¤.

```
saveCovariateData(covariateData, "covariates")
```

loadCovariateData() í• íŒ í•"œ ì ¬ìŠ©í• ì — í•"œ ì ì ¼ëŠ" ë°°ì ´í„°ë¥¼ ì¡°í•œì • ì ^ ìž^ë‹¤.

3.3.3 ê°„í•ì ì € ê³ ìœ í•ŽÂ° ì œê±°, ì •ê · œí•™ ë° ì¤'ë°µ ì œê±°

1ì ¸ì¹ ê³ ìœ í•ŽÂ°ì „ í•í•±ì „"œ ì ˜œ ìœ˜ ì • í•"œ ê'ê³ ì í•™ìŠµì—ì"œ ê³ ìœ í•ŽÂ°ì „ ì ¬ìŠ©í•˜ê° ë°¸ë¬¬ì €í• í˜ ë‹¤. ì ˜ ê²½ìš° í•í•¼ì„ ë°°ì ´í„°ë¥¼ ì •ì¦í• ì • ì ^ ìž^ë‹¤.

tidyCovariateData í• íŒ"œ ì„ê°°ì§€€ ìž'ì—…ì" ì • ì ^ ìž^ë‹¤.

1. **ê°„í•ì ì € ê³ ìœ í•ŽÂ° ì œê±°** : ì¢…ì¢… ëŒ€ë¶€ë¶„ì "œ íŠ¹ì§•ì„"ê°°ì € ì½"í•"œ ì € í•œ ê°°ìˆ œ ì í"¼ì<í—ìž€ì¦í•"œ ë° ëŠ" ë"ë¤ì " í"¼ì<í—ìž€— ëŒ€í•˜ ´ 0ì ´ ì •ë‹^ Œ ê°°ì§€€,ë"ë¤ìœ ì œê±°í• ˜ ì©´ ì·±ëŠ¥™ í•Œí•¥ ì œê±°<ë‹¤.

2. **ì •ê·œí•™** : ë°"ë• ê³ ìœ í•ŽÂ° ê°°ì „ (ê° ê³ ìœ í•ŽÂ°ì ~ ì¼µœ ê°°ì ° ìœ¼¼ë¡œ ë…ˆ„–¬ìœ„) 0ê³¼ 1 ì— ì '¸ ê°°"ìˆ ¼œ œ ì¡° í• í• œë‹¤.

3. ì¤'ë³µ ì œê±° : ì½"í˜¸íŠ¸ì˜ ë"¨ë" ì,¬ëžŒì'´ ê³µê³€ëŸ‰ì— ëŒ€í•´´ ë™ì¼í•œ ꪇ— ëŠ" ê²½ìš°(ì˜ˆ: ì—¬ì"±ì— ê²Œ ì œí•œëœ œ ì½"í˜¸íŠ¸ëŠ" ëª¨ë"ì— ëŒ€í•´´ ë™ì¼í•œ ì"±ë³"ꪇ³µê³€ëŸ‰ê°'ì"¬ ê°'ì§)ì— ëŠ" ê³µê³€ëŸ‰ì´ ì¤'ë³µ œë‹¤. ì¤'ë³µ œ ê³µê³€ëŸ‰ì€ ì¼ë¶€ ê¸°°ê³,í•™Šµ ì•Œê³ ë¦¬ì¦˜ì— ë¬¸ì œë¥¼ ìœ ë°œìœ¼ë¡œ ì˜¥ë¥¼ ë"¤ì–´´ ë‹¬ì˜^ìœ íŠ©ê · 궄,ì, ì— ì,œ ë"ì 'ìf ë‹ì¼½˜ ˜ëµì ´ ˜ì§€€ ìŠ€Š"ë‹¤. ꧈ì˜°ê³€ì§€€ë§œ 공곀ëŸ‰ ê·,경ô£¹ë", ì¤'ë³µë ì^³ ìž̂ë‹¤.(ì˜^: ëª¨ë" ì,¬ëžŒì€€ ì ì´ë"¨ í•œ ì—ºì ¹ ê·,경ô£¹ì— ì† í• ê²fi ´ê³ ,í•˜ë,˜ì ˜ ê·,경ô£¹ì€€ ë'¤ëŸ¥ ê·,경ô£¹ì ˜ ë¶€ì¦¬žëœ¡¡œ ì•ì ˜ ì˜^ ìž̂ì¼ë̂´ëŒ€ê¡œ ì¤'ë³µ œë‹¤.)

tidyCovariates <-tidyCovariateData(covariateData,minFraction = 0.001, normalize = TRUE, removeRedundancy = TRUE)

ëª‡ ê°œì˜˜ ê³,í—ì ì ¸ 공곀ëŸ‰ì ´ ì œê±°°ë "ì—"ëŠ"ì§€€ ì•Œê³ ì‹¶¶ë‹¤ë©´ì´´ metaData ꪰ ì²´ë¥¼ ì°°ì©©í´ ë œë‹¤

deletedCovariateIds <- tidyCovariates$metaData$deletedInfrequentCovariateIds
head(deletedCovariateIds)

```
## ê²°ê³¼
## [1]        3 22274210 22274212 22288210 22340212 22350210
```

ë¹„ì¤ëžìí ˜ê²Œ,ì¤'ë³µë ꪵ곀ëŸ‰ì´ ì œê±°°ë ˜ì—"ëŠ"ì§€€ ì•Œê³ ì‹¶¶ì§€œ© metaData ꪰ ì²´ë¥¼ ì°°ì©©í´ ë œë‹¤.

deletedCovariateIds <- tidyCovariates$metaData$deletedRedundantCovariateIds
head(deletedCovariateIds)

```
## ê²°ê³¼
##       covariateId                        covariateName analysisId conceptId
## 1        8527004                        race = White           4       8527
## 2        8532001                     gender = FEMALE           1       8532
## 3        2015006                    index year: 2015           6          0
## 4           1007                      index month: 1           7          0
## 5   900000010802  ...tance Abuse Coverage Indicator         802  900000010
## 6   21603933412  ...EUMATIC PRODUCTS, NON-STEROIDS        412   21603933
## 7   21603932412  ...ORY AND ANTIRHEUMATIC PRODUCTS        412   21603932
## 8   21603932410  ...ORY AND ANTIRHEUMATIC PRODUCTS        410   21603932
## 9    1118084412  ...s relative to index: celecoxib        412    1118084
## 10   1118084410  ...s relative to index: celecoxib        410    1118084
## 11  21603991410  ...days relative to index: Coxibs        410   21603991
## 12  21603931412  ...index: MUSCULO-SKELETAL SYSTEM        412   21603931
## 13  21603991412  ...days relative to index: Coxibs        412   21603991
## 14  21603931410  ...index: MUSCULO-SKELETAL SYSTEM        410   21603931
## 15  21603933410  ...EUMATIC PRODUCTS, NON-STEROIDS        410   21603933
## 16  21603933413  ...EUMATIC PRODUCTS, NON-STEROIDS        413   21603933
## 17  21603932413  ...ORY AND ANTIRHEUMATIC PRODUCTS        413   21603932
## 18   1118084413  ...s relative to index: celecoxib        413    1118084
## 19  21603931413  ...index: MUSCULO-SKELETAL SYSTEM        413   21603931
## 20  21603991413  ...days relative to index: Coxibs        413   21603991
```

3.4 ì½"í˜¸íŠ¸ì— ëŒ€í•œ ì§'ê³„ ê³µë³€ëŸ‰ ìf ì„±

ì¢…ì¢… ìŠ°ë¦¬ëŠ" 1ì ë¹' ê³µë³€ëŸ‰ì " ë§Œë"¤ í•,ìŠ"ëŠ" ì—†ì§€§Œ, ëŒ€ì‹ ì— ìŠ°ë¦¬ëŠ" í†µí•• í©í•µê³,ì "ë—†ì "ì ˜ë" ,ì— ìž̂ë‹¤. ì˜^ë¥¼ ë"¤ì––, ìŠ°ë¦¬ëŠ" ì-'ë–¤ ì,¬ëžŒì ´ ë,"ì œ ì §€§ ì• ꪰ í•,ìŠ"ê²€ ì—†ì " ì˜^ë " ,ìž̂ê²Œ§Œ, ì½"í˜¸íŠ¸ì ˜ ë¹',ìœ ì ´ ë,"ì œ ì §€§ ì• ꪰ³ ì¦ì î̂˜ë " ,ìž̂ë‹¤.

ë°"ì— ì½"ë“œ¡œ 1ì ë¹' 공변ëŸ‰ì "ì§'ê³„,í• ì˜^ ìž̂ë‹¤.

covariateData2 <-aggregateCovariates(covariateData)

ë¬¼ë¡ ìš°ë¦¬ê°€ ì›í•˜ëŠ" ëª"ë" ꪃ´ ì§'ê³„ë œ í†µê³„ì˜€ë‹¤ë©´, ìf ì„± ꪼì •ë—ì˜œ ꪷ ¸ê²ffë"¤ ¸
ìœ…í•©í•˜�¸ëŠ"ê²Œ ë " íš"ìœ"ì ì¼ ê²ff ´ë‹¤.

```
covariateSettings <-createDefaultCovariateSettings()
covariateData2 <-getDbCovariateData(connectionDetails = connectionS,
                                    cdmDatabaseSchema = cdmDatabase
                                    cohortDatabaseSchema = resultsD
                                    cohortTable = "cohorts_of_inter
                                    cohortId = 1118084,
                                    covariateSettings = covariateSe
                                    aggregated = TRUE)
summary(covariateData2)

## ê²°ê³¼
## CovariateData object summary
##
## Number of covariates: 41330
## Number of non-zero covariate values: 41330
```

aggregated = TRUEë¡œ ì§€ì •í–ˆë‹¤. ì˜ í•´ ë " ì ´ìf personId ë˜°ì ´„°ë¥¼ ë°©ì§€ì•Šê¸° ë•Œë¬ì— ë "
ì ´ìf rowIdFieldë¥¼ ì •í•˜ ì •ìˆ"ê²€ ì—†ë‹¤.

3.4.1 ì§'ê³„ ê³µë³€ëŸ‰ ì¶œë ¥ í˜• ì‹

ì§'í•© ë "œ covariateData ê° ì²´ì ˜ ë " ê°€ì§€€ ì£¼ìš" ꪹ¬ì„± ìš"ì†ŒëŠ" ê° ê° ì ì§„ ë ì—°ì† ꪹ ë Y‰ì—
ëŒ€í•œ œ ꪹ ë Y‰ ë° ꪹ³ ì—°ì† ´ë‹¤.

```
covariateData2$covariates

## ê²°ê³¼
## ffdf (all open) dim=c(41326,3), dimorder=c(1,2) row.names=NULL
## ffdf virtual mapping
##              PhysicalName VirtualVmode PhysicalVmode AsIs VirtualIsMatrix PhysicalIsMatrix Physical
## covariateId        list..       double        double FALSE           FALSE            FALSE
## sumValue           list...1     double        double FALSE           FALSE            FALSE
## averageValue       list...2     double        double FALSE           FALSE            FALSE
## ffdf data
##         covariateId      sumValue averageValue
## 1     1.978072e+08 4.000000e+00 8.457911e-05
## 2     4.133018e+09 6.000000e+00 1.268687e-04
## 3     4.369452e+08 3.000000e+01 6.343433e-04
## 4     3.141032e+08 2.000000e+00 4.228956e-05
## 5     1.343122e+08 9.000000e+00 1.903030e-04
## 6     4.181339e+09 6.000000e+00 1.268687e-04
## 7     4.054827e+09 1.200000e+01 2.537373e-04
## 8     4.389632e+08 1.000000e+00 2.114478e-05
## :                :            :            :
## 41319 2.211822e+09 6.000000e+00 1.268687e-04
## 41320 2.211878e+09 2.000000e+00 4.228956e-05
## 41321 2.313614e+09 5.000000e+00 1.057239e-04
## 41322 2.313664e+09 1.000000e+00 2.114478e-05
## 41323 2.314111e+09 8.000000e+00 1.691582e-04
## 41324 2.514442e+09 2.370000e+02 5.011312e-03
## 41325 2.514467e+09 2.300000e+01 4.863299e-04
## 41326 2.721532e+09 1.000000e+00 2.114478e-05
```

```
covariateData2$covariatesContinuous
```

```
## ê²°ê³¼
## ffdf (all open) dim=c(4,11), dimorder=c(1,2) row.names=NULL
## ffdf virtual mapping
##                   PhysicalName VirtualVmode PhysicalVmode AsIs VirtualIsMatrix PhysicalIsMatrix Phys:
## covariateId               list..        double        double FALSE           FALSE            FALSE
## countValue                list...1      double        double FALSE           FALSE            FALSE
## minValue                  list...2      double        double FALSE           FALSE            FALSE
## maxValue                  list...3      double        double FALSE           FALSE            FALSE
## averageValue              list...4      double        double FALSE           FALSE            FALSE
## standardDeviation         list...5      double        double FALSE           FALSE            FALSE
## medianValue               list...6      double        double FALSE           FALSE            FALSE
## p10Value                  list...7      double        double FALSE           FALSE            FALSE
## p25Value                  list...8      double        double FALSE           FALSE            FALSE
## p75Value                  list...9      double        double FALSE           FALSE            FALSE
## p90Value                  list...10     double        double FALSE           FALSE            FALSE
## ffdf data
##     covariateId    countValue      minValue      maxValue averageValue standardDeviation   medianValue
## 1   1901.000000 33013.000000      0.000000     21.000000     2.321168          2.657008      1.000000
## 2   1904.000000 31301.000000      0.000000      6.000000     1.307128          1.144543      1.000000
## 3   1902.000000 27951.000000      0.000000     13.000000     2.646671          2.723938      1.000000
## 4   1903.000000 30898.000000      0.000000      5.000000     1.234157          1.055174      1.000000
```

covariatesì ˜ ì»¬ëŸ¼ì € ë‹¤ì Œê³¼ ê°™ì ´ ì •ì˜ë œë‹¤

- covariateIdëŠ” ê³µë³€ëŸ‰ì „ ì‹ ë³„í•˜ëŠ”©°, ê³µë³€ëŸ‰ì ˜ ì •ì˜ëŠ” cohortData$covariateRef ê° ì²´ì—
  ì„œ ì³¾ì „ ì^˜ ìž^ë‹¤
- sumValueëŠ” ê³µë³€ëŸ‰ ê°'ì ˜ í•©©ì ´ë‹¤.   ì ´ê²fë“¤ì € ì ì§„ íŠ¹ì„±ë"¤ì ´ê¸° ë•Œë¬¸ì—,
  ê³µë³€ëŸ‰ ê°'ì ´ 1ì „ ì,¬ëžŒêµ"ì ˜ì^ˆ^ìTM€ ê°ºTMë‹¤.
- averageValueëŠ” í‰ê· ê³µë³€ëŸ‰ ê°'ì ´ë‹¤.   ì ´ê²fë“¤ì € ì ì§„ íŠ¹ì„±ë"¤ì ´ê¸° ë•Œë¬¸ì—,
  ê³µë³€ëŸ‰ ê°'ì ´ 1ì „ ì,¬ëžŒêµ"ì ˜ ë¹„ìœ¨ê³¼ ê°ºTMì Šµë‹^ë‹¤

covariatesContinuousì ˜ ì»¬ëŸ¼ì € ë‹¤ì Œê³¼ ê°™ì ´ ì •ì˜ë œë‹¤

- covariateIdëŠ” ê³µë³€ëŸ‰ì „ ì‹ ë³„í•˜ëŠ”©°, ê³µë³€ëŸ‰ì ˜ ì •ì˜ëŠ” cohortData$covariateRef ê° ì²´ì—
  ì„œ ì³¾ì „ ì^˜ ìž^ë‹¤
- countValueëŠ” ê°'ì ´ ìž^ëŠ” ì,¬ëžŒì ˜ ì^ˆìž…ë‹¤(ì—°ì†† ë œ ì²´ì^^ì ¼ ê²½ìš°)
- minValue,maxValue,averageValue,standardDeviation,medianValue,p10Value,p25Value,p75Value,p90Value
  ë"¤ì € ê³µë³€ëŸ‰ ê°'ì ˜ ë¶„í¬ë¥¼ ì•Œê²Œ í•´ì¤€ë‹¤. ì ¼ë§€ ê³µë³€ëŸ‰(ì^: Charlson comor-
  bitity index)ì ˜ ê²½ìš° 0 ê°'ì € 0ìœ¼ë¡œ í•´ì„, ë ˜ì§€ê³Œ ë¶„ë¥¸ ê³µë³€ëŸ‰(ì^: í~ˆì••)ì ˜
  ê²½ìš° 0ì € ë^,ë ¼ê²œ ê°°ì¼ìœ¼ë¡œ í•´ì„, ë ˜ì©° ë¶„í~ ì†µê³³,ëŠ” ë¹,ë",ë ½ ê°°ì— ë Œ€í• ì,œœ§Œ
  ê³,ì,ºë œëŠ"ëŠ" ì ì "ìœ ì•Œ í•∙í•¼ í•œë‹¤. ì—´ë-¤ ì—°ì†† ê³µë³€ëŸ‰ì ´ ì—´ë-¤ ë²²"ì£¼ì—
  ì† í•´ ìŠ"ì§€ ì• Œì•"ë³"ë ¤ë©´ covariateData$analysisRef ê° ì²´ì ˜ missingMeansZero í•"ë"œœ¥¼
  ì°,ì¡°í•˜©´ ë œë‹¤.

3.5 í…Œì ´ë¸" ë§Œë“¤ê¸° 1

íŠ¹ì§•ì¶"ì¶œ íŒ¨í‚¤ì§€ê³€ ì§€ì› í•˜ëŠ" í•œ ê°€ì§€ ê³¼ì ê¾€ì œëŠ" ë…¼ë¬¸ì— í ¬í•¨ë³€ì^ ì˜ì ^ ìž^ëŠ" ì •ì¼ í•™ì´Šµ ì,ë¡ ìŠ"ì„± í…Œì ´ë¸ì „ ë§Œë"œ°ëŠ"ê²fì ´ë‹¤. ì ´ê²fì € ì ¼ë°˜~ì ìœ¼ë¡œ í…Œì ´ë¸ì ¼ í ¼ë¶^ í•©ê°"ë^ì•⁴ë ¤ì— í…Œì ´ë¸"1ì ´ë¼ê³ í•©ë³€ì"ë‹¤. íŠ¹ì§•ì¶"ì¶œ íŒ¨í‚¤ì§€ì—— ì,œ ê,ºë³¸ í…Œì ´ë¸" 1ì „ ì,¬ìš© í•´ ì ^ ìž^ë‹¤.

```
result <-createTable1(covariateData2)
print(result, row.names = FALSE, right = FALSE

## ê²°ê³¼ëŠ" ë°'ì- ì,¬ì§„
```

Characteristic	% (n = 47,293)	Characteristic	% (n = 47,293)
Age group		Ischemic heart disease	7.0
00-04	0.1	Peripheral vascular disease	13.3
05-09	0.2	Pulmonary embolism	1.2
10-14	0.7	Venous thrombosis	3.2
15-19	2.4	Medical history: Neoplasms	
20-24	2.0	Hematologic neoplasm	1.2
25-29	3.8	Malignant lymphoma	0.4
30-34	5.8	Malignant neoplasm of anorectum	0.2
35-39	7.1	Malignant neoplastic disease	6.3
40-44	9.2	Malignant tumor of breast	1.7
45-49	11.5	Malignant tumor of colon	0.4
50-54	14.2	Malignant tumor of lung	0.4
55-59	13.0	Malignant tumor of urinary bladder	0.2
60-64	11.4	Primary malignant neoplasm of prostate	0.3
65-69	5.6	Medication use	
70-74	4.3	Agents acting on the renin-angiotensin system	47.0
75-79	3.4	Antibacterials for systemic use	82.6
80-84	3.0	Antidepressants	63.9
85-89	2.2	Antiepileptics	77.1
90-94	0.1	Antiinflammatory and antirheumatic products	100.0
Gender: female	72.4	Antineoplastic agents	27.4
Race		Antipsoriatics	9.2
race = Black or African American	18.6	Antithrombotic agents	21.7
race = Other Race	18.9	Beta blocking agents	54.0
race = White	62.4	Calcium channel blockers	47.8
Ethnicity		Diuretics	62.6
ethnicity = Hispanic or Latino	1.5	Drugs for acid related disorders	83.2
ethnicity = Not Hispanic or Latino	17.4	Drugs for obstructive airway diseases	59.6
Medical history: General		Drugs used in diabetes	31.3
Acute respiratory disease	37.0	Immunosuppressants	3.7
Attention deficit hyperactivity disorder	2.4	Lipid modifying agents	55.0

Characteristic	Value	Characteristic	Value
Chronic liver disease	5.0	Opioids	79.8
Chronic obstructive lung disease	19.1	Psycholeptics	85.6
Crohn's disease	0.6	Psychostimulants, agents used for adhd and nootropics	77.7
Dementia	2.0		
Depressive disorder	34.4	Characteristic	Value
Diabetes mellitus	25.2	Charlson comorbidity index	
Gastroesophageal reflux disease	25.1	Mean	2.3
Gastrointestinal hemorrhage	4.3	Std. deviation	2.7
Human immunodeficiency virus infection	0.7	Minimum	0.0
Hyperlipidemia	35.8	25th percentile	0.0
Hypertensive disorder	50.4	Median	1.0
Lesion of liver	0.7	75th percentile	3.0
Obesity	17.2	Maximum	21.0
Osteoarthritis	44.9	CHADS2Vasc	
Pneumonia	6.2	Mean	1.3
Psoriasis	1.0	Std. deviation	1.1
Renal impairment	4.6	Minimum	0.0
Rheumatoid arthritis	4.7	25th percentile	0.0
Schizophrenia	3.5	Median	1.0
Ulcerative colitis	0.4	75th percentile	2.0
Urinary tract infectious disease	15.2	Maximum	6.0
Viral hepatitis C	3.6	DCSI	
Visual system disorder	35.7	Mean	2.6
Medical history: Cardiovascular disease		Std. deviation	2.7
Atrial fibrillation	2.6	Minimum	0.0
Cerebrovascular disease	5.7	25th percentile	0.0
Coronary arteriosclerosis	9.8	Median	1.0
Heart disease	26.2	75th percentile	5.0
Heart failure	6.9	Maximum	13.0

ì ìš© ê²€ìŠ¥í•œ ê²½ì½, ì ´ëŸ¬í•œ íŠ¹,±ì € ê¸´ ì‹œê°"ëŒ€ìˆ€ ê´ë "ë œ ë¶„ì„ ì— ë „ì¶œëì ˜ë©°, ì½"î~¸íŠ¸ ì‹œœìž' ë, ì§œë¶í•í,° 365ì ¼ê¹Eì§€ì ~ ê´€ì°°ë œ ê°œë… ì „ í ¬í• ¨í-ˆ‹¤. createTable1 í•¨ì^¨ëŠ" í…Œì´ë¸"ì— í ¬í• ¨í• ë³Eì^¨ë¥¼ ê°"ë¨í• ꬃ ì§€ì •í• ì•¼ í• ¼ í•œœë‹¤. í¯,¤ì§€ì— í ¬í• ¨ ë œ ê¸ë³ìŠ¤íŽ™ì € getDefaultTable1Specifications í• ¨ì^¨ë¥¼ í^¸ì¶œí• ˜—¬ ë³¼ ì^¸ìž^ë‹¤. ìŠ¤íŽ™ì î°¸ì¼° ꬄ ì, ID ì™€ ꬵê³€ëŸ‰ ID ë° ꬀¸ì ꬵ꤀ ꬵꟀ IDꥀ°¨ ꬀¸ ꬵꟀꟀ Ꟁ°Ꟁ ì,¤ì •ì ~ IDꥀ Ꟁ°Ꟁí•œëë‹¤. ëŒ€ì‹ ì— ì,¬ìš©ìž ꬵꟀ • í…Œì´ë¸" 1 ìŠ¤íŽ™ì „ ꬵꟀë"¤ì-"¸ì,¬ìš©í• ì^¨ ìž^ë‹¤.

ì—¬ê¸°ì„œ í…Œì´ë¸"ë¬ ꬵꟀ€ëŸ‰ì ~ ìž'ì €ë¶ë¶¶ ꬵꟀ ì,¬ìš© ë ë "ë ¼ë "Ꟁ€¸ ꬵꟀꟀ ꬵꟀꟀ, í¯,¤í•¼ë"Š" covariateData ê° ì²´í— ëŒ€í•• í…Œì´ë¸" 1ì "¸ ꬵꟀ~ìœ¼ë¡œ í•œëë‹¤. í…Œì´ë¸"ì— í•ìš"í•œë ꬵꟀ€ëŸ‰ ꬵꟀ ꬵꟀꬵꟀ~ë Ꟁ¤Ꟁ€½ìš° ê²½ì²½ì° createTable1CovariateSettings í•¨ì^¨ë¥¼ ì,¬ìš©í• ì^¨ ìž^ë‹¤.

```
covariateSettings <-createTable1CovariateSettings()
covariateData2b <-getDbCovariateData(connectionDetails = connectionDetails,
                                      cdmDatabaseSchema = cdmDatabase
                                      cohortDatabaseSchema = resultsh
```

```
                                                          cohortTable = "cohorts_of_inter
                                                          cohortId = 1118084,
                                                          covariateSettings = covariateSe
                                                          ;aggregated = TRUE)
```

```
summary(covariateData2b)
```

```
## ê²°ê³¼
## CovariateData object summary##
## Number of covariates: 90
## Number of non-zero covariate values: 90
```

4. ë‘ ê°œì˜ ì½”í˜¸íŠ¸ ë¹„êµ

íŠ¹ì§•ì•¶ì¶¶ìœœ íŒ¨í„¤ì§€ê°€ ì§€ì› í•˜ëŠ” ë˜ ë‹¤ë¥¸ ê¸°ëŠ¥ì€ ë‘ ê°œì§€ ê´€ì‹¬ ì½”í˜¸íŠ¸ë¥¼ ë¹„êµ í•˜ëŠ”
ê²ƒì…´ ´ë‹¤.

ê¸°ëŠ³ í…Œì´ ´ë¸” 1ì— í¬í•¨ë˜ì´ ê³ì±¶ì´—ë§Œ ë‘ ê°œì§€ ì½”í˜¸íŠ¸ë¥¼ ë¹„êµ í•œ ¤ëŠ” ê²½½ìš°ë¥¼
ê°€ì •í•´ ´ë³´ìž

```
settings <-createTable1CovariateSettings(excludedCovariateConceptIds =c(1118084, 1124300),
                                              addDescendantsToExclude
```

```
covCelecoxib <-getDbCovariateData(connectionDetails = connectionDetails,
                                          cdmDatabaseSchema = cdmDatabaseScher
                                          cohortDatabaseSchema = resultsDataba
                                          cohortTable = "cohorts_of_interest",
                                          cohortId = 1118084,
                                          covariateSettings = settings,
                                          aggregated = TRUE)
```

```
covDiclofenac <-getDbCovariateData(connectionDetails = connectionDetails,
                                          cdmDatabaseSchema = cdmDatabaseSche
                                          cohortTable = "cohorts_of_interest"
                                          covariateSettings = settings,
                                          aggregated = TRUE)
```

```
std <-computeStandardizedDifference(covCelecoxib, covDiclofenac)
```

ìœ„„ì ˜ì˜ì— ì„œëŠ” celecoxib (1118084)ì™€ diclofenac (1124300)ì ˜ ë‘ ê°€ì§€ ì½”í˜¸íŠ¸ë¥¼ ì •ì˜í•˜ëŠ"
ë° ì,¬ìš© ë œ ë‘ ê°€ì§€ ê°œ… ì—ì„œ íŒŒì if ë œ ê³µë³€ëŸ‰ì „ ì œì™í• ê¸°ìœ í• ì˜ì—. ë˜ë˜ ì§€
ê³µë³€ëŸ‰ ê°„ì ˜ í‘œì¤€í™" ë œ ì°¨ì ë¥¼ ê³„ì‚°í• ìˆˆì«ë‹¤.

```
head(std)
```

```
##      covariateId    mean1       sd1       mean2       sd2        sd        stdDiff                    covaria
## 69 21601387410 0.27393060 0.5233893 1.0000000 1.0000021 1.1286897  0.6432852 ...o index: ANTINEOPLAST
## 74 21601853410 0.54965851 0.7413974 1.0000000 1.0000021 1.2448591  0.3617610 ... index: LIPID MODIFYI
## 901904 1.30712790 1.1445434 0.9100477 1.0777444 1.5721045 -0.2525787                    CHADS2VASc
## 43003 0.02404161 0.1550552 0.1183165 0.3439724 0.3773051  0.2498638                    age group: 15
## 2480180210 0.44879792 0.6699309 0.2619693 0.5118305 0.8430764 -0.2216034         ...ative to index: Os
## 891902 2.64667075 2.7239376 1.8278085 2.6827256 3.8231993 -0.2141825         ...orbidity Severity Inc
```

stdDiff ì—´ì— ëŠ" í‘œì¤€í™" ë œ ì°¨ì ê°€ ìž^ìŠµë‹ˆ^ë‹¤. ê¸°ë³³ì ìœ¼¼ë¡œ ì´„°ëŠ" í‘œì¤€í™"ë œ ì°¨ì ´ˆ
ì ˆŒ€ ê°"ì „ ê¸°ì¤€ìœ¼¼ë¡œ ë,´ë¦¼ì°¨ìˆœìœ¼¼ë¡œ ì •ë ¬ ˜ë©, ê°€ìž¥ í ˆì˜ ´ê°€ ìž^ëŠ" ê³µë³€ëŸ‰ë‹ ´
ë¨¼ì € í‘œì‹œ ë©ë‹ˆ^ë‹¤.

ìš°ë¦¬ëŠ" ë˜ í•¬ ë‹„êµ ë¥¼ í‘œì¤€ í…Œì´ ´ë¸" 1ë¡œ ë³´ì——ì„„ë‹ìˆ " ì^^ ìž^ë‹¤.

Characteristic	% (n = 47,293)	% (n = 240,761)	Std.Diff
Age group			
00-04	0.1	0.0	-0.01
05-09	0.2	0.2	0.00
10-14	0.7	2.5	0.11
15-19	2.4	11.8	0.25
20-24	2.0	5.8	0.14
25-29	3.8	8.8	0.14
30-34	5.8	10.1	0.11
35-39	7.1	10.0	0.07
40-44	9.2	9.2	0.00
45-49	11.5	9.4	-0.05
50-54	14.2	9.9	-0.09
55-59	13.0	8.7	-0.09
60-64	11.4	6.1	-0.13
65-69	5.6	2.2	-0.12
70-74	4.3	1.7	-0.11
75-79	3.4	1.3	-0.10
80-84	3.0	1.1	-0.10
85-89	2.2	0.9	-0.07
90-94	0.1	0.1	-0.01
Gender: female	72.4	72.8	0.00
Race			
race = Black or African American	18.6	28.1	0.14
race = Other Race	18.9	15.3	-0.06
race = White	62.4	56.6	-0.05
Ethnicity			
ethnicity = Hispanic or Latino	1.5	2.1	0.03
ethnicity = Not Hispanic or Latino	17.4	13.2	-0.08
Medical history: General			
Acute respiratory disease	37.0	39.4	0.03
Attention deficit hyperactivity disorder	2.4	4.4	0.07
Chronic liver disease	5.0	3.7	-0.05
Chronic obstructive lung disease	19.1	11.3	-0.14
Crohn's disease	0.6	0.5	-0.02
Dementia	2.0	1.0	-0.06
Depressive disorder	34.4	29.3	-0.06
Diabetes mellitus	25.2	17.4	-0.12
Gastroesophageal reflux disease	25.1	19.1	-0.09
Gastrointestinal hemorrhage	4.3	3.2	-0.04
Human immunodeficiency virus infection	0.7	0.8	0.01
Hyperlipidemia	35.8	25.2	-0.14
Hypertensive disorder	50.4	37.4	-0.14
Lesion of liver	0.7	0.5	-0.02
Obesity	17.2	16.6	-0.01
Osteoarthritis	44.9	26.2	-0.22
Pneumonia	6.2	4.0	-0.07
Psoriasis	1.0	0.8	-0.02
Renal impairment	4.6	3.3	-0.05
Rheumatoid arthritis	4.7	2.4	-0.09
Schizophrenia	3.5	2.5	-0.04
Ulcerative colitis	0.4	0.2	-0.02
Urinary tract infectious disease	15.2	16.0	0.01
Viral hepatitis C	3.6	2.6	-0.04
Visual system disorder	35.7	31.2	-0.06
Medical history: Cardiovascular disease			
Atrial fibrillation	2.6	1.4	-0.06
Cerebrovascular disease	5.7	3.7	-0.06
Coronary arteriosclerosis	9.8	5.5	-0.11
Heart disease	26.2	16.6	-0.15
Heart failure	6.9	4.1	-0.09

Characteristic	% (n = 47,293)	% (n = 240,761)	Std.Diff
Ischemic heart disease	7.0	4.0	-0.09
Peripheral vascular disease	13.3	7.8	-0.12
Pulmonary embolism	1.2	0.5	-0.05
Venous thrombosis	3.2	1.9	-0.06
Medical history: Neoplasms			
Hematologic neoplasm	1.2	0.6	-0.04
Malignant lymphoma	0.4	0.2	-0.02
Malignant neoplasm of anorectum	0.2	0.1	-0.02
Malignant neoplastic disease	6.3	3.7	-0.08
Malignant tumor of breast	1.7	1.0	-0.04
Malignant tumor of colon	0.4	0.2	-0.02
Malignant tumor of lung	0.4	0.2	-0.03
Malignant tumor of urinary bladder	0.2	0.1	-0.02
Primary malignant neoplasm of prostate	0.3	0.2	-0.02
Medication use			
Agents acting on the renin-angiotensin system	47.0	32.5	-0.16
Antibacterials for systemic use	82.6	100.0	0.13
Antidepressants	63.9	53.2	-0.10
Antiepileptics	77.1	73.9	-0.03
Antiinflammatory and antirheumatic products	100.0	100.0	0.00
Antineoplastic agents	27.4	100.0	0.64
Antipsoriatics	9.2	7.3	-0.05
Antithrombotic agents	21.7	11.4	-0.18
Beta blocking agents	54.0	39.6	-0.15
Calcium channel blockers	47.8	32.4	-0.17
Diuretics	62.6	47.8	-0.14
Drugs for acid related disorders	83.2	79.9	-0.03
Drugs for obstructive airway diseases	59.6	57.0	-0.02
Drugs used in diabetes	31.3	21.0	-0.14
Immunosuppressants	3.7	2.1	-0.07
Lipid modifying agents	55.0	100.0	0.36
Opioids	79.8	79.3	0.00
Psycholeptics	85.6	81.0	-0.04
Psychostimulants, agents used for adhd and nootropics	77.7	77.8	0.00

Characteristic	Value	Value	Std.Diff
Charlson comorbidity index			
Mean	2.3	1.6	-0.19
Std. deviation	2.7	2.5	
Minimum	0.0	0.0	
25th percentile	0.0	0.0	
Median	1.0	1.0	
75th percentile	3.0	2.0	
Maximum	21.0	27.0	
CHADS2Vasc			
Mean	1.3	0.9	-0.25
Std. deviation	1.1	1.1	
Minimum	0.0	0.0	
25th percentile	0.0	0.0	
Median	1.0	0.0	
75th percentile	2.0	1.0	
Maximum	6.0	6.0	
DCSI			
Mean	2.6	1.8	-0.21
Std. deviation	2.7	2.7	
Minimum	0.0	0.0	
25th percentile	0.0	0.0	
Median	1.0	0.0	
75th percentile	5.0	3.0	
Maximum	13.0	13.0	