Validation Report of Proposed Simplified Truveta Data Model (TDM) System

Overview

Patient mining in Truveta Data Model (TDM) core tables, requires multiple joins on multiple tables including core, map and concept tables. Joining these large tables is often at the expense of cluster resources, leading to delayed or slow execution sometimes, especially, when multiple users are performing similar tasks. Hence, the proposed simplified TDM data model system.

The function of the simplified TDM data model system is to create a one-time join of the multiple tables and store the simplified final table in the exploratory folder. Users can then, access the stored final table to perform script tasks. The advantages of developing such system are to minimize execution time, optimize cluster bandwidth and improve overall efficiency in scripts mining patients' information from Truveta core tables.

The goal of this validation process is to compare the proposed TDM simplified data model with the existing system of extracting patients' information from the TDM core tables. The table below summarize the findings.

Summary of Findings

Item	Existing TDM System	Proposed Simplified TDM System	Comments
Test Validation Script	%sql select count(distinct personid) as rep from cdh_truveta.condition@v15 c left join cdh_truveta.conditioncodeconceptmap@v15 m on c.codeconceptmapid=m.id left join cdh_truveta.concept@v15 cp on m.codeconceptid=cp.conceptid where conceptcode='U07.1'and recorddeddatetime>='2022-08-01' --Outcome=2,788,360	%sql select count(distinct PersonId) as rep from edav_prd_cdh.cdh_truveta_exploratory.ddp_conditions where ConditionConceptCode="U07.1" and RecordedDateTime>='2022-08-01' --Outcome=2,788,360	✓ Similar counts ✓ Proposed simplified TDM system is faster (1sec), compared to existing TDM system(3min) ✓ Simplified TDM system utilized lesser cluster resources
Columns:			
1	AbatementDateTime	AbatementDateTime	✓ Validated ✓ Similar (6540398) ✓ Retain Column
2	CategoryConceptId		✓ Retain Column ✓ Column may be useful in shedding more light on an encounter info
3	ClinicalStatusConceptId	ClinicalStatus	✓ validated ✓ Interpretated to ClinicalStatus ✓ Retain Column and its Conceptid

			✓ Retain its Conceptid for ease of reference in scripts
4	CodeConceptMapId	CodeConceptMapId	✓ Don't retain column ✓ Serves no useful purpose after it's been used for linkage
5	EncounterId	EncounterId	✓ Validated(14836327) ✓ Retain Column ✓ Encounterid links to encounters in other TDM tables
6	Id(from condition table)	ConditionId	✓ Validated ((15516280) ✓ ID from condition table ✓ Retain Column
7	OnsetDateTime	OnsetDateTime	✓ Validated:(7096510) ✓ Retain Column
8	PatientId	PatientId	✓ Validated:(15516280) ✓ Retain Column
9	PersonId	PersonId	✓ Validated:(15516280) ✓ Retain Column
10	PrimaryDiagnosisConceptId	PrimaryDiagnosisFlag	✓ Recoded into 1=primary and 0= Not primary ✓ Retain Column
11	RecordedDateTime	RecordedDateTime	✓ Validated:(15516280) ✓ Retain Column
12	SequenceNumber	SequenceNumber	✓ Validated:(12930816) ✓ Retain Column
13	SeverityConceptId	Severity	✓ Interpreted to Severity, ✓ Retain Column and its Conceptid ✓ Retain its Conceptid for ease of reference in scripts
14	SourceProvenanceConceptId		✓ Don't Retain Column ✓ No valuable info

15	SourceRecordId		✓ Don't Retain Column ✓ No valuable info
16	VerificationStatusConceptId	VerificationStatus	✓ Interpreted to verificationStatus ✓ Retain Column and its Conceptid ✓ Retain its Conceptid for ease of reference in scripts
17	BodySiteMapId		✓ Don't Retain Column ✓ No valuable info
18	Id(from map table)		✓ Don't Retain Column ✓ No valuable info ✓ Serves no useful purpose after it's been used for linkage
19	CodeConceptId	CodeConceptId	✓ Validated (15516280) ✓ Like ConceptId ✓ Retain Column
20	SourceConceptId		✓ Don't Retain Column ✓ No valuable info
21	ConceptId		✓ Don't Retain Column ✓ Redundant ✓ Same as CodeConceptID
22	ConceptName	ConditionConceptName	✓ Interpreted to ConditionConcept Name ✓ Retain Column ✓ Provides condition description
23	ConceptDefinition		✓ Don't Retain Column ✓ ConditionConcept Name(22) above is already providing sufficient description

			✓ This may be redundant
24	CodeSystem	ConditionCodeSystem	✓ Renamed to ConditionCodeSystem ✓ Validated (15516280) ✓ Relevant for code taxonomy ✓ Retain Column
25	ConceptCode	ConditionConceptCode	✓ Renamed to ConditionConceptCode ✓ Validated (15516280) ✓ Retain Column
26	Domain		✓ Retain Column ✓ May provide valuable info about what core table to use for patient search when combined with ConditionCodeSystem (24) above.
27	ConceptClass		✓ Don't Retain Column ✓ No significant valuable info
28	NA	ConditionOccurrence	✓ id generated for each record in given patient panel ✓ Not available as original TDM column ✓ Validated ✓ Retain Column

Conclusion

In summary, the existing system and the proposed simplified TDM system produced similar counts. However, for increased efficiency, faster execution and cluster resource optimization, the proposed simplified TDM is a right step in right direction. Nonetheless, not all the columns

resulting from the join should be included in the final table. Some columns are either redundant or does not contribute to a meaningful or significant understanding of the mined dataset. Therefore, such columns may be removed. It is my opinion that the columns that should be considered for inclusion in the final simplified TDM data are: AbatementDateTime, CategoryConceptId, ClinicalStatusConceptId, ClinicalStatus, EncounterId, ConditionId, OnsetDateTime, PatientId, PersonId,PrimaryDiagnosisFlag, RecordedDateTime, SequenceNumber, SeverityConceptId, Severity,VerificationStatusConceptId, VerificationStatus, CodeConceptId, ConditionConceptName, ConditionCodeSystem,ConditionConceptCode,Domain and ConditionOccurrence .

Highlighted columns are the recommended additions to the already selected columns