

Release Notes

eMERGE Network Phase III Clinical Sequencing: eMERGEseq Panel Clinical Data in OMOP

Last Updated: 5/21/2026

dbGaP Study Page:

https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001616.v2.p2

OMOP Transformations

- The person table only contains those who had a year_of_birth value and if race, gender, or ethnicity information were missing, concept_id is set to 0.
- Each participant is assigned an “estimated date of birth” in the person table: 15 June of the person’s year_of_birth. For example, a participant with a year of birth of 1970 would have a date of birth of “1970-06-15”. This is in alignment with the OMOP Convention: https://ohdsi.github.io/Themis/missing_birthdatetime.html
- When an “age at event” is available in the eMERGE data, eg, for longitudinal BMI measures, dates for OMOP tables are calculated as the “estimated date of birth” plus the age. For example, a person with an “estimated date of birth” of 1970-06-15 and a BMI at age 40 would have that record at 2010-06-15. Decimal age at events are used where possible, eg, for that same person a BMI at age 40.03 would have an event date of 2010-06-26.
- When no age at observation is available in the eMERGE data (eg, case/control data, specimen data), the date (specimen_date, condition_date etc.) was set to 2020-01-01 as an estimated extraction date for the eMERGE data.
- Table assignment is based on which domain the OMOP concept_id belongs to (e.g., a concept code of 4245997 for Body Mass Index is classed as a Measurement so that eMERGE data goes into the measurement table)
- No pedigree information was mapped into OMOP even if it was provided in the input tables
- Data are segmented into different folders based on the information in the “Subject Consent” file (e.g., code 1 = General Research Use [GRU])
 - Participants with consent code = 0 are not available in the OMOP tables because they do not have complete demographic history and they do not appear in the demographics table along with consented participants.
- The observation table has two added columns at the end of the table, value_source_value and value_source_concept_id which are linked out to a custom concept table containing further information about the case/control data.
- Case/control data are provided as observations and no assertions of being “case” are put into condition_occurrence.
- Custom concept table with IDs 2002000000 - 2002000048

Input to Output Table Mappings

Input tables to OMOP:

- eMERGEseq_BMI_DS_20200925.csv
- eMERGEseq_IndicationsforSequencing_DS_20200925.csv
- eMERGEseq_CaseControl_DS_20200925.csv
- eMERGEseq_Phecode_DS_20200925.csv
- eMERGEseq_Demographics_DS_20200925.csv
- eMERGEseq_SampleAttributes_DS_20200925.csv
- eMERGEseq_SampleSubjectMapping_DS_20200925.csv
- eMERGEseq_SubjectConsent_DS_20201020.csv

Output tables ([OMOP v5.3](#)):

- care_site.csv
- concept.csv
- condition_occurrence.csv
- measurement.csv
- observation.csv
- person.csv
- specimen.csv

Vocabularies in Datasets

- ICD9CM
- ICD10CM
- SNOMED
- LOINC