**Summary of Steering Committee Meeting: February 2025**

February 12 - February 13, Zoom & In-Person

|  |  |
| --- | --- |
| **eMERGE Day 1: Wednesday, February 12, 2025** | |
| **Time** | **Event** |
| 9:00-9:20 AM | eMERGE progress and timelines | Rex Chisholm (SC Chair, Northwestern) |
| 9:20-9:45 AM | NHGRI Program Official Report | Robb Rowley (NIH/NHGRI) |
| 9:45-10:45 AM | Panel: Data Sharing & Utilization | Emma Perez (MGB) |
| 11:05 AM-12:20 PM | Panel: GIRA EHR Data | Josh Peterson (CC) |
| 1:00-1:20 PM | Broad multisample VCF updates and Invitae BAMs | Katie Larkin (Broad) & Sienna Aguilar (Invitae) |
| 1:20-2:50 PM | Working breakout session   * Data needs for upcoming manuscripts (Phenotyping, CARE, Outcomes) * Data quality and sharing (GRID & QA/QC) |
| 3:20-3:40 PM | ELSI: External data sharing | Ingrid Holm (BCH), Maya Sabatello (Columbia) & Anna Lewis (MGB) |
| 3:40-4:20 PM | Scientific priorities of eMERGE: Capturing additional data & utilization | Josh Peterson (CC) |
| 4:20-4:30 PM | Discussion and closing remarks | Rex Chisholm (SC Chair, Northwestern) |

|  |  |
| --- | --- |
| **eMERGE Day 2: Thursday, February 13, 2025** | |
| **Time** | **Event** |
| 9:00- 9:45 AM | Outcomes data plans: Unanswered questions and next steps | Nita Limdi (UAB) & Dave Veenstra (UW) |
| 9:45- 10:05 AM | HCP survey results | Ingrid Holm (BCH) and Georgia Wiesner (VUMC) |
| 10:25- 11:25 AM | Panel: Manuscript updates |
| 11:25 AM - 12:00 PM | Panel & discussion: Y6 Workgroups and major network milestones | Rex Chisholm (SC Chair, Northwestern) |
| 12:00 - 12:15 PM | Discussion & closing remarks | Rex Chisholm (SC Chair, Northwestern) |

|  |
| --- |
| **Other Links** |
| Action Items |
| Decisions Made |

**eMERGE Day 1: Wednesday, February 12, 2025**

1. **eMERGE progress and timelines | Rex Chisholm (SC Chair, Northwestern)** 
   1. The overall goal of the meeting relates to data status, capture, and utilization.
      1. Day 1 will focus on the progress on data tools & format as well as planned outcomes data capture. Day 2 will focus on the remaining barriers to network productivity, including publications.
   2. There have been several Network accomplishments since September 2024.
      1. 100% of the return goal has been achieved. The global IRB amendment that included provider interview & waiver of minor consent has been approved. QA/QC manual chart review testing has been completed and data analysis is ongoing. The Invitae reclassification flagging has been implemented in R4 and an SOP for network review has been established. The EHR data dictionary draft is ready for network review. This should allow us to be able to move forward with data pulls. The eMERGE study has been registered in dbGap.
   3. There are 27,048 Consented participants (not withdrawn).
      1. Of those, 24,012 are Active (not lost to follow up or missing sample) with 19,613 samples processed at Invitae (adults only) and 23,865 samples processed at The Broad.
      2. The expected total GIRA return is 23,889.
         1. Currently 23,426 GIRA are marked as returned in R4. It will be necessary for sites to reconcile Recruitment and Return numbers.
         2. Returned GIRA include 19,171 adults (81.8%) and 4,255 children (18.2%). 8,248 (35.2%) are male and 15,174 (64.8%) are female.
      3. 8,045 (34.3%) of returned GIRA are high-risk.
         1. The number of high risk GIRA is higher than expected.2.3% of high risk GIRA are due to monogenics, 36% are due to PRS, 23.4% are due to FHx, and 3.4% are due to BOADICEA.
   4. There are several network goals for the next 6 months (February 2025 - October 2025).
      1. Site Recruitment and Return numbers need to be reconciled in R4 with a target of February 2025.
      2. Current goals related to Outcomes analysis have a targeted completion by April 2025. Sign off on the final Outcomes framework is needed from condition leads. Sign off on the EHR data dictionary is needed from PIs. Data pull should be initiated by sites.
      3. Final multisample VCF & BAMs should be available on AnVIL by May 2025.
      4. EHR, CDP, and genomic data should be available on AnVIL for both network and external controlled access by July 2025.
      5. Assessment and QC of outcomes data and variables has a targeted completion of October 2025.
2. **NHGRI Program Official Report | Robb Rowley (NIH/NHGRI)** 
   1. Accomplishments
      1. The NIH wants to recognize and applaud the network to change how we practice healthcare. This network is accomplishing great things and we will overcome challenges.
   2. Changes at NHGRI
      1. Erin Ramos, Ph.D., M.P.H. Director of the Division of Genome Sciences
      2. Simona Volpi, Ph.D. Deputy Director of the Division of Genomic Medicine
   3. Future of eMERGE
      1. eMERGE One-Year Extension, FY25: NHGRI has received the supplement request for the extension with funds. Robb will be the point of contact until grants management is available. They are looking through the request and may send emails for clarification.
3. **Panel: Data Sharing & Utilization | Emma Perez (MGB)** 
   1. **Finalizing data location, format, and timing | Jodie Jackson (CC)** 
      1. The goal of the panel is to discuss data locations, format, and tools to ensure that the data are necessary and sufficient for network analysis needs.
      2. The network is primarily done collecting identified data, and is now focused on the non-human subjects designated data. The Clinical Data Platform is a non-human subjects designated version of the R4 Platform and will be pushed onto AnVIL.
      3. AnVIL is the platform for conducting the analysis and research on any of the eMERGE data. AnVIL will contain EHR, clinical (CDP), and genetic data of the current and legacy eMERGE datasets. AnVIL also allows for external controlled sharing and pairing of other data sets.
      4. Tanagra should be utilized for record counts and summary analysis. Tanagra primarily includes legacy eMERGE data (EHR) and current GIRA EHR data. In the future, limited GIRA clinical data can be added.
      5. i2b2 primarily includes CDP data from the current study. i2b2 may include EHR data in the future.
      6. eMERGE timeline:
         1. As of February 2025, the eMERGE IV study has been registered with dbGAP and some genetic data is available on AnVIL.
         2. April 2025: EHR 6 month data pull kick off
         3. May 2025: Multisample VCF on AnVIL
         4. June 2025: Final Invitae BAMs on AnVIL
         5. Year 6 focused on working with the data, data QC, and ensuring that the 12 month data pull is sufficient for analysis needs.
      7. A summary of the data types, locations, and timelines can be found here. The points of contact for each of the data types can be found here.
         1. The different eMERGE data types are located on AnVIL, Google Cloud Platform, and Tanagra.
         2. Tanagra (GCP) is a user interface for data exploration, i2b2 is a combination of user interface and tables, and AnVIL has data in a CSV format.
      8. Both i2b2 and Tanagra have user interfaces to query the data but i2b2 primarily has clinical data from the CDP while Tanagra primarily has the EHR data. In the future, some EHR data may be available on i2b2 and some high level GIRA data may be available on Tanagra but the timeline is unclear.
      9. The common variables are from the legacy datasets and are available on AnVIL and Tanagra. Additional datasets can be added to the internal eMERGE AnVIL but not necessarily the external controlled access which mirrors dbGAP.
      10. It is encouraged to do analysis on AnVIL if possible to demonstrate use cases for the platform. If there are issues with using the AnVIL for analysis, the NIH and CC should be informed of any barriers.
      11. A centralized version of the manual chart review data can be put onto AnVIL but will likely not be put onto the CDP. The manual chart review data will likely not be put on the CDP due to the number of free text fields and that it is a static dataset. It may be helpful to put the manual chart review data on i2b2.
      12. The 6 month EHR data pull is expected to be available to the network in July 2025 but will depend on the data quality submitted by the sites.
   2. **Tanagra | Alanna DiVietro (CC)** 
      1. Tanagra is being launched as the replacement for the eMERGE Record Counter that was decommissioned last year. The tool is very user friendly with drag and drop functionality and easy cohort visualization. Tanagra will be available to internal network members and will allow sorting and filtering by a variety of data elements. External access can also be granted to affiliates and collaborators but it will not have unrestricted public access.
      2. The Tanagra record counter application currently houses legacy and GIRA EHR data (demos, CPT & ICD, labs, etc.).
         1. The legacy data is common variable data from site EHRs collected on all eMERGE legacy sets.
            1. The last refresh was done in 2021. The legacy data includes eMERGEseq, GWAS, Exome Chip, PGRNseq, and WGS .
         2. The GIRA EHR data from the pilot data pull (Fall 2023) is also included.
            1. These data will be updated after the next GIRA refresh in Spring of 2025.
      3. Tanagra does not include specific dates but instead uses age at event.
      4. Since running into issues with the previous load onto Tanagra, the CC has added some additional QC checks:
         1. Rules and reference ranges around lab and med units, numeric and text value checks, minimum and maximums included in the data dictionary, more OMOP standardization.
      5. In the future, there are several additional features that will be explored for depending on need.
      6. There are some functionalities of Tanagra under consideration for the future:
         1. Currently we do not have eMERGE ID list export ability turned on but in the future may be able to enable that depending on individual permission type (internal investigator vs external).
         2. Tanagra is hosted on GCP, discussions on if/how to integrate with AnVIL are ongoing.
         3. Need to synergize with i2b2 tool to avoid duplication and highlight data uniqueness.
      7. A demonstration of the Tanagra user interface showed how users can create a cohort based on inclusion and exclusion criteria and breakdowns (bar graphs and tables) of the cohort. Cohorts are automatically saved and can be edited.
      8. Tanagra is open for beta testing and will be open to the network soon.
      9. The new eMERGE IV data dictionary was mapped to OMOP as much as possible to match the legacy data set. Not all data elements could be mapped to OMOP concepts.
      10. Mapping the monogenic data between eMERGE\_seq and eMERGE\_GIRA will consist of creating P/LP flags for specific genes.
      11. Adding the eMERGE IV GIRA data on Tanagra was necessary for the data to be complete and is not a heavy lift.
          1. Having the eMERGE IV data on Tanagra is redundant with having the data on i2b2 which has similar data querying abilities.
          2. It is possible to filter by eMERGE dataset in Tanagra.
          3. Putting the eMERGE IV and legacy EHR data on i2b2 has an unclear timeline while Tanagra will be ready more immediately to the network. The goal is to get the clinical data into i2b2 by summer 2025. Putting data on i2b2 is funded by a separate AnVIL Clinical Resource (ACR) grant.
      12. Tanagra is intended to be used as a record counter and visualization tool, not to export data. In the future, the ability to export eMERGE IDs from Tanagra for use in AnVIL should be available.
   3. **i2b2 | Jeff Klann (MGB)** 
      1. An i2b2 Portal was added to the cloud, with the Clinical Data Platform data available for users to interrogate the data. will be added to the data. By summer 2025, the GIRA Outcomes data should also be available on i2b2 along with the new data export feature.
      2. i2b2 is an approach for organizing and transforming person-oriented clinical data that is optimized for clinical and genomics research.
      3. i2b2 is a platform for clinical data warehousing and analytics. i2b2 has a graphical query tool, used for investigator-initiated data query portals, and allows export of custom analytic tables.
      4. The data are organized as ontologies, a concept dictionary that organizes the facts through hierarchies.
      5. The data are queried through the i2b2 webclient, a graphical query tool that enables advanced data exploration and cohort selection. The CDP data are organized by instrument. The i2b2 webclient can support record counts and breakdowns of simple analyses. Breakdowns can include customizable bar charts to show cohort breakdown by e.g., gender, race, and age.
      6. Data can be exported from i2b2 through a specified workflow allowing for further analysis.
         1. The ability to do custom analytics is being built in i2b2. Currently, exports can be done by domain but custom analytics will allow export by specific variables.
         2. Custom analytics can be performed by:
            1. Designing the data table, previewing the data table format, saving the data table, requesting the data export, managing the data export request.
      7. i2b2 will be ready for internal pilot testing by the end of February, with the goal to release to the network by the end of March.
      8. i2b2 stores all the backend data as one large table while OMOP stores data in domain specific tables. Custom analytics tables pivots the data into a wide Excel spreadsheet. It will be important that there are no differences between data extracted using the two methods.
      9. The goal of i2b2 is to provide cohort building functionality for multiple eMERGE GIRA data types.
      10. The data will include eMERGE GIRA Clinical Data Platform data and ICD outcomes data, as well as more GIRA EHR data in the future.
      11. i2b2 will be available to eMERGE members, with external access TBD.
      12. i2b2 will be hosted on Google Cloud Platform (GCP) at VUMC or MGB. There are plans for AnVIL integration in ~2026.
      13. Timeline:
          1. 2/28/25 - Internal pilot, working prototype, with a few eMerge users.
          2. 3/31/25 - Release to the network with refreshed data.
          3. 4/25/25 - Usability testing begins.
          4. 6/30/25 - Data export feature added.
          5. The CDP ontology will be revised and more outcomes data added during this time (3/25-6/25).
          6. With FISMA Moderate funding, a portal inside the Terra security boundary could be offered in the future (e.g., 2026).
      14. A demonstration of the i2b2 user interface showed how cohorts could be defined based on eMERGE IV GIRA data, review a breakdown of the cohort (bar graph or table), and utilize the data export tool.
      15. A process and timeline for providing feedback to the i2b2 team needs to be established.
      16. The source of the i2b2 data will be the 6 month and 12 month EHR data pulls, and will include referrals.
   4. **GRID analysis ready data set and format | Adam Gordon (NU) & Matt Lebo (MGB)** 
      1. CDP data is comprehensive and complex and some aspects of the R4 Portal do not automatically translate well to a ‘flat file’.
      2. The goal of the analysis-ready dataset is to improve usability of the CDP data in AnVIL for network and future users. Additional goals of the analysis ready dataset include integration with i2b2 and providing documentation for key variables.
      3. The GRID and QA/QC workgroups have been collaborating to create the analysis ready dataset. The GRID workgroup is focused on column-level changes (collapsing/renaming variables, recoding data). The QA/QC workgroup is focused on row-level changes (transforming values, identifying missingness).
      4. The process of creating the analysis-ready dataset includes:
         1. Review of all R4 instruments exported from CDP (completed),
         2. Compile recommendations for instrument-level changes (completed),
         3. Review of all variables exported from CDP,
         4. Compile recommendations for variable-level changes, Examples of variable level recommendations may include renaming and collapsing variables.
         5. Develop computational pipeline to implement these changes, the original ‘raw’ CDP export file will remain unchanged.
         6. Document changes and key variables for analysis.
      5. The new derived variable names should be used in i2b2 for ease of use.
      6. Creating a shared library of analysis scripts would be helpful to prevent redundant work and should be discussed further. In particular, standardized scripts to generate tables which will be used by multiple manuscripts (ex. demographics) would be helpful.
      7. Documentation of user requirements for the initial data collection (R4 Portal) would be helpful in future rounds of eMERGE.
      8. Partitioning the data into what can be standardized versus not and the ontologies that can be used could be helpful. There is ongoing work in AnVIL to map data onto standards.
      9. Ideally, the analysis-ready dataset will be utilized for analysis but the raw export of the CDP will always be available on AnVIL as well. The raw export of the CDP data is already available on AnVIL.
4. **Panel: GIRA EHR Data | Josh Peterson (CC)** 
   1. The goals of the panel are to understand EHR capture mechanisms, approaches, and progress of the QA/QC of the data. Additionally, the goals include identifying barriers to the site and network wide EHR data collection.
   2. **Manual Chart Review Findings | Lisa Martin (CCHMC) & Jen Pacheco (NU)** 
      1. The primary goal of chart review is to refine algorithms for extracting data from EHR for outcome analysis, reducing the need for manual data collection.
      2. An iterative process will be used in the analysis: initial keyword and algorithm development is followed by chart review, refining algorithms, and incorporating machine learning models for data extraction from encounter notes.
      3. A beta test was conducted across three sites, ensuring representation from pediatric and adult institutions and including one non-Epic site. Following beta test feedback, instructions and forms were refined and implemented in R4.
      4. R4 reports were created and shared to assist in selecting a broad range of participants. Subjects had to be active participants with at least six months since their RoR to ensure outcome availability. The majority of selected subjects were high-risk individuals, with some non-high-risk participants included as controls. Selection included at least one individual per condition and a mix of monogenic and polygenic conditions. Sites were encouraged to include subjects who had discussed results with their providers to increase the likelihood of finding relevant EHR data. This was selected by using post-RoR variables.
      5. As of today, 293 chart reviews have been completed across all sites.
      6. The chart review process examined not only outpatient encounters, but included MyChart messages and phone encounters.
      7. Of the 243 subjects fully reviewed so far:
         1. 70% had at least one order, encounter, or referral recorded in the EHR.
         2. 28% had a direct mention of the GIRA in their EHR records.
            1. ACTION ITEM: The QA/QC Task Force will break down the count and percentages of participants in the manual chart review who had the GIRA mentioned in their EHR by risk status.
      8. The average number of orders per subject was two, though there were variations by condition.
      9. Referrals were present for many conditions but were not expected for all.
      10. Reviewers were instructed to focus on encounters, referrals, and orders related to the condition of interest (e.g., oncology for breast cancer) and limit searches to two years before the RoR.
      11. The next step is to examine the variables of interest further, specifically where items were found, how they were coded and with which codes, and what keywords were found.
      12. Variability in EHR systems, even among institutions using Epic, posed challenges in data extraction.
      13. Subjects may receive healthcare outside the healthcare system. Sites had different policies regarding data access, such as restrictions on querying "Care Everywhere" data.
      14. Differences in how providers document conditions in problem lists and encounter notes led to inconsistencies in manual review.
      15. Some sites lacked sufficient resources for inter-rater reliability testing, limiting consistency checks.
      16. Lists of relevant ICD and CPT codes provided were found to be incomplete, requiring real-time adjustments during the review process. The forms and instructions asked reviewers to record codes found if the code was not listed.
      17. Lessons learned include that some sites were returning Invitae positive results prior to GIRA RoR, that it was difficult to distinguish between routine screening and testing due to GIRA findings, and that some sites add diagnosis codes to problem lists.
      18. The next step is to compare the results to the EHR data pulls. Comparisons must be done using the same timeline, with pulls based on either the last encounter or the order date recorded during chart review for each subject.
      19. For discrete data (e.g., checkboxes, yes/no), export must be done in a way that is easily comparable to the EHR data on a 1:1 basis. For free text, each site is responsible for extracting the data.
      20. The comparison process will start with a pilot comparison, reviewing data from only 10 sites. Sites will be responsible for comparing the original data.
      21. EHR algorithms would be updated based on the pilot review. The full comparison would be conducted following the pilot. The algorithms and keyword lists would be updated again using the full comparison results.
      22. There was a question about whether reviewers should have access to Care Everywhere data. If sites had backend access, they could query Care Everywhere, but some sites, like NU, could not. MGB is not allowed to use Care Everywhere data for research purposes for unconsented patients.
      23. The manual chart review forms collected what triggered the referral, order, or encounter, and had a question asking specifically if the GIRA was mentioned.
      24. Unless the provider explicitly mentioned the GIRA, we do not want to assume the GIRA was a factor in the order, referral, or encounter.
   3. **EHR data dictionary plan | Wei-Qi Wei (VUMC) & Shawn Murphy (MGB)** 
      1. Shawn and Wei-Qi presented on the EHR DD plan, focusing on refining methodologies for extracting and validating EHR data while ensuring consistency across all sites.
      2. Some outcomes were collected in the surveys, while others need to be collected from the EHR. There are many overlaps as well.
      3. It is important to define methodologies and measures to be standardized across sites in order. Some sites classify orders differently based on their internal processes. A structured format for defining and validating the variables should be established prior to the 6-month data pull. Codes used are likely different for every EMR, however, the order, procedure, and encounter likely will have a free text associated with it, which can be used to define the code at each site.
      4. It was proposed to develop a document mapping different terminologies across sites. It was also proposed to perform periodic audits to ensure definitions are consistently being applied.
      5. A placed order does not mean that it was completed. There are two key outcomes. Recommended action ordered (provider action) and recommended action received (patient action). Result is not a key outcome, but will be evaluated by some phenotypes. It also can be used as a proxy for the key outcomes.
      6. There are three levels of variables: Test ordered, test completed, and test result.
         1. If a result was found but not the order, the outcome would still be counted. The result is being used as a proxy for the patient's action. The reason for the inclusion of both results and orders for some conditions may be due to condition groups developing their outcomes list separately from each other. The order must be placed after RoR in order for it to be counter in outcomes analysis.
      7. There is a question of if there are multiple results over time, and what the best practice should be to handle these.
      8. The EHR data pull approach should be flexible enough to accommodate the differences between institutions while still being structured enough to allow for meaningful comparisons.
      9. eMERGE will capture all structured information.
         1. We will only capture medications after 2017 to determine any new or changed prescriptions. We will only capture certain labs in structured information after 2017. There are no time constraints on procedures being collected. We will examine the EHR for the order information and search keywords.
      10. Currently, the Phenotyping workgroup is collecting keywords from condition leads. The keyword list will be used to search for order and referral information.
      11. The Phenotyping workgroup had previously introduced the LLM plan, and will share how to install the closed source LLM on local servers.
      12. Clinical and phenotype experts are working together to update and finalize the procedure lists.
      13. ACTION ITEM: Following the conceptual review, condition leads, clinical experts, and phenotype experts should finalize all keywords, orders, and results in order to complete the EHR DD.
   4. **Site(s) Outcomes & referrals data | Kavi Wagholikar (MGB)** 
      1. The goal of this presentation is to discuss the different methods, and select one, to map site-specific terms to terms used in the outcomes analysis.
      2. Kavi diagrammed the life cycle of an order including a data trail, broken up into individual components.
      3. There are multiple methods to identify order names, including a manual review, keyword search, LLM, or code search. Any of these methods must be validated by manual chart review.
      4. There are almost 60,000 different free text entries for orders in a system.
      5. To perform a keyword approach, an inclusion list of keywords and an exclusion list of keywords is required.
      6. Kavi performed a trace back to examine the different texts used to describe a serum LDL cholesterol measurement. All sites should map order text to the lab code.
      7. eMERGE should link the vocabulary or ontology to the extracted data tables.
      8. Concerns were raised that the orderable lists would become unwieldy over time. The keyword approach is the easiest and fastest approach. It is estimated that the keyword approach would be sufficient 90% of the time.
      9. We should be able to use hierarchy information in loinc to assist in mapping.
      10. There was a suggestion to use a complementary approach involving using the broader repositories. For example, starting with asthma as a keyword in patients with asthma, and examining what codes are enriched in the asthma dataset versus not.
      11. It was previously discussed that the next step for this process would be to select two-to-three assessments of interest (mammograms, breast MRIs, and LDL measurement) to use as test cases to examine free text.
      12. ACTION ITEM: The phenotyping group will investigate the best approaches to pull accurate mappings for order keywords.
5. **Broad multisample VCF updates and Invitae BAMs | Katie Larkin (Broad) & Ed Esplin (Invitae/Labcorp)** 
   1. Broad Results Summary
      1. 24,005 samples have been resulted. 13,389 samples were blood and 10,616 were saliva.
      2. 991 samples (4%) resulted in Test Not Performed (TNP). Blood samples had a .8% TNP rate and saliva samples had a 8.3% TNP rate.
   2. AnVIL PRS Data Transfer
      1. 23,159 samples have been transferred to AnVIL. There are approximately 100 outstanding samples that are still in the process of being transferred.
      2. Withdrawn participants have been removed.
      3. Once all samples have been staged, it will take approximately 1 week for the merged VCF to be generated. VCF includes imputed and array.
   3. Invitae (now Labcorp) BAM transfer
      1. Coordination of the BAM file transfer is underway. The Labcorp team is working with the CC and the Broad to set up a secure cloud to cloud transfer. The first batch of BAM files has been transferred as of February 12th, 2025.
      2. The transfer will include data for genes sequenced on the eMERGE panel. Completion is expected by Q1 of 2025.
6. **Working breakout session** 
   1. **Data needs for upcoming manuscripts (Phenotyping, CARE, Outcomes)**
      1. The phenotyping, CARE, and outcomes workgroups are considering the formation of phenotype specific workgroups to take a closer look at condition measures/assessments of interest.
      2. There are two ways order and referrals could be collected from sites. One way is for sites to map orders to keywords based off of the list of outcomes of interest and the other way is for sites to send all order data and then it is sorted centrally. An issue with sorting through the orders centrally is that some sites may have oddly worded tests that are very site specific and someone sorting centrally would not be aware of those.
      3. There is a federally funded AI tool that was created at VUMC called Brim Analytics with a 98% success rate that is able to go through 800 charts in one day. This tool could be useful for sorting through orders collected from sites.
         1. The tool can be locally hosted at other sites besides VUMC and can be run with no external network connection. More information on the brim tool can be found here and here. For more information on the Brim tool or to see a demo, please reach out to Kate Mittendorf at kate.mittendorf@vumc.org.
      4. All conditions are interested in provider and patient actions as the primary endpoint.
      5. When thinking about clinical utility, it may be helpful to analyze or focus analysis on completed orders and encounters and not contemplation of orders/encounters. In the charts, it will be difficult to figure out intentions compared to completed orders.
      6. An example of how sites should think through this process using type II diabetes - sites are looking for an order for hemoglobin A1c which is a provider action. The completion of this test is the patient action. Both come from the EHR and from patient surveys and the point is to see if lifestyle changes were made.
         1. If type II diabetes also wants to look at a participant’s blood glucose in the last year versus after the GIRA return, that is a separate task.
      7. The outcomes workgroup should be deciding how to combine various variables for different kinds of analyses.
      8. It would be helpful if sites could see three lists - one for orders, one for orders received, and on for the actual test result. Taking this a step further, there can be 3 lists for each condition.
      9. Another consideration for the group is where to find the line between the overarching outcomes paper and the condition specific outcomes paper.
      10. There needs to be a master file of what the core outcomes are the core variables for the cross site analysis which has been created here.
      11. It is very important for the phenotyping workgroup and the condition leads to confirm the labs and meds lists are accurate. The outcomes group should confirm all outcomes of interest are accurate.
      12. The main outcomes of interest are variable meaning lipid profile can mean total cholesterol, HDl, LDL, etc and whether it was ordered.
      13. The vast majority of outcomes are coming from the EHR. The patient survey is only used for information that cannot be derived from the EHR.
      14. ICD codes can come from lots of different places and can have different meanings depending on where they come from. If they come from billing, those are often used as rule out codes. They are less reliable coming from orders and more reliable coming from problem lists. Also, if a code will not yield insurance coverage, it likely will not be sewn. A computed phenotype approach is used since that granularity is not available for analysis.
   2. **Data quality and sharing (GRID & QA/QC)** 
      1. The QC measures will be implemented on the raw CDP export.
      2. It was proposed to have three files on AnVIL: The raw CDP export, the QCed version, and the analysis-ready data set.
      3. ACTION ITEM: A working call will be scheduled to identify overlaps between variables in the missingness manuscript and analysis ready data set variables.
      4. It was proposed for the QC process to occur on the exported csv file for more consistency. The API may result in versions that differ from the version on AnVIL.
         1. Updates were made to the CDP to address an issue with character values being lost depending on data export method, and Megan He will confirm after the next push.
         2. ACTION ITEM: The QA/QC Task Force will assemble a group to write scripts to perform the QC measures on the raw CDP data.
      5. Static and derived survey variables require QC only once, and some recalculated variables will be finalized at the initial dataset release and remain unchanged.
      6. We aim to establish a stable pipeline for quarterly CDP updates in AnVIL, preventing overwrites of QCed variables, targeting completion by May/June. Outcomes-associated data from the first three surveys should be ready by July.
      7. If a mid-level cleaned analysis file serves as the source of truth, updates should exclude static, previously QCed instruments.
      8. A process diagram is in development to track files and workflow steps, which will be refined over time.
      9. A proposed plan includes documenting all changes with “what” (recoding, transformation, etc.) and “where,” aiming to be complete by June 2025.
         1. Transformations occur between raw and cleaned data, while collapsing and renaming happen between cleaned and analysis-ready datasets. QC coding will identify errors and add derived variables. Recoding involves column-level changes (e.g., true/false to 1/0), while transformations modify row-level data based on algorithms.
         2. Some variables in the analysis-ready dataset may not be QCed, such as those that don’t require QC or new derived variables.
      10. It is undecided whether variables not yet reviewed for QC should be included in the analysis-ready dataset. It was proposed to first release an analysis-ready dataset with defined columns, then release versions with QCed variables as they become available.
      11. Column labeling is locked, so work on the analysis ready dataset can begin and will not rely on the QCed data.
      12. The group reviewed the data milestones and noted that "analysis-ready" differs from "ready and QCed." May 2025 is too aggressive a timeline for anything beyond the pre-RoR surveys. MeTree could be ready by fall 2025. The GRID workgroup can complete the column level changes prior to the 6 month EHR data pull.
      13. It was proposed to break milestones further down, such as into column-level transformations and annotations, as well as different QC process stages (e.g., pre-RoR and baseline surveys vs. post-RoR surveys).
      14. Megan has started compiling a list of CDP variables that the QC group is annotating with needed information. The group found 250 more fields in the AnVIL CDP export than in the missingness variables spreadsheet.
          1. ACTION ITEM: Elisabeth Rosenthal will add a color legend to a new tab in the variables selected for the missingness manuscript spreadsheet.
      15. ACTION ITEM: Individuals interested in understanding the CDP data or performing outcomes analysis should join the GRID and QA/QC groups to help shape the product they will use.
7. **ELSI: External data sharing | Ingrid Holm (BCH), Maya Sabatello (Columbia) & Anna Lewis (MGB)** 
   1. The Network has been discussing data sharing and how it should be handled within the Network.
      1. Broad data sharing is expected, encouraged, and even required as part of the eMERGE funding.
      2. There are some concerns that come with data sharing, including re-identification risk and harm that could result from this for both individuals and groups. Loss of trust in research is also a concern. There is increased risk for harm for individuals belonging to small groups and socially-stigmatized populations.
      3. Researchers have a responsibility to protect participants they enroll into their studies. eMERGE has focused on enrollment of underrepresented populations.
   2. The Network has addressed data sharing limitations.
      1. To address reidentification concerns the Network made the decision to “roll up” categories with less than or equal to 10 participants prior to sharing them externally to reduce the chances of re-identification. Granular race and/or ethnicity data is not being shared externally.
      2. To address harms that could potentially result from re-identification the Network decided to redact certain codes (elective pregnancy termination, gender dysphoria/gender transition, and suspected/actual child abuse).
   3. Guidelines for the use of eMERGE data (particularly by secondary users) are in the process of being developed.
      1. The goals of the guidelines are to give details of the available data and guidelines for the data’s use. These guidelines would be made available alongside the eMERGE data.
      2. The guidelines are currently organized under the following categories: Race, ethnicity, ancestry, genetic ancestry, genetic similarity, participants with disabilities, sexual and gender minorities. There has been discussion about adding additional groups, such as participants with HIV or participants with substance use issues.
      3. Case examples will be included in the guidelines, indicating which variables it would be appropriate (or not) to use. The Network is encouraged to contribute case examples.
   4. NT499 (Data Sharing Tradeoffs) is currently being drafted.
      1. The manuscript outlines the issues and challenges related to data sharing, as well as the related decisions and reasoning for those decisions within eMERGE.
      2. The manuscript also highlights that we do not have the ability to influence how the data is used by secondary users and advocates for the need for data governance, informed by the right voices.
         1. There needs to be a balance between openly shared data that impacts science and participant protection.
         2. Users do have to apply for access to the data and have a legitimate research purpose. However, this does not apply to use of language after access for example.
      3. The authors have concerns about drawing attention to this work in the current environment.
      4. The consensus opinion from the Network is that the manuscript is valid and important academic work that should be submitted. The data collected in eMERGE is very important.
8. **Scientific priorities of eMERGE: Capturing additional data & utilization | Josh Peterson (CC)** 
   1. The goal of this session is to discuss requirements for data and analyses that extend the scope of eMERGE Aims.
      1. What additional data on our existing cohort would be valuable? Do we want to leverage the cohort for ancillary studies? Are there additional services that would increase access or impact?
   2. There are many data milestones already defined for Year 6.
   3. Potential domains of new work could include:
      1. New use of existing eMERGE data: Refine or create new PRS, refine or create new integrated risk instruments\*
         1. In eMERGE IV, breast cancer (BOADICEA / CanRisk) and coronary artery disease (Pooled Cohort Equation - PCE) were returned as integrated risk scores but integrated risk scores are available for other eMERGE phenotypes which include PRS.
         2. There is interest from the PRIMED consortium for further collaboration with the eMERGE consortium for clinical contextualization of PRS and in silico analysis with longitudinal data.
         3. Integrated Risk Score calculations exist for many of the eMERGE phenotypes.
         4. Screening for Abdominal Aortic Aneurysms can be considered, which was not selected as a phenotype. Incremental steps toward comprehensive prediction can be considered.
      2. Extend follow-up of cohort for existing outcomes, administer current surveys at later time point.
      3. Expand eMERGE data to enable new analyses:
         1. Mine participant EHRs for additional exposures or outcomes
         2. Actively gather data from recruited cohort (w/consent). For example, claims and payor data could be considered to track outcomes data outside of the health network, expand medication outcomes, and allow economic analyses.
         3. Connect other data sources
      4. Overlay an ancillary study with new intervention
   4. The timing of new interventions needs to be considered, since the effect of the intervention fades over time for a single intervention.
   5. Several other domains of research were proposed during the discussion:
      1. Determining the role of environment in family history, which is shared genetics and environment, in risk calculations.
      2. More depth in family history data (interaction between family history and PRS, communicating risk to family members).
      3. Harmonizing data for social determinants of health in PRS.
      4. Disease penetrance.
      5. Incidental diagnoses.
      6. The value of the dataset could be enhanced through passively collecting data from participants through syncing with local EHRs or participant smartphones.
9. **Discussion and closing remarks | Rex Chisholm (SC Chair, Northwestern)** 
   1. Following the September 2024 ESP/SC the ESP recommended that the Network's high-level outcomes paper include both adults and children, since the study was designed to cover the lifespan.
   2. Points to consider regarding the composition of the Outcomes paper include:
      1. The study is a single study and protocol. All data will be included in the Outcomes study design & RoR paper (NT466 & NT510). T2D and Obesity span the whole lifespan and are not unique to pediatric participants. Pediatric components may be hard to explain without the context of the larger study. Combining adult and pediatric outcomes may dilute the amount of attention for either. Analysis of pediatric conditions analysis will be combined, which may allow for more space for rationale and details in an independent manuscript.
   3. The high-level outcomes manuscript will not be the final paper about outcomes, but it will be the marker paper for primary and secondary outcomes. Phenotype papers will look at adults and pediatrics separately.
   4. The inclusion of pediatrics in the initial Outcomes paper had been discussed previously. The consensus was re-confirmed.
   5. DECISION MADE: The high-level Outcomes manuscript will include analysis on both adult and pediatric participants.

**eMERGE Day 2: Thursday, February 13, 2025**

1. **Outcomes data plans: Unanswered questions and next steps | Nita Limdi (UAB) & Dave Veenstra (UW)** 
   1. The first marker paper led by Jodie Jackson (CC) laid out what eMERGE was about and another paper led by Niall Lennon (Broad) laid out the polygenic risk scores and the pipelines, thresholds, and odds associated with the risk thresholds.
   2. The study design and analysis framework paper will focus on screening, outreach, engagement, enrollment, biospecimens collected, surveys completed, and family health history.
   3. The analysis paper is going to assess the adoption of the recommended actions and the intervention and control groups.
   4. Since there has been a communication gap between the roles of the phenotyping and outcomes workgroups, the groups will work together in the coming weeks to fill out an excel sheet that lists what will be measured.
   5. A comparison will be done between high risk and not high risk and whether providers adopt actions more for high risk participants and if those high risk participants complete actions more will be analyzed.
   6. A condition specific analysis will also be done across additional longer term actions.
   7. One of the first things that will be in the study design paper is a consort diagram that will lay out how many patients were screened, how many were enrolled, and how many were withdrawn. Additionally, the diagram will show how many participants had a sample, consented, and the sample was sent to the Broad and Invitae. Also, how many of the samples failed and if there was recollection.
   8. The final data set will be loaded into the CDP as the source of truth so all sites really need to update their CDPs for accurate numbers.
      1. Part of the process of updating the CDP at sites will be identifying where sites can go back and either update instruments or look at the accuracy of what is being reported compared to what is in the CDP.
      2. Sites will sign off on a comparison of the numbers reported in the Google spreadsheet being used for this task and the CDP numbers. .
   9. This spreadsheet has been created listing every condition lead site, co-lead site, condition lead, clinician lead, outcomes person of contact, and phenotyping person of contact.
   10. This spreadsheet will be used for sign off of remaining tasks.
   11. Although the outcomes heavily rely on EHR data, it is important to note that survey data will be used for some things like smoking where the ICD code for smoking is rarely ever used. It is important to reevaluate the outcomes spreadsheet to make sure all variables of interest are included.
   12. The group has agreed that a deadline (for by the end of April EHR data refresh kick-off) is in place to come up with a complete list of variables that sites need to capture.
2. **HCP survey results | Ingrid Holm (BCH) and Georgia Wiesner (VUMC)** 
   1. Overview of HCP survey results were shared.
      1. Background: Survey of Healthcare Providers (HCPs) who received a high-risk GIRA opportunity to document the impact of real-world genetic information on clinical care.
      2. Primary aim: Do HCPs perceive the GIRA report as useful for clinical care? Primary hypothesis: HCPs who are more confident in their abilities to understand PRS test results will more likely view the GIRA report as valuable for their patients.
      3. Sub aims were to understand why some HCPs do not review the GIRA and to understand why some HCPs do not act on and/or meet with their patient about the GIRA.
   2. Overall response rates were shared.
      1. The response rate was 16%: 1250 surveys sent and 195 completed at least part. Of those who completed the survey, 148 (78%) recalled receiving the GIRA. 90% of the providers reviewed GIRA with patients (132/146). Of those who did not review, the main reason was waiting for the next visit and no change in management.
   3. Results were shared about the demographics of the providers.
      1. The largest number of surveys completed by site were from UW and VUMC.
      2. Most providers were internal medicine with varying years in practice.
   4. The confidence of the providers in returning the GIRA was reported.
      1. Very few people were very confident in their knowledge about PRS, their ability to explain and their ability to make recommendations. Less than 10% of HCPs were very confident in their knowledge of PRS and ability to use PRS in practice.
   5. Providers were asked about how understandable and valuable the GIRA was.
      1. Most providers answered “somewhat” or “very” to the question about understandability. There was a spread in the response regarding value.
   6. Providers were asked about the actions they and the patient were likely to take based on the GIRA.
      1. Providers had varied responses to the question regarding their likelihood to act. Most were somewhat or very likely to act on the GIRA. Most providers said the GIRA did not affect workflow. There was an equivocal response regarding whether the patient would make changes with most responding “a little” or “somewhat” likely to make changes.
   7. Providers were asked about their communication and recommendation plans with the patient.
      1. About half said they had communicated with their patient, and of those who had not, about half said they were planning to.
      2. The most common reason that providers did not communicate was that there was no change in management.
      3. About 75% said that they did make a recommendation, and of those who hadn’t 2/3rds said they planned to make recommendations.
      4. The most common reason that the HCP did not plan to make a recommendation was that the patient already knew about the condition.
   8. Analysis was performed to look at correlation and associations between variables.
      1. There were strong correlations between their knowledge of PRS and their ability to explain a PRS as well as between their ability to explain with their ability to make recommendations.
      2. The understandability of the GIRA was associated with confidence in 1) knowledge of PRS, 2) ability to explain PRS results to patients, and 3) ability to make recommendations based on PRS. How valuable they thought results of the GIRA report are for their patient’s medical care. How likely they were to act on the GIRA reports. How likely they think their patients are to make healthcare changes on the GIRA recommendations.
   9. Next Steps
      1. Additional analyses will be performed to look at additional questions and demographic differences. Provider interviews will also be performed soon as a result of the recent amendment.
   10. Questions from others were shared.
       1. Comments were shared regarding site differences and level of engagement among providers. In addition, there are concerns about the low response rate and the investment of the provider in the study.
       2. A suggestion was made to consider analysis stratified by disease as recommendations will differ by the different diseases.
       3. A concern was shared that providers may have differing confidence in recommendations based on age.
       4. The central message is that providers need support to feel confident in returning GIRA.
3. **Panel: Manuscript updates** 
   1. CIRT
      1. NT519 (Emma Perez (MGB) & Bob Freimuth (Mayo)) eMERGE IV: Experience integrating genomic-based risk assessment reports into the medical record
         1. This manuscript outlines what the sites did to put results in the medical record. The group is working with CARE and Outcomes workgroups to make sure there is not too much overlap with other manuscripts.
         2. Barriers/Dependencies - Coordination with other workgroup papers and confirming site workflows
         3. Status/Timeline - Writing in progress (intro and methods are drafted); Submission target of May 2025
      2. NT520 (Bob Freimuth (Mayo) & Nephi Walton (NHGRI)) Integrating Polygenic Risk Score into the Electronic Health Record: An eMERGE Network Perspective.
         1. This manuscript looks at the future state of technologies needed to support networks like eMERGE going forward as well as lessons learned.
         2. Barriers/Dependencies - Potential restrictions on NIH personnel/bandwidth
         3. Status/Timeline - Writing in progress; Submission target of June 2025
      3. NT527 (Jeritt Thayer (CHOP) & Bob Freimuth (Mayo)) eMERGE: Using Socio-technical Models for Translational Genomic Research
         1. This manuscript looks at a framework for implementation that can be applied to genomic medicine broadly and uses lessons learned from prior phases of eMERGE.
         2. Barriers/Dependencies - Connecting with content experts to gather information,
         3. Status/Timeline - Writing in progress; Submission target of May 2025
   2. CARE
      1. NT426 (Sabrina Suckiel (Mt. Sinai) & Noura Abul-Husn (Mt. Sinai)) Genetic counselors' perspectives on returning high-risk PRS results in eMERGE
         1. A survey was done to look at perspectives of different site personnel who have returned results and their perceptions of the process.
         2. Barriers/Dependencies - None (survey results have been collected)
         3. Status/Timeline - Submission target of early
         4. NT507 (Emily Miller (MGB) & Matt Lebo(MGB)) Challenging and interesting cases in returning Genetic Informed Risk Assessments to patients (eMERGE Edge Cases) Sites contributed challenging or interesting cases related to the return of GIRA. Pediatric cases are now being included in a separate paper.
         5. Barriers/Dependencies - None.
         6. Status/Timeline - The writing group is currently going through the cases to determine which cases will be included Submission target of
         7. NT510 (Lucinda Lawson (CCHMC)/Cindy Prows (CCHMC) & Gail Jarvik (UW)/Iftikhar Kullo (Mayo)) The Return of Genome Informed Risk Assessments for Common Conditions to 20,000 Adults and Children: The eMERGE experience of incorporating polygenic risk scores in routine clinical care This is the main CARE workgroup manuscript. It will be a descriptive process manuscript and start where the main Outcomes paper finishes. The title should potentially be altered to indicate that the manuscript will focus on distribution of risk.
         8. Barriers/Dependencies -
         9. Status/Timeline - Data analysis and writing in progress; Submission target of
         10. NT511 (Jasmine Purcell (CHOP)/John Lynch (CCHMC) & Ingrid Holm (BCH)) Parents' responses to their children's PRS-based genomic risk results These interviews were conducted under separate IRBs at CCHMC and CHOP, however the methods and interview guides are identical. Barriers/Dependencies - None
   3. Status/Timeline - The writing group is in the process of analyzing the qualitative data; Submission target of Additional potential manuscripts should be focused on outcomes (clinical, lifestyle, psycho-social). Outcomes
      1. NT466.1 (Nita Limdi (UAB) & Dave Veenstra (UW)) Study design and analysis framework [cross-condition]
         1. Barriers/Dependencies - None.
         2. Status/Timeline - Writing in progress, currently confirming numbers in CDP for accurate consort diagram.
      2. Cross-condition key outcomes analysis (Nita Limdi (UAB) & Dave Veenstra (UW))
         1. Barriers/Dependencies - Outcome measurement, data availability.
         2. Status/Timeline - MCS is in progress with an aim to submit within the next month.
   4. Additional potential Outcomes manuscripts
      1. There will be several Post RoR related manuscripts
         1. Patient Post-RoR survey related manuscripts. The Network will need to differentiate from outcomes used in cross-condition and condition-specific analyses.
         2. Provider Post-RoR survey and Interview related manuscripts.
      2. Outcomes will collaborate with condition leads on Condition-specific manuscripts. Outcomes is a broad space and investigators should coordinate with the Outcomes workgroup regarding potential publications.
   5. Phenotyping
      1. NT490 (Xin Yi & Shawn Murphy (MGB)/Wei-Qi Wei (VUMC))
         1. NT504 (Kavi Wagholikar (MGB) & Wei-Qi Wei (VUMC)/Shawn Murphy (MGB)) The use of the EHR to capture Outcomes in eMERGE; Significance: Summarization of lessons-learned for using EHR to capture outcomes This manuscript involves work being done with the Phenotyping group to collect variables needed by the Outcomes group. One lesson learned outlined in the manuscript relates to challenges with order in the EHR. A letter on this topic is being submitted to JAMA. The JAMA letter final draft is ready and will be submitted February 20th.
         2. The full paper will cover the challenges faced in extracting EHR data and the plan to resolve those challenges.
            1. The point was made that the authors should be cautious about sending too much of a negative message regarding the use of EHRs.
   6. QA/QC
      1. NT516 (Elisabeth Rosenthal (UW) & Maya Sabatello (Columbia) Scope of Data Missingness, evaluation of potential non-random missingness, and Variables correlated with Missingness in the CDP.
         1. Barriers/Dependencies - The group is working to identify and determine the complete list of variables to include.
         2. Status/Timeline - In progress; A group is meeting regularly and preliminary analysis underway.
      2. NT525 (Jennifer Pacheco (NU) & Lisa Martin (CCHMC)) Utility of structured EHR data for capturing referrals and encounters: Findings from eMERGE 4 Manual Chart Review.
         1. Barriers/Dependencies - Chart reviews fully completed & EHR data pull(s).
         2. Status/Timeline - Chart review is almost fully complete; Initial data analyses have started and a plan has been drafted for full analyses.
      3. Tentative: Multisite QA/QC: Successes, Challenges, and Opportunities
         1. Through eMERGE we have learned to start the QC sooner rather than later.
         2. Barriers/Dependencies - It will be necessary to avoid overlap with the Data sharing tradeoffs MCS as well as site specific papers.
         3. Status/Timeline - Pre-MCS.
   7. Additional potential QA/QC manuscripts: While additional manuscripts are not planned at this time - we expect that the work from the QA/QC group will help address questions/concerns faced by others within the Network.
   8. GRID
      1. NT522 (Adam Gordon (NU), Matt Lebo (MGB)) Harmonization and analysis of eMERGE participant-reported family health history data derived from the MeTree tool. Part of the analysis will look at the missingness and patterns across the MeTree data. MeTree was implemented differently at different sites so it will be important to tease apart the site implementation effects. The plan is to incorporate rescue survey data where possible, but it will be limited to eMERGE conditions. There may be potential for another paper investigating the degree to which family history is collected within health systems and what was seen with the MeTree tool.
         1. Barriers/Dependencies - Requires bulk de-identified MeTree export
         2. Status/Timeline - MCS submitted and in review
      2. NT524 (Adam Gordon (NU) & Georgia Wiesner (VUMC)) Pan-cancer polygenic risk conferred by cancer-type-specific PRS, involves all 3 cancer groups. This manuscript involves all three of the cancer phenotype groups. It is mostly focused on the eMERGE IV data, with eMERGE 1-3 data being complementary.
         1. Barriers/Dependencies - Requires eIV genotyping data
         2. Status/Timeline - Analysis of eI-III data is underway
         3. Jordan Smoller has been leading a PRIMED analysis of PRS in context, stratified by biological sex, using eMERGE III data that could be a good resource for this manuscript.
   9. ELSI
      * 1. NT532 (Jackie Odgis (Mt. Sinai), Priya Marathe (Mt. Sinai), Naama Elefant (Columbia) & Noura Abul-Husn (Mt. Sinai), Ingrid Holm (BCH), Eimear Kenny (Mt. Sinai), Maya Sabatello (Columbia), Sabrina Suckiel (Mt. Sinai)) Impact of Genome-Informed Risk Assessments for Common Conditions on Perceptions and Emotional Response in a Diverse National Cohort This is in collaboration with the RoR and Outcomes workgroups. The manuscript looks at the psychosocial impacts of return of results. Barriers/Dependencies -
        2. Status/Timeline -
        3. NT534 (Minerva Nong (Columbia) & Maya Sabatello (Columbia)/Ingrid Holm (BCH)) Behavioral impacts of GIRA results: Experiences of eMERGE-4 Participants (with Outcomes and CARE workgroups - working through overlap). The manuscript looks at the behavioral impacts of return of results. The authors will work with the other groups regarding overlap. Barriers/Dependencies -
        4. Status/Timeline -
      1. NT483 (Laura Rasmussen-Torvik (NU); Maya Sabatello (Columbia) Complications of covering preventive care costs in research studies.
         1. Barriers/Dependencies -
         2. Status/Timeline -
         3. NT483.1 (Kate Bonini (Mt. Sinai) & Laura Rasmussen-Torvik (NU)) Uptake of and experience with medical care cost coverage in eMERGE. Barriers/Dependencies -
         4. Status/Timeline -
         5. NT516 (Elisabeth Rosenthal (UW) & Maya Sabatello (Columbia)) Scope of Data Missingness, evaluation of potential non-random missingness, and Variables correlated with Missingness in the CDP. Barriers/Dependencies - Working to identify and determine the complete list of variables to include.
         6. Status/Timeline - A group is meeting regularly and preliminary analysis underway.
         7. NT487 (Ingrid Holm (eMERGE) & Genevieve Wojcik (PRIMED)) Social and ethical considerations for the use of race, ethnicity, and genetic ancestry in the development, validation, and clinical implementation of polygenic scores [PRIMED/eMERGE collaborative project]. Barriers/Dependencies -
         8. Status/Timeline -
         9. NT499 (Anna Lewis (MGB) & Maya Sabatello (Columbia)) Data sharing tradeoffs Barriers/Dependencies -
         10. Status/Timeline -
      2. NT456 (Naama Elefant (Columbia), John Connolly (CHOP), Jackie Odgis (Mt. Sinai), Maria Riano (Mt. Sinai), Emma Henricks (MGB), Maya Sabatello (Columbia)) Inclusion and participation of individuals with disability in eMERGE IV
         1. eMERGE has a unique disability cohort.
         2. Barriers/Dependencies - Analyses are almost completed; Completed post-RoR surveys are needed.
         3. Status/Timeline - Writing of paper in progress; 2025.
      3. NT480 (Makenna Martin (VUMC)/Harris Bland (VUMC) & Kate Mittendorf (VUMC)) Clinical and psychological impact of sex- and gender-related data collection and usage in genetics research: A qualitative analysis within a large genomics network
         1. Most necessary contracts are in place and DUAs were needed.
         2. Barriers/Dependencies - Insufficient power; lapse in funding.
         3. Status/Timeline - Data collection is in process; Interview completion anticipated by June 2025; Submission planned for Summer 2025.
      4. NT481 (Makenna Martin (VUMC)/Harris Bland (VUMC) & Kate Mittendorf (VUMC)) Design of an improved sex- and gender-related data model and usage principles for genetics research.
         1. Most necessary contracts are in place and DUAs were needed.
         2. Barriers/Dependencies - Insufficient power; lapse in funding.
         3. Status/Timeline - Data collection is in process; Interview completion anticipated by June 2025; Submission planned for Summer 2025.
      5. Participants with disabilities - RoR cases (with RoR subgroup) (Maya Sabatello (Columbia).
         1. Barriers/Dependencies -
         2. Status/Timeline - MCS in progress;
      6. Additional potential manuscripts - Planning for ELSI-focused manuscripts using the participant survey data Peds vs adults
         1. Enrollment & recruitment (Molly Hess (CHOP)).
         2. RoR (Shannon Terek (CHOP)).
      7. Participants with disabilities - RoR edge cases (with RoR subgroup & Maya Sabatello (Columbia))
         1. Only two cases have been identified so far. The Network is asked to send the group more cases if possible.
         2. The MCS is in progress and should be developed
   10. Transgender participants (on hold).  Potential overlap between Network-wide manuscripts and condition-specific manuscripts is a concern.
       1. The Network papers should explain background and provide wholeness of data. Individual condition papers that are pushed out before the data is finalized may provide an incorrect picture of the data. Finalized numbers are crucial and are in process.
       2. Related to outcomes, network papers should come first.
       3. Junior investigators should be encouraged to publish and carry out analyses. It is a disservice to postdocs to not be able to publish on the research they work on.
       4. The Network should consider what is included in the condition-specific manuscripts so as to not take away/detract from the overall Network manuscripts. Some condition papers don’t look at outcomes.
       5. Timing of publications and overlap are separate issues.
       6. Overlap should be discussed prior to the finalization and submission of MCS. It is also critical for Network members to voice concerns during the MCS approval process.
          1. Comments and concerns regarding an MCS are facilitated by the CC. Co-authors are contacted and connected with the individuals with concerns. If resolution cannot be achieved at this level, the concerns can be brought to leadership.
       7. Concepts may evolve over time. If the scope of an MCS changes it is the responsibility of the authors to notify the CC so the MCS can be updated and the Network can be notified.
       8. Guidelines related to publications need to be flexible.
       9. Conversation is needed throughout the entire publication process, starting with concept development prior to submission for approval.
4. **Panel & discussion: Y6 Workgroups and major network milestones | Rex Chisholm (SC Chair, Northwestern)** 
   1. Rex Chisholm shared the major areas of work in eMERGE
   2. A new discovery project focuses on MeTree harmonization. Discussion included how it should look and whether it can be improved using rescue survey data.
   3. It was suggested that joint meetings be held between the Outcomes and Phenotyping groups. The network should consider whether additional people need to be identified to lead or support these projects.
   4. **CARE**
      1. The focus is on completing manuscripts.
      2. The workgroup decided to meet once a month to track RoR and address the issue of non-responders.
      3. In eMERGE 3, there was a substantial number of non-responders, and the same trend is observed in eMERGE 4.
      4. Some individuals did not respond because they already had a disease diagnosis and did not see a need for follow-up.
      5. ACTION ITEM: All sites should reconcile their site reported data with the CDP data and sign off using the CDP Sign-off spreadsheet by March 17th, 2025.
   5. **Outcomes**
      1. The workgroup aims to collaborate with the Phenotyping workgroup to ensure outcomes include lab values and test orders.
      2. They will work with CARE on documenting post-RoR survey completion.
      3. They will develop a study design paper and outline cross-condition outcome analysis papers. Manuscripts will be reviewed to ensure minimal overlap. The team will collaborate with condition leads and site representatives.
      4. Each condition must ensure that data elements are accurately captured and aligned with the CDP.
      5. The workgroup must carefully consider overlap between working group papers.
      6. ACTION ITEM: The outcomes concepts (the “what”) must be reviewed and confirmed by the end of March prior to the “how” is defined.
   6. **Phenotyping**
      1. The workgroup is finalizing variable sign-offs, pulling orders, and mapping them to variables.
      2. Data will be uploaded in two steps, including ICD codes, RxNorm, CPT codes, and LOINC/labs.
      3. The workgroup wants to add a step where records are pulled to compare with the manual chart review. This will first be done for encounter and order data.
         1. ACTION ITEM: The Phenotyping workgroup should collaborate with the QA/QC Task Force to compare orders and referrals data from the manual chart review to a pilot EHR pull of the same participants.
      4. Pulling orders involves extracting them from the EMR system and creating a list for mapping to variables.
      5. The goal is to determine which orders correspond to relevant assessments (variables).
      6. Once a preliminary list is created, condition leads will review it before final approval by the Steering Committee.
      7. Some time will be built into the data pull timeline to allow for iterations.
   7. **QA/QC**
      1. The workgroup is completing an evaluation of the eMERGE 4 CDP variables for quality issues using the list of variables used in outcomes analysis.
      2. There is significant collaboration between the Phenotyping and Outcomes workgroups.
      3. A proposal was made for the Phenotyping, Outcomes, QA/QC, and GRID co-chairs to meet monthly for check-ins. The already scheduled co-chair meeting could be used for ongoing coordination.
      4. The workgroup must use the same dataset for analysis to avoid confusion from using different versions. There was a discussion on dataset evolution due to participant withdrawals. If participants withdraw after data is released to AnVIL, the CDP will be updated, but their data will not be removed.
         1. ACTION ITEM: The QA/QC Task Force should collaborate with the GRID workgroup to develop a schema for version control of the CDP data in AnVIL.
      5. Users of the data are encouraged to assist with quality control to ensure proper understanding of the dataset.
      6. Once the six-month surveys are complete, the dataset should become static.
      7. A six-month data freeze will include multisample, CDP, and EHR files, planned for April 2025. This version should be used until November 2025 when a 12-month update occurs.
      8. Pulling directly from the CDP instead of from a data freeze on AnVIL may lead to inconsistencies.
      9. There will be two final datasets: A six-month dataset and a twelve-month dataset. One of these datasets should be used for analysis, preferably the twelve month.
      10. ACTION ITEM: The CC will investigate creating a static version of the CDP in REDCap for each data freeze.
   8. **GRID**
      1. The workgroup wants to release the Analysis Ready Data Set and integrate it with i2b2.
         1. ACTION ITEM: The CC will work with Jeff Klann and the GRID workgroup on setting up i2b2 beta testing.
      2. Structured MeTree data will be released.
      3. A new initiative involves kinship analysis of eMERGE 4, aiming to understand genomic data relationships and provide a resource for other researchers. PRS for new conditions are being analyzed in eMERGE 4, using Venous Thromboembolism (VTE) as an example. Several synergies exist between these initiatives.
      4. Work on the kinship analysis and PRS for new conditions are pending the release of the joint call dataset.
   9. **ELSI**
      1. Work continues on manuscripts.
      2. The workgroup is discussing best practices for sharing racial and ethnic data, including how to report and describe categories.
      3. A publication policy is being developed.
      4. ELSI is expected to be more involved in Year 6 due to emerging challenges.
      5. There is collaboration with other workgroups on manuscripts.
      6. The workgroup is exploring ways to study populations at risk, provided current language and terminology restrictions.
   10. The SC considered whether the co-chair call should have dedicated time for coordinating work between workgroups and collaborations.
   11. While cross workgroup discussions are helpful, combining workgroups may hinder productivity.
   12. Leadership will determine the best way to structure discussions during co-chair calls. Co-chair calls should include updates and feedback requests from different workgroups.
       1. ACTION ITEM: eMERGE Leadership will facilitate structured Co-Chair calls in order to foster collaboration and coordination work.
       2. ACTION ITEM: The network must prioritize finalizing the comprehensive order list spreadsheet.
5. **Discussion & closing remarks | Rex Chisholm (SC Chair, Northwestern)**
   1. There was a lot of great discussion regarding the data (location, format, documentation) during this Steering Committee meeting.
      1. Any questions or concerns about data access can be discussed at a PI call.
   2. DECISION MADE: The Co-chair and PI call can be used to adjudicate conflicts about manuscripts when necessary.
   3. The Network should implement versioning of the CDP data to ensure that analyses are performed on the same dataset(s).
   4. The Co-chair call is an open meeting such that any workgroup members can attend as needed.

**Action Items:**

1. Following the conceptual review, condition leads, clinical experts, and phenotype experts should finalize all keywords, orders, and results in order to complete the EHR DD.
2. eMERGE Leadership will facilitate structured Co-Chair calls in order to foster collaboration and coordination work.

CARE

1. All sites should reconcile their site reported data with the CDP data and sign off using the CDP Sign-off spreadsheet by March 17th, 2025.

QA/QC & GRID

1. A working call will be scheduled to identify overlaps between variables in the missingness manuscript and analysis ready data set variables.
2. Individuals interested in understanding the CDP data or performing outcomes analysis should join the GRID and QA/QC groups to help shape the product they will use.
3. The QA/QC Task Force should collaborate with the GRID workgroup to develop a schema for version control of the CDP data in AnVIL.
4. The QA/QC Task Force will break down the count and percentages of participants in the manual chart review who had the GIRA mentioned in their EHR by risk status.
5. The QA/QC Task Force will assemble a group to write scripts to perform the QC measures on the raw CDP data.
6. Elisabeth Rosenthal will add a color legend to a new tab in the variables selected for the missingness manuscript spreadsheet.

Outcomes

1. The outcomes concepts (the “what”) must be reviewed and confirmed by the end of March prior to the “how” is defined.

Phenotyping & QA/QC

1. The Phenotyping workgroup should collaborate with the QA/QC Task Force to compare orders and referrals data from the manual chart review to a pilot EHR pull of the same participants.
2. The network must prioritize finalizing the comprehensive order list spreadsheet.
3. The phenotyping group will investigate the best approaches to pull accurate mappings for order keywords.

Coordinating Center

1. The CC will investigate creating a static version of the CDP in REDCap for each data freeze.
2. The CC will work with Jeff Klann and the GRID workgroup on setting up i2b2 beta testing.

**Decisions Made:**

1. The Co-chair and PI call can be used to adjudicate conflicts about manuscripts when necessary.
2. The high-level Outcomes manuscript will include analysis on both adult and pediatric participants.