**Summary of eMERGE SC Meeting: June 22 & 23, 2023**

|  |  |
| --- | --- |
| **Day 1 - June 22, 2023** | |
| **Time** | **Event** |
| 8:30 - 8:45 | [Announcements & Opening Remarks | Rex Chisholm (SC Chair, Northwestern)](#63rn21ssy2t) |
| 8:45 - 10:25 | [Panel: Data Quality and Assessments](#9irw5uurnyn6)   * Interim Data Freeze: Timing and Strategy (20 min) | Wei-Qi Wei (VUMC) * Data Complexity: Clinical Variables & Outcomes (30 min) | Priya Marathe (Mt Sinai) & Jen Pacheco (NU) * Network & Site Data Quality Assessment (25 min) | Lisa Martin (CCHMC) * Q&A and Discussion (25 mins) | Chunhua Weng (Columbia) |
| 10:45 - 11:05 | [Clinical Operations Update | Niall Lennon (Broad) & Sienna Aguilar (Invitae)](#po1tnzoruats) |
| 11:05 - 12:25 | [*Workgroup Breakout Session One*](#kix.7ei375oior7)   * **Comprehensive Risk Assessment & Return** (Banneker Ballroom) * **PRS & Clinical Operations** (Lyra/Lynx) |
| 1:05 - 1:25 | [A Case of Atypical Progeroid Syndrome | Katherine Bonini & Ayuko Iverson (Mt Sinai)](#zf163z77uyv1) |
| 1:25 - 2:45 | [*Workgroup Breakout Session Two*](#kix.ynxkoq63m6jj)   * **Outcomes & Phenotyping** (Banneker Ballroom) * **R2/sIRB/ELSI** (Lyra/Lynx) |
| 3:05 - 3:25 | [Edge Cases in the Wild | Emma Perez & Matt Lebo (MGB)](#9ihh78g7vvo4) |
| 3:25 - 3:55 | [Recruitment Challenges & Successes | Digna Velez Edwards (VUMC) & Ingrid Holm (BCH)](#xozd3twkgc92) |
| 3:55 - 4:10 | [Closing Remarks & Discussion | Rex Chisholm (SC Chair, Northwestern)](#m14uothfb93w) |

|  |  |
| --- | --- |
| **Day 2 - June 23, 2023** | |
| 8:30 - 9:10 | [Breast Cancer Edge Cases & Family History | Wendy Chung (Columbia) & Georgia Wiesner (VUMC)](#kix.9zki66bqeq3e) |
| 9:10 - 10:00 | [GIRA Return Challenges & Successes | Margaret Harr (CHOP) & Nita Limdi (UAB)](#kix.qbxi7b7pdgd3) |
| 10:00 - 10:45 | [Program Report & Network Timelines | Leadership](#kix.h2lq2gbonk5t) |
| 11:05 - 11:25 | [Clinical Decision Support & Integration Progress | Emma Perez (MGB) & Eta Berner (UAB)](#kix.820yccm9p91j) |
| 11:25 - 11:45 | [Polygenic Background Affects the Penetrance of Monogenic Kidney Disease | Atlas Kahn (Columbia)](#kix.sczmvzk4w5km) |
| 11:45 - 12:00 | [Closing Remarks | Rex Chisholm (SC Chair, Northwestern)](#kix.a919jqjhdvat) |

1. **Announcements & Opening Remarks | Rex Chisholm (SC Chair, Northwestern)**
   1. Goals: Data QC & GIRA return progress
      1. Day 1 will focus on data quality & harmonization and discussions will be held about data quality and assessment as well as edge cases & recommendations.
      2. Day 2 will focus on GIRA return successes & challenges and discussions will be held about breast cancer and BOADICEA edge cases and recommendations, GIRA return challenges & successes and timeline discussions.
   2. Enrollment progress and metrics: 12,682 enrolled.
      1. We are more than halfway through enrollment. Congratulations to everyone!
      2. 90% of participants have completed baseline surveys and 85% have completed Pre-ROR surveys.
      3. Demographic breakdowns show 66% are women, with 51% white and 10% pediatric.
   3. GIRA generation & return progress
      1. Eight sites are actively generating and returning GIRA and all sites are sending data to partners.
      2. 1,059 GIRA reports have been generated and 820 have been returned.
      3. There have been 303 high risk out of those generated. This is a little higher than expected but may be due to bias in who had reports generated.
   4. Goals for the next 3 months
      1. Kick off EHR data pull. Remaining hurdle is to sort out the exclusions.
      2. Move out of beta phase with all 10 sites actively returning GIRA.
      3. Reach 17,500 recruited individuals and all sites begin recruiting pediatrics.
      4. Assessment of back up survey utilization and 6 month Post-RoR survey sent to initial GIRA recipients.
      5. Increase confidence in GIRA return and have more infrequent issues. Sites should take advantage of the change log and if there is concern that the response to the change request hasn’t been documented, reach out to the CC.
   5. Challenges ahead
      1. Finalizing redaction codes, collating and cleaning interim data freeze. The ESP will ask us about this so this should be a high priority.
      2. Ensuring data we collect is useful and sufficient for outcomes analysis. We created a new workgroup tasked with this item.
      3. Establish access to the eMERGE data for eMERGE investigators and familiarize the Network on how to use AnVIL for data analysis.
      4. Moving out of the beta phase for GIRA return so that at the next meeting, we’ll be able to look at the returned data more closely.
2. **Panel: Data Quality and Assessments**
   1. **Interim Data Freeze: Timing and Strategy | Wei-Qi Wei (VUMC)**
      1. After suggestions from the ESP, the network has implemented a plan for an interim data freeze.
      2. The phenotyping workgroup has been most recently working on finalizing a data dictionary and overall myplan for an outcomes data pull including a data quality control process.
      3. There will be two separate categories to the data pull - one data pull that will include ICD & CPT codes, specific labs & meds, and visit information and one data pull that will include specific referral and order data. All will come from participant EHRs, however the referral and order data will be pulled separately at a different time to allow for more focus on consistency across sites since it is a more complex data pull.
      4. All diagnosis and procedure codes will be collected in addition to a list of medications and labs. The visit, medication, and lab data will be collected after 2017 to ensure there is sufficient data to analyze before and after enrollment.
      5. Age at event will be collected for all data in place of event date (for example date of CPT code or date of lab will be excluded and participant age at that event will be included instead) so there will be no identifiable information based on age included with the data pull. Age at event will be pulled to 3 decimal places to allow for granularity.
      6. The referral data will not have a one size fits all process for every site since not all sites use Epic and some sites use the Epic reporting workbench or a different way of obtaining referral information.
      7. The data will be collected via PheKB which is the same process as the previous eMERGE data collection. A PheKB page will be created for the specific data pull files where the data will go through basic quality control (checked against the data dictionary) directly on PheKB before going to the CC where all data will be thoroughly checked and compiled.
      8. The targeted medications and labs will be mapped to standard terminology with plans to do a keyword search to make sure nothing was missed due to mismatching.
      9. For this outcomes data pull, codes will be redacted from 3 major categories: elective pregnancy termination, gender affirming care, and suspected child abuse.
         1. These codes are being reviewed by the sites and a final list will be compiled, approved by the PI group, and released into the data dictionary for implementation.
      10. For the quality control process, 10 medications and labs of interest will be randomly selected and a chart review will occur to make sure nothing is being missed.
          1. Labs and medications with extreme values will be left in the dataset unlike with previous data pulls.
      11. Once the formal data pull request is sent to the network, the timeline for the data pull process will be shared.
      12. The referral data will be pulled at a separate time however the data may not be consistent.
   2. **Data Complexity: Clinical Variables & Outcomes | Priya Marathe (Mt Sinai) & Jen Pacheco (NU)**
      1. Discussions are ongoing around clinical variables and how data is extracted from REDCap and R4.
      2. Study staff at Mt Sinai uses a trigger variable to verify MRN in internal study status form.
         1. API pulls clinical variable data from the EHR into Mt Sinai’s REDCap.
         2. This is ultimately populated in R4.
      3. More than 250 GIRAs have been reviewed. These have been reviewed for accuracy and their internal data have been populated into R4.
      4. A few issues have been detected during this process.
         1. Due to a lag in enrollment and GIRA review, sometimes patients receive updated labs.
            1. The decision was made to not update these values in R4.
         2. Some participants have had labs from outside institutions that are not found in the labs tab in EPIC and API is not able to access these.
            1. It is not feasible to do chart review for each patient to address this.

If the date and time for each lab included in the GIRA is not annotated, it should be considered how to identify which lab was used (since we may not be able to use the latest lab value or external lab data).

* + - 1. The language in the GIRA is the same if a participant doesn’t have lab values for a specific variable and if they have lab values that don’t meet criteria.
      2. Some labs, like allergies, are not all numerical and may be missed.
         1. Prevalences can be compared to expected prevalence to identify discrepancies.
      3. Some vital labs, like BP, were not being pulled or being pulled in reverse order.
         1. Sites should check these before putting in R4.
    1. Proposed solutions:
       1. The updated clinical factor language “clinical factors (information) may be outdated or unknown” was amended to end of Clinical Risk Factors text.
       2. Trigger clinical variables to pull at time samples are sent to Broad and/or Invitae to be closer to time GIRA is generated.
       3. Update implementation guide to not only exclude outliers but also impossible values such as DBP>SBP.
    2. Ongoing review is occurring to verify accuracy of data and flagging problematice data.
    3. Data missingness for the important GIRA variables is being looked into.
    4. Outcomes are being compared to GIRA recommendations including:
       1. Tests, encounters, referrals, imaging, Rx or lab orders.
       2. Analyzing if outcomes are initiated or changed.
       3. Encounters and referrals added to post RoR survey.
       4. Estimates for new diagnoses and health care action uptake must be recalculated based on prevalence findings in enrolled populations.
    5. Review of more difficult to obtain EHR data is still in progress.
       1. Not all CPT codes are mapped and most sites don’t have Rx fill data. CAC and tumor biopsy feature grades may require NLP however are deferred for now.
    6. Three new labs are being added to the existing total. Some labs are not mapped to LOINC or have different units so that will need to be resolved (labs can possibly be mapped to LOINC codes using keywords searches in the EHR).
    7. More than 12 classes of meds are being added and there are 13 total classes now. OMOP IDs are being used, but 2.5% are missing those IDs. Many of these codes are only at ingredient level and there is no dose or unit information.
    8. Starting this summer, issues with site-specific EPIC instances where codes may not be consistent between sites will be addressed.
       1. Each site needs to determine which specific clinics (for example, specialty clinics site health systems use) they want to include so this can be taken into account when each site is searching for and pulling data.
       2. Data is stored differently between sites within EPIC.
       3. There are issues with determining PCPs and with PCPs leaving institutions, participants may be referred outside of the system and there won’t be an order for that.
       4. Telehealth visits may not be adequately pulled, some sites may not be able to view future encounters.
    9. Solutions:
       1. Sites review 10 or more charts each to make sure Rx, lab results, etc are in the charts and in OMOP standardization.
       2. Both lab name and LOINC code should be used if necessary and a search should be done by generic Rx names. This can also resolve issues when codes do not have one-to-one mapping.
       3. Sites should find and confirm with phenotyping leads that relevant clinics are selected.
       4. PCP should be confirmed during the GC appointment.
    10. In instances where there is no PCP, like a resident “run” clinic, the clinic could be entered in the PCP free-text field.
    11. How to treat cases where patients with a GIRA that would warrant a cardiologist visit see a cardiologist for another issue should be treated.
  1. **Network & Site Data Quality Assessment | Lisa Martin (CCHMC)**
     1. Lisa Martin is a co-chair of the eMERGE QA/QC Task Force. All sites and workgroups are represented in the QA/QC Task Force. This group first met on May 3rd, 2023.
     2. The task force is still closely working with the sites and workgroups to identify the key goals.
     3. There is difficulty related to implementing QA/QC procedures when the cohort is over halfway recruited.
     4. The CSER consortium also implemented data quality processes midway through their project, and published a paper on lessons learned.
     5. The task force is trying to engage with stakeholders in not just sites, but also in workgroups to ensure their work is not duplicative and informed by their work.
     6. eMERGE is collecting a lot of data, and the task force aims to identify which data elements are the most critical for regular evaluation.
     7. The long term goals of the task force are to make data-informed changes to the eMERGE process. A prior example of this is when the network noticed participants were having a difficult time completing MeTree, the rescue survey was developed.
     8. The task force members completed a spreadsheet of [current QC measures](https://docs.google.com/spreadsheets/d/1QqS7SR2aO-UUjg8vIDbvwUp0xh_e49SCqHHg5e5V6Lo/edit#gid=490652922) being used by each site, which revealed a wide range of variability between sites. There are opportunities to harmonize and share resources and lessons learned.
     9. The task force would like to maximize shared resources and minimize the amount of time spent on manual review and processes to create the best quality data.
     10. Sites expressed a preference to not use SaS due to cost.
  2. **Q&A and Discussion | Chunhua Weng (Columbia)**
     1. It is beneficial to have more consistency through electronic methods, versus manual review.
     2. Collecting data missingness by site may show variability but would be informative for each site.
        1. The QA/QC Task Force is planning to generate reports across the network and then dividing them by site.
        2. This is important as different sites may have different levels of missingness among different groups of participants.
     3. Chart reviews will be time consuming and expensive to appropriately train staff.
     4. ACTION ITEM: The outcomes group should prioritize the most essential data elements into high, moderate, and low priority.
     5. The reason for specialist visits does not need to be recorded, there just needs to be a mention of a visit.
     6. There are tools built into REDCap that UAB uses to track outliers and missing data weekly. These tools are for both survey data and EHR data. The data fields can be updated by study staff in R4 within the two weeks before the record locks.

1. **Clinical Operations Update | Niall Lennon (Broad) & Sienna Aguilar (Invitae)**
   1. Broad ClinOps Production Update
      1. Metrics:
         1. The Broad is now delivering for all sites.
         2. Currently, 4,545 samples have been received and 3,979 samples have been processed.
         3. The sample call rate is 99.01%.
         4. There have been 3,841 reports issued.
      2. Fail Rates:
         1. The number of fails/failure mode is 82.
         2. Broad monitors fail modes. The predominant fail mode is sample QC (60). This is related to concentration amount, quality, or a similar factor. They also have drop-out due to genotyping (10) and a few that require some investigation with the metadata. For example, reported sex discordance (11).
      3. High Risk Results:
         1. The team reviews where high risk results come from in the ancestry space to look for patterns. They look for outliers. So far everything is as expected.
         2. There have also been statistical models used to assess whether the results are different from expected. So far, some of the conditions are showing more high risk results than expected. As the population increases, the noise is accounted for and the outliers become less significant.
         3. There is some grouping and correlations with the participants who have multiple high risk PRS. There are 7 participants who have been high PRS for 3 conditions. T2D is a risk condition in 6 out of 7 and T2D and CHD is present in 4 out of 7.
      4. Research Data Release to AnVIL
         1. Retrospective validation: Broad is working to pull existing datasets into AnVIL. Legacy eMERGE data is being imported into AnVIL workspaces.There are some blocked studies due to data access issues.
         2. For this eMERGE study, there are currently over 3,600 studies brought into AnVIL. Broad is also working with Invitae to take their files and de -identify in the appropriate ways so they can be pushed into AnVIL as well.
   2. Invitae Updates
      1. Sample Updates
         1. As of June 9, they have had 3,620 total orders with 2,300 completed.
         2. When looking at orders by month, there is a spike around January.
         3. Per month, 250-300 orders have been sent per month.
      2. Results Updates
         1. As of May 5, there have been 32 positive results (1.6%) and 2 carriers. 1 result was an indeterminate result that couldn’t be disambiguated.
         2. There was 1 individual who had 2 findings in both MLH1 and PALB gene.
         3. The most frequently reported findings are BRCA2 (7) and LDLR (6).
      3. Sample Failure Updates
         1. There have not been any complete failures. There have been 5 sequencing only reports for saliva and 17 for GDNA. This failure rate isn’t a deviation from expected.
   3. Q&A
      1. Do we know if any of the BRCA2 results were already known? That would be anecdotal from a site. There were some results that participants already knew about. This will be discussed more in the CARE group.
      2. What are common causes of failures? Sex discordance, insufficient sample, missing metadata. Contact partners to see if the sample can be rerun.
      3. There were individuals having 2 or 3 high risk PRS. Are these genetically correlated? There are individuals with more than one cardio metabolic condition. When we have more data, the PRS overlap will be reviewed by the CARE group.
      4. Invitae reported a sex chromosome abnormality. Does Broad look at that as well? What is expected? Broad does not. In the future, CARE may review sex chromosomes mismatch in Invitae. There is preliminary data that the mismatch is about what is expected in the population. Invitae will investigate more for future meetings.
      5. For Invitae: What do we do if a clinical comment in the Invitae report says deletions could not be completed and an alert isn’t sent? A notification should be sent. Sienna will follow-up offline about the discrepancies.
2. **Workgroup breakout session one (Notes can be found in the workgroup google docs, linked below for reference)**
   1. **Comprehensive Risk Assessment & Return**
      1. A question for pediatrics, do we require the child to be present at RoR and how strictly should this be enforced? Are there age thresholds for this? How to address this?
         1. Everything has been geared toward the parent. Since this is not a pediatric study, we are comfortable with the child not being there. The child can be encouraged to be there, but not required. We could ask the parents on a phone call if they’ve discussed being present for RoR with the child. There are difficulties when participants turn 18 as they have to re-fill out their survey and it doesn’t always match up.
      2. A question on what to do with participants who decline genetic counseling and in-person RoR?
         1. In the IRB, there has to be 4 attempts to contact participants before returning to the PCP.
         2. To avoid participants screening “unknown” calls, the calls should come from hospital numbers and calls should occur during evenings. It appears that a substantial proportion of participants who don’t respond to calls get results.
         3. Some patients already have the disease they’re considered high-risk for, but still are willing to discuss results. For certain conditions, like obesity, this group may represent a large portion of high-risk returns.
      3. A question surrounding how to handle incidental findings, for instance a gain of X chromosome in males and if that should stop RoR.
         1. If something besides the participant’s status as transgender or a sample mix-up is present (i.e. Klinefelter syndrome), that is something medically actionable and should be returned.
         2. A second test is reasonable in these cases.
         3. In a situation of non-paternity, is there any concern for returning results?
            1. We don’t have incidental findings in our consent form and this is something that should probably go through the IRB to be amended. These findings should not be returned until that is agreed on. There is no scope for incidental findings currently.
         4. It may make sense to log these incidental findings as they come in, and then as a group the decision can be made on what/how to report these moving forward.
         5. It is argued that there is a responsibility to bring attention to clinically actionable findings. Contacting the PCP directly to discuss is an option.
      4. Currently the date of RoR is the date they were returned to participants, but we may want to change that to be the date the report was uploaded to EHR. This is causing a mismatch on our dashboard metrics.
      5. Cincinnati Children’s Hospital is returning their results to family units at the same time.
      6. High-risk prostate cancer reports have not been sent out while issues with medical care costs are being resolved.
         1. When patients have high integrated risk scores, insurance may cover MRIs. If a test is ordered and an MRI is ordered and turned down by insurance, we can’t pay for it. The decision must be made at the time of GIRA return on whether screening test will paid for by the NIH study funds or Insurance. PI can discuss with Laura.
         2. We have to schedule the appointment in order to attribute it to the study. This impacts outcomes.
         3. A letter is being drafted at some sites to outline for participants what they need to do to qualify for medical care costs and this is something we could consider adding as a global IRB amendment going forward.
         4. Most physicians won’t understand this process without clear instructions, as it is much different than the typical workflow.
         5. It is important to note that this is still an intervention and will impact outcomes.
         6. If participants already know their monogenic results, that will impact how they are sorted for outcomes.
   2. **PRS & Clinical Operations**
      1. The Genotyping workgroup had previously combined with the PRS workgroup.
      2. The main PRS related work is currently occurring in collaboration with PRIMED.
      3. The clinical operations subgroup will continue running independently as long as samples are in production.
      4. Based on the prior co-chair call, the PRS workgroup goals have been met, and it would be more beneficial for the network to develop a workgroup focusing on how to perform discovery analysis on the data that is being generated.
      5. There is an AnVIL Clinical Resource starting soon, and one of their key milestones is to work with the eMERGE network to demonstrate the utility of AnVIL. There will be one PI from eMERGE, and one from AnVIL.
      6. PRS is not contextualized. The GIRA is attempting to contextualize and integrate the PRS with other factors. This is an open area, and Robb Rowley had previously distributed a funding notice regarding research on this.
      7. The imputation pipeline used by the Broad is available in AnVIL.
      8. Self reported ancestry does impact the PRS. It is ideal to collect both self reported and genetic ancestry.
      9. An SDOH group on EHR data just started in PRIMED. PRIMED is focused on method development and calibration of PRS.
      10. Sites have members comfortable working in AnVIL, but there will be a learning curve across the network. There was a consensus that members do want to work with this data in AnVIL.
      11. The group should define what they want to achieve and where, and Robb can look into resources for funding. NIH is interested in integrated risk scores. More money goes into the development of AnVIL but will result in equity.
      12. The full data set from the Broad would be complete a few months after the Broad receives the samples. Invitae is providing a large panel and vcf files. The Broad is calling arrays by batches but have individual vcfs.
      13. Operational support and progress will continue to be addressed in the clinical operations group.
          1. There is an R4 report/notification when Broad samples fail. Currently, emailing the Broad to ask why the sample failed is the only way to know. There are plans to improve this system.
      14. The new workgroup will focus on discovery science, stemming from the prospective data. It was proposed to be project based. The workgroup would work stepwise to develop what the project will be.
          1. The focus will be on using data generated in eMERGE for discovery research to better understand clinical context integration of PRS in diverse populations. This group will examine data coming out of the clinical labs and ready the data for discovery science research applications.
          2. This group would also be the contact for cross-consortium work with PRIMED. This group will be distinct, and would attract different members than the PRS workgroup.
          3. The group name is Genomic Risk Innovation and Discovery (GRID).
3. **A Case of Atypical Progeroid Syndrome | Katherine Bonini & Ayuko Iverson (Mt Sinai)**
   1. eMERGE IV is used to identify individuals at risk for common, medically actionable conditions, but it also has the potential to incidentally identify and diagnose rare disease.
   2. The case presented was found early in the return of results process, in which a participant was found to have a novel *LMNA* variant (via Invitae) and a phenotype consistent with atypical progeroid syndrome (APS). LMNA variant is included on screening panels due to the risk of cardiomyopathy, arrhythmia, or neuromuscular disease.
   3. Case Participant: 32 y/o female, recruited via PC clinic; completed all steps of process, PRS = not high risk; limited family history available.
   4. Monogenic results: LMNA positive (indicating change in laminate protein/amino acid).
      1. Sherloc Scoring of variants also obtained, indicating likely pathogenic classification.
      2. Variants in LMNA are associated with hereditary cardiac, neuromuscular conditions, and accelerated aging conditions.
   5. Literature was reviewed for case studies with similar variants related to APS.
   6. APS is a subtype of Hutchinson-Gilford Progeria Syndrome (HGPS). HGPS is a rare and progressive condition characterized by early onset of accelerated aging and premature death.
      1. APS is considered milder (than HGPS) and is associated w/later age of onset and longer survival.
      2. Comparison of HGPS and APS physical features and characteristics were discussed.
   7. Results were returned via telehealth. Further information was collected at that time from the participant, and referrals made to cardiovascular genetics and clinical genetics for further evaluations.
      1. Cardiovascular genetics evaluation revealed normal for participant, but father has cardiac history and possible lipodystrophy. Father was adopted, with limited family history.
      2. Parental cascade testing is in progress for both father and mother of the participant.
      3. Additional imaging testing was recommended, and an audiology evaluation referral given to the participant.
   8. Regarding GIRA return (pending): PRS not high risk; did not trigger any high-risk results; this was discussed with the CARE group, and it was determined to be classified as not high risk.
   9. Findings are currently pending, and further information/results will be shared in the future.
   10. Genetics professionals should be aware of the clinical associations of LMNA related disease and be prepared to counsel participants when identified.
4. **Workgroup breakout session two (Notes can be found in the workgroup google docs, linked below for reference)**
   1. **Outcomes and Phenotyping**
      1. It is important to prioritize variables that, if they were missing or incorrect, would introduce bias into analysis.
      2. Sites are using different methods to generate QC reports. This is something that can be harmonized across sites. R4 has reports that can be used for much of this.
         1. Key reports in R4 could be emphasized to make them more accessible. Some metrics are tracked locally but are not things that would impact outcomes.
      3. The GIRA report is still generated when key clinical variables are missing.
         1. It is suggested that we should focus on BOADICEA and other conditions that require those elements. We typically don’t know the reason for an element’s missingness.
         2. It is suggested to look at the first 200 participants for each site and assess data completeness, with specific focus on high impact elements in PCE and BOADICEA.
         3. The R4 dashboards could be expanded to show missingness of specific elements as well.
         4. Is one site having better outcomes/data quality due to doing something differently and can that be shared with other sites?
         5. There could be sociodemographic differences in data completeness. Remote enrollment isn’t as effective as working with a coordinator that can help a participant complete information.
      4. There was discussion around code redactions that should be considered. The proposal is to drop these codes which include elective pregnancy termination, substance abuse, and transgender status. The initial list of codes to redact is too broad. This will be reviewed by the phenotyping workgroup. In addition, we need clinicians to review. Eliminating codes in some contexts but not others is not an easy task.
         1. Are these redacted only when sharing data or will they be redacted from our OMOP databases internally? These codes will be redacted at each individual site from all data files before the files are sent to the CC.
         2. The proposed codes to be redacted should be turned into a google sheet and shared to allow for comments. There are lists of codes that are used in legal determinations for specific states, maybe these lists can be cross-referenced.
         3. We should be careful to not take any action that would prevent us from doing a high quality PheWAS analysis in the future.
         4. Site representatives who can help with these determinations need to be identified.
         5. Drugs taken by patients can be used to help infer some of these codes.
         6. These codes should be prioritized, ranked high to low risk, to help guide discussion.
      5. Referrals tell more about how a clinician is interpreting rather than what’s actually happening with a patient’s health which is more of an implementation science.
      6. Outcomes aren’t just patient outcomes but what the providers are doing. Did the patient act on a referral?
      7. In looking at the time-window for referrals, time 0 is the upload of the GIRA into the EHR. Decision is to look for referrals for 6 months. And then by 12 months would look to see if that resulted in a new diagnosis.
      8. A PCP is only pinged currently for high-risk results. There is some disagreement on if this is best practice or not. PCPs always say they want fewer pings, but the volume of these is low enough to potentially not be a burden.
   2. **R2/sIRB/ELSI**
      1. Conversations in previous meetings regarding data sharing determined that codes that won’t be shared include codes for elective pregnancy termination, trangender care, and suspected child abuse.
      2. There is concern about possible repercussions around certain data being collected in some areas.
      3. Survey data is covered by the common rule and EHR data is protected by HIPAA laws.
      4. There is also disconnect from data provided in surveys (participant volunteered) vs EHR data (participant not necessarily aware of what is being accessed).
      5. One caveat of not getting these data is not advancing science in these groups, which contradicts having an inclusive study.
      6. Discovery data will be omitted in other populations as more codes are limited.
      7. MeTree is eMERGE specific and only contains information regarding eMERGE conditions.
      8. A possible manuscript on eMERGE coverage of medical care costs was discussed.
         1. Sites likely have different methods on how to cover these costs for participants.
         2. It is important that the [conditions](https://docs.google.com/document/d/1nhxSJTLDn4GVN7CL8EzvcefASvByg0NX6f-jOeIRUUI/edit) agreed upon will be paid from eMERGE for everyone.
      9. Defining passive withdrawal/disenrollment is set at 12 months of not receiving a sample from the participant.
      10. There is a need for updated REDCap reports that show more than “Complete” and “Not Complete” for better reporting of data.
      11. Many of the issues with the sIRB have been resolved by close work with the CC and sIRB leadership.
      12. Sites are reminded to consider creating global documents to make submissions easier to submit and for more timely approval.
      13. Concern about the participant burden when a parent enrolls multiple participants due to the necessity of completing separate studies for each was discussed.
5. **Edge Cases in the Wild | Emma Perez & Matt Lebo (MGB)**
   1. Case 1
      1. The first case was a 63-year-old white woman AFAB with a personal history of hypertension, hypercholesterolemia, obesity, and high risk for CHD. There was limited family history info.
      2. Her PRS score was high for Breast cancer at 99.7%.
      3. Although this participant was high risk based on PRS, based on her age and family history, the integrated risk was below 25%. This is expected based on the participants age and other factors. The integrated score incorporates multiple factors and there will be participants with a high PRS who do not meet the threshold defined in this study.
   2. Case 2
      1. The second case was a 41-year-old white woman AFAB with a personal history of hypertension, obesity, type 2 diabetes with metformin, low LDL values, and high triglyceride values.
      2. This participant was high risk for type 2 diabetes and also had a monogenic finding associated with APOB. She was a heterozygous carrier and the override function will be used.
      3. The counseling around this participant is going to be different and will focus on liver testing.
   3. Case 3
      1. The third case was a 66-year-old white woman AMAB with personal history of prostate cancer and family history of prostate cancer (among other conditions). She is also on hormonal treatment with testosterone blockers and estrogen.
      2. PRS was high risk for prostate cancer so this participant received a high risk GIRA.
      3. There is gendered language in the GIRA including risk relative to men opposed to people. An amendment is recommended. Breast cancer is only validated in people assigned female so caution should be taken if modifying the language in a similar manner.
   4. Case 4
      1. The fourth case was a 50-year-old AMAB. This participant was born blind and uses a screen reader. He alerted the study team that 2 questions were not accessible on the pre-ROR survey.
      2. This participant was not at high risk and did not require 1:1 return.
      3. There were several difficulties encountered when returning the GIRA including accessibility of images of text in the GIRA for FAQ, methods, and disease risk information. The CC provided word documents of the text to the site. In addition, the meTree pedigree was converted to a text file.
      4. The edge case group recommends reviewing the disability section of the pre-screen to assist with identifying participants who may be blind or deaf and need additional accommodations.

1. **[Recruitment Challenges & Successes | Digna Velez Edwards (VUMC) & Ingrid Holm (BCH)](https://docs.google.com/presentation/d/1vDqmhQ6jPwapdah9c7WYY6LzADVbJY_lCw31dQ3h8YA/edit" \l "slide=id.g2294b73f4e3_1_0)**
   1. Digna and Ingrid provided an update on the recruitment challenges and successes. They provided an overview on current recruitment compared to predicted recruitment. The first enrollment occurred in February 2022. Initially, the first return was targeted to occur in June 2024.
   2. The new updated targets are an enrollment end date of June 2024, which provides the network with 12 months from this month to complete recruitment, and a return end date of October 2024.
   3. A participant retention process was proposed, which consisted of reminder emails and other materials depending on site preference.
   4. Enrollment has been steadily increasing and the network is at over 50% enrolled. The majority (66%) of participants are assigned female at birth. The recorded baseline survey completed is 90%, and pre-RoR surveys completed is 85%. This contains missingness. The QA/QC Task Force will work on identifying the different levels of missingness in the surveys.
   5. The samples collected, as defined by samples at the Broad and Invitae, is 38%. Only 6% of GIRAs have been returned.
   6. The upcoming year should be used to push participants through the stages of the study.
   7. eMERGE sites have done better than expected when collecting MeTree information.
   8. There have been administrative hurdles including IRB/amendment timelines. This also includes the healthcare providers survey.
   9. The errors and updates to R4 have led to a reluctance among sites in sending samples. This could be the reason that only 38% of consented individuals have samples at the Broad.
   10. The group has been strategizing on how to recruit pediatric participants. There must be a large enough pediatric population in order to assess the pediatric phenotypes.
   11. Within the ELSI group, Maya Sabatello has been leading the effort to understand the barriers to enrollment for participants with disabilities. The ELSI group has also been working on ensuring clarity in the language in the surveys for transgender participants.
   12. Sites are encouraged to send samples through to the Broad and Invitae, and to let the CC know any specific concerns. Sites have been waiting for the beta phase to be complete, and it was unclear when the beta phase would end. The beta phase will end when all ten sites are returning, and the R4 errors have decreased. Currently, the errors have dropped to once a month.
   13. Sites are concerned regarding effort available to manually review the GIRA.
   14. UAB is sending samples to the Broad, but holding samples from Invitae as a way to control the rate of return of results. They returned 175 GIRAs and identified glitches, then closely worked on the next 100 GIRAs to ensure the solutions were working as expected.
   15. CHOP is pacing their sending of samples with their return of results. Margaret Harr has done internal modeling of how many participants to recruit, how many samples to be sent per month, and how many returns must be performed per month. They are sending samples based on these deadlines at the intervals to prevent a backlog.
   16. ACTION ITEM: Sites to create projections for sample processing and GIRA return rates. (PIs/sites; CHOP to share code; due July 31st)
   17. The Broad can run 10,000 samples a week. Invitae can only accept 50 gDNA samples a day.
2. **Closing Remarks & Discussion | Rex Chisholm (SC Chair, Northwestern)**
   1. If any sites are holding on to samples, it is encouraged they be sent to the eMERGE partners and let the CC know if there are any concerns. Currently, there are over 12,000 participants enrolled and 5,000 samples received. Sites should be sending these samples as quickly as possible so that there is not a large backlog at the end of the recruitment period.
   2. The goal of the network is to be out of the Beta phase by the end of the summer.
   3. Sites can divide how they send samples to the Broad and Invitae so they are not overwhelmed. Some sites are pacing the sending of results, returning one plate at a time, so the results are not backlogged. Some sites are not comfortable yet returning the GIRA for every condition, so have been slow and gradual in sending samples.
   4. Two-hundred GIRAs have been returned and a lot of time has been spent on quality control. Deeper dives have been done to identify issues and amendments have been implemented. Amending the GIRA can change the status for a patient so sites have paused on returning results.
   5. Sites should try and look at how many returns/samples need to be generated/returned to model the best way to divide/handle them.
   6. In closing, eMERGE continues to be cutting edge on a lot of the work being done and getting publications out will be really important.
3. **Breast Cancer Edge Cases & Family History | Wendy Chung (Columbia) & Georgia Wiesner (VUMC)** 
   1. GIRA uses Integrated Score for breast cancer risk including personal, family, and genetic.
      1. Risk is something that changes over the lifecourse with data.
      2. A 25% lifetime risk or higher triggers high risk GIRA. A 24.9% would be considered “not high.”
   2. Several “glitches” in BOADICEA risk score variables were flagged and fixed.
   3. Combined, full, model shows improvements in risk prediction, and specifically in identifying high risk individuals.
   4. For context, family history and polygenic risk values pale in comparison to monogenic risks with BRCA1/2. It is important to include this perspective in educating others.
   5. Aggregated risk prediction models outperform individual inputs with estimating risk.
   6. eMERGE high risk participants are managed by enhanced surveillance including breast MRI starting at age 40.
   7. Scores from Broad PRS and Integrated scores are not always in agreement which is expected.
   8. After code review of BOADICEA, sites are being asked to regenerate GIRA for those not returned and re-return any GIRA with altered risk status.
   9. If a participant has a personal history of breast, ovarian, or pancreatic cancer, BOADICEA is not run.
   10. There is a question of whether it would be possible to offer counseling to people with discordant PRS and GIRA risks.
   11. It is important to be clear in what we’re saying to the patient with “high risk” vs “not high risk.”
       1. There are no other conditions that show this incongruence with the Broad report and GIRA.
       2. One possible solution is to do in-person returns for those with these incongruencies as long as the numbers of affected participants is low. Can consider sending the corresponding PCP an additional letter to explain this.
       3. **Decision**: The protocol will be amended to do in-person returns for those with incongruencies in high risk PRS and GIRA (BOADICEA) for Breast Cancer and offer counseling to those with discordant PRS as long as the numbers of affected participants is low.
       4. Note that this can cause complications with outcomes analysis and this will be discussed further in the outcomes workgroup meetings.
4. **GIRA Return Challenges & Successes | Margaret Harr (CHOP) & Nita Limdi (UAB)** 
   1. Data element availability, data accuracy, and fidelity of the algorithms will be addressed.
   2. UAB implemented a 2-week pause to analyze withdrawals, missing samples, and missing surveys.
   3. UAB found participants that scanned a QR code to enroll did not complete surveys (10%) vs. those enrolled in person (0.9%).
   4. A lesson learned in this was how much work it takes to recover data compared to being able to get data up front.
   5. GIRA generation is working fairly well for most conditions.
      1. BOADICEA for breast cancer complicates GIRA for this condition. Data needed to be imputed for BOADICEA family history components. UAB did a comparison between BOADICEA and CanRisk and found them to be comparable. Examples of cases with different scores were discussed, showing where some glitches may be found if scores are discordant with each other.
      2. Family health history for CKD was not triggering risk initially but has now been fixed.
      3. For CHD family history, onset of disease was not known so the date of onset was imputed to 1 year prior to death. Assumption of age imputed is not noted in the report but it does not impact the risk score.
      4. Participants have trouble discerning between type 1 and type 2 diabetes in the clinical factors.
      5. UAB has returned 175 GIRA reports with 96 pending returns.
      6. ACTION ITEM: The CC will work with Mt. Sinai and the Breast Cancer team to resolve any final questions regarding score generation between the API for BOADICEA and the web based tool.
   6. CHOP has returned approximately 300 reports, 93% of those for children. CHOP uses a 14-point checklist to ensure all data points are collected prior to return. The vast majority of these returns are for obesity.
   7. Most of the “adult” participants are just over 18 years old, which limits data obtained in this age group.
   8. Participants with high risk are initially contacted through MyCHOP (MyChart) in Epic with instructions on how to schedule an appointment. CHOP is able to link the study to a participant in Epic, which also links a lot of pertinent information within. Providers will get a best practice alert and CDS when the participant comes for a well visit.
   9. CHOP can monitor when parents/PCPs have viewed results, opened PDFs, capture CPT codes, etc.
   10. An RoR modeling spreadsheet was developed to project capacity for most eMERGE functions (return capacity, sample plates needed, recruitment goals, etc).
5. **Program Report & Network Timelines | Leadership**

**Robb Rowley (NHGRI)**

1. The NHGRI is thankful for and recognizes the incredible amount of work from the Network.
2. The [Marker](https://pubmed.ncbi.nlm.nih.gov/36621880/) paper was presented at the May Council.
3. Nephi Walton joined the NHGRI as a part-time clinical informatic consultant.
4. Funding opportunities were discussed:
   1. AGMR (Advancing Genomic Medicine Research)
      1. Topics funded include undiagnosed diseases, EHR integrated tools, evidence for clinical and economic value, newborn screening, and promoting access for underserved populations. This is a Notice of Intent to Publish with a plan to publish a Notice of Funding Opportunity (NOFO).
   2. Data Integration and Statistical Analysis Methods
      1. Individuals are invited to use these data sets and analyze in terms of the genomic changes with developmental periods.
   3. Genomics and Health Equity
      1. Notice of Funding Opportunities for [Investigator-Initiated Research in Genomics and Health Equity (R01 Clinical Trial Optional)](https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-23-017.html) and [Investigator-Initiated Research in Genomics and Health Equity (R21 Clinical Trial Optional).](https://grants.nih.gov/grants/guide/rfa-files/RFA-HG-23-018.html) The goal is to increase the equitable use of genomics in clinical care across the US.
      2. Applications are due November 8th, 2023.
   4. Network of Genomics-Enabled Learning Health Systems (gLHS)
      1. This is a Notice of Intent to Publish, not a notice of funding opportunity. The focus is on how to implement evidence into practice.
      2. There is one notice for [clinical sites](https://grants.nih.gov/grants/guide/notice-files/NOT-HG-23-044.html) and another for [coordinating centers](https://grants.nih.gov/grants/guide/notice-files/NOT-HG-23-045.html).
      3. Applications are due around Nov 7th, 2023.
5. The NHGRI is continuing to work with the CC to make sure recruitment projections are monitored and resources are applied across sites where needed.
6. Based on previous projections NHGRI estimated we would be short approximately 1900 participants by June of 2024. With modifications in plans recruitment goals should be reachable.
7. NHGRI will keep track of recruitment numbers to help determine how to best distribute funds. Any sites with supplements should email Robb (copy Jahnavi when 50 of the extra individuals that you are recruiting are recruited.

**Jodell Jackson (Coordinating Center)**

1. Alpha phase began with UAB in December 2022. Beta phase added Mt. Sinai and Columbia in January and then was opened up to the remaining sites in March 2023.
2. There are currently 8 sites in the beta phase.
3. There have been 3 major issues in the beta phase.
   1. These were found through extensive review of codes and trigger logic.
   2. These were centered around coronary heart disease, breast cancer (BOADICEA), and chronic kidney disease.
4. When an issue is added to the GIRA change log Jennifer Morse and the R4 team begin investigating. The CC reaches out directly to condition leads and co-leads to talk through issues and determine a fix.
5. In instances where a fix will cause a change in risk, the CC audits all the GIRAs that have already been returned. The CC works with affected sites and condition leads to determine next steps. Fixes are implemented for current (unreturned) and future GIRA generation.
6. Network emails are sent out once the error has been identified (including summary and scope) and guidance is obtained from condition leads. There has been a very small number of people affected by changes in risk requiring re-return.
   1. Typically, a non-compliance is submitted to the IRB for any GIRA needing to be re-returned with a change in risk status.
7. The CC encourages network members to examine the GIRA error log and the GIRA production log.
8. The R4 code is accessible for site review. This includes links to PCE, BOADICEA, MeTree and Invitae parsing, and location of GIRA trigger logic code. Broad parsing will be included soon.
   1. ACTION ITEM: Sites should review partner and GIRA code and provide any questions to the CC by the end of August.
9. The goal is to move out of beta testing by the end of August 2023.
   1. In order to achieve that goal, the last sites who have not started to return need to start staging and returning GIRAs.
10. The error rate has significantly slowed over the last few months.
11. The CC is looking for suggestions to ensure confidence in GIRA - meetings with condition leads to discuss code and variables, further breast cancer discussions variable mapping between API and online tool, open forums with genetic counselors and study staff, lessons learned from more experienced sites.
12. The issue of missing data on family history was brought up. It was recommended that sites go into MeTree and impute data themselves.
    1. It was suggested that official guidance for each condition would be beneficial.
    2. There needs to be consensus of how family history is used and what to do in cases of missing family history information.
    3. ACTION ITEM: This issue will be taken back to CARE and discussed on an upcoming PI call.

**Rex Chisholm (SC Chair, Northwestern)**

1. Network Timelines - We are currently in the beginning of Year four.
   1. Current recruitment is 12,682 participants, Current GIRA returned is 820.
2. The Network is still committed to last participant enrollment by June of 2024, last GIRA returned by October of 2024, and analysis completed by April of 2025.
3. NHGRI is proposing an additional year (Year 6) that would run from May of 2025 to April of 2026 to collect outcomes data. The final data freeze would be in October of 2025 to allow for additional data analysis.
4. Progress: Currently projected to have 22,500 enrolled participants by the end of June 2024.
   1. Adjustments may need to be made to ensure the goal of 25,000 participants is met.
5. Return progress and issues to address include:
   1. Moving out of the beta phase by the end of August 2023.
   2. Network confidence in GIRA process and return with all sites returning.
6. Currently there have been 1059 GIRA generated and 820 returned across 8 sites.
7. There is a need to address barriers to site return, how to handle need for re-return for GIRA errors, and examine rates of return at September Steering Committee meeting.
8. Margaret Harr will share the tool CHOP used for rates of return projections.

**Action Items:**

* Sites should review partner and GIRA code and provide any questions to the CC by the end of August.
* The CC and CARE group should investigate if a disclaimer box can be added to the first page of both not high risk and high risk GIRA when no family history data is obtained to indicate this risk was not assessed. This should also be explored for monogenics if an Invitae sample is not returned.

1. **Clinical Decision Support & Integration Progress | Emma Perez (MGB) & Eta Berner (UAB)**
   1. The CDS group is a sub-group of the EHRI working group in which sites share what they are doing and that work is tracked and struggling sites are identified and assisted by other sites.
   2. The current goal of the CDS subgroup is overall RoR workflow. The CDS workgroup is working with the EHRI workgroup to track progress across sites. Sites have been asked by EHRI to document full technical workflow.
   3. Examples of data variation collected from sites include where the GIRA lives in the EHR, the type of notification done, and whether people are notified for all results or only some results.
   4. eMERGE network tasks include clinical decision support, network protocol for issues and harmonizations, return progress and lessons learned, and the provider survey (including planned interviews, feedback from providers, and informal interviews by sites while in the beta phase).
   5. Future directions of the CDS subgroup include:
      1. A clarified definition of the CDS subgroup including alerts and notifications, best practice alerts, and education.
      2. Examining processes and the use of automation and barriers to learn from other sites about possible processes and automation.
      3. What to begin thinking about now in eMERGE IV to prepare for sustainability of PRS return. This strongly relates to how monogenic testing may or may not already have CDS in the EHR.
   6. If GIRA and/or counseling notes were opened for outcomes purposes and issues were solved, including amended reports, it would be known which providers actually looked at the GIRA, whether high risk or low risk.
   7. There was limited feedback in the REDCap survey with 8 participants and the highest priority was examining processes and barriers to automation, a definition of CDS in eMERGE IV.
      1. A later priority is the sustainability of incorporating PRS into the EHR and considering if it can go in the lab section or in the media tab.
      2. Another consideration is if sites can determine if the GIRA was opened by a provider.
   8. The collaboration with PRIMED includes the EHRI workgroup, the CDS subgroup, and PRIMED data sharing. The aims include developing a tool that can compute the distance between the populations used in the development datasets vs. the patient itself, the accuracy of PRS with genetic distance, and displays in the report. The collaboration is anticipating input from the GIRA design group.
2. **Polygenic Background Affects the Penetrance of Monogenic Kidney Disease | Atlas Kahn (Columbia)**
   1. Atlas Kahn presented his work on the genetics of kidney disease.
   2. Autosomal dominant polycystic kidney disease (ADPKD) can be explained by two genes PKD1 and PKD2.
   3. 60% of ADPKD patients develop kidney failure before age 60.
   4. COL4A- Associated Nephropathy (COL4A-AN) is another homogeneous kidney disease. 80% of cases are inherited by an X-linked pattern.
   5. The research question posed was if polygenic background is contributing to variable penetrance in common forms of monogenic kidney disease.
   6. The group tested the effects of a genome-wide polygenic score (GPS) for CKD among carriers of ADPKD and COL4A-AN using biobank data.
   7. The CKD algorithm was developed in eMERGE III. The algorithm has three steps. The steps are pre-filtering, G-staging, and A-staging. This allows for flexibility to define cases based on the specific stages. The control is individuals with an eGFR>90 with no CKD-related billing codes.
   8. The GPS for CKD was published last year. The GPS was tested across four major ancestries.
      1. The odds ratio for the top 2% was 3-5, which is equivalent to family history risk.
   9. They tested the GPS as a predictor for the ADPKD monogenic carrier.
   10. There were 600,000 individuals with ES/GS from the UK biobank and All of Us. Through filtration, 206 carriers were found (0.03%).
   11. A PheWAS analysis was performed to show the positive control, that the carrier definition was adequate.
   12. By comparing the GPS tertiles and ORs, the group was able to conclude that accounting for the polygenic background can increase the accuracy of kidney failure in the ADPKD carrier group.
   13. The same exercise was repeated for COL4A-AN. The same effect was seen as in ADPKD when examining the GPS tertiles, but not as high an effect.
   14. The findings concluded that GPS is predictive of ADPKD and COL4A-AN and contributes to variable penetrance, and that polygenic and monogenic risks of CKD are additive.
   15. Optimization was done using the UK biobank African ancestry group.
   16. The one used in the calculation was from the effect of the Apol1 regional variable divided by the effect of the GPS.
   17. *Apol1* has a different frequency in African-Americans in America and those of African ancestry in the UK. Typically in their research, they report the analysis by ancestry.
       1. In this analysis, they did not have the power to perform the analysis by ancestry, so they used the ancestry correction.
       2. The group uses the same method to adjust for Apol1 high risk genotype and use the methods developed by the Broad to standardize the variance and mean by ancestry based on the reference across all the cohorts.
       3. The GPS performance is validated in All of Us.
   18. There is a striking difference in the upper tail tertile of the polygenic risk versus the lower tertile. This degree of difference in risk will have implications for genetic counseling of patients diagnosed with ADPKD.
   19. ADPKD is an autosomal dominant disorder, and the most severe presentations are under the recessive mode of inheritance. The group reanalyzed the data under the recessive mode and saw increased effects, and the sample size drops dramatically with recessive and X-linked coding.
   20. While this is currently in the discovery stage, there is potential for real clinical significance.
3. **Closing Remarks | Rex Chisholm (SC Chair, Northwestern)**
   1. The network saw a bit of the future in terms of having to sort out some issues with GIRA return and the CC has worked hard to make a lot of fixes and resolve code bug errors.
   2. A lot of the topics discussed during this meeting will have continued discussion in the future and the group has a concrete list of action items including what the workflow needs to look like for sites throughout the next year in order to make sure recruitment and completing returns occurs efficiently so the network gets as close to the 25,000 participants as possible.
   3. The majority of sites are using saliva samples but it would be helpful to get updated data for future discussions.

**ACTION ITEMS**

**Conceptual AIs**

* Develop guidance on imputation for missingness across each conditions (Condition leads & CARE/CC to coordinate; **July 31st**)
  + Data Taskforce to focus on assessment of missingness for key data variables (above)
* Develop high level guidelines and documentation for GIRA creation and R4 usage (CC & condition leads for guidance; **due July 31st)**
* Sites to create projections for sample processing and GIRA return rates. (PIs/sites; CHOP to share code; **due July 31st**)

**Data and analytic AIs**

* Phenotyping to lead code redaction collation to be reviewed both within the workgroup and by the PIs. (Chunhua & site leads; Completely signed off by **August 17th)**
* Finalize the [metric](https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.google.com%2Fpresentation%2Fd%2F1HTdWVzWL5cRUNyZaMfQB1pV-RFgd9wBDb1mbtnjSbqI%2Fedit%23slide%3Did.g251a5b4c228_0_1156&data=05%7C01%7Csophie.forman%40vumc.org%7C9f375223a9ee4710efe608db766caeed%7Cef57503014244ed8b83c12c533d879ab%7C0%7C0%7C638233981382535186%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=vy1unIv4IZQdkby99eP0TcujdtZWCJYHSfNza4tTAeg%3D&reserved=0) definitions (Data taskforce & CARE; **Due August 17th**)

**New subgroup:**

* Formation of data analysis subgroup: Genomic Risk Integration Discovery (GRID) (CC/leadership confirm co-chairs: Adam Gordon & co-lead; **Due August 3rd**)
* Establish lead & infrastructure group for Outcomes data analysis leverage eMERGE & AnVIL expertise to demonstrate clinical genomics research on the cloud (AnVIL and eMERGE co-chairs; **Due August 17th**)

**Protocol related AIs**

* Change protocol to clarify FHH only return by mail, and that complicated non high risk return can be in person (CC, included in current amendment, **completed pending IRB**).
* Language changes for male/female on Pca and Bca for GIRA/PRS language (CC, included in current amendment, **completed pending IRB**)
* Language changes for male/female on Pca and Bca Patient education pages (CARE & condition leads; **Due August 15th**)
* Sites should review partner and GIRA code and provide any questions to the CC by the end of August.
* The CC and CARE group should investigate if a disclaimer box can be added to the first page of both not high risk and high risk GIRA when no family history data is obtained to indicate this risk was not assessed. This should also be explored for monogenics if an Invitae sample is not returned.

**Decision**: The protocol will be amended to do in-person returns for those with incongruencies in high risk PRS and GIRA (BOADICEA) for Breast Cancer and offer counseling to those with discordant PRS as long as the numbers of affected participants is low.