Summary of External Scientific Panel/Steering Committee Meeting: October 2021

October 27-28, Zoom & In-Person

eMERGE Day 1: Wednesday, October 27th, 2021

- 9-9:20 am NHGRI program official report | Robb Rowley (NIH/NHGRI)
- 9:20-9:30 am Announcements, opening remarks | Rex Chisholm (SC Chair, Northwestern)
- 9:30-10 am PRS Diversity Consortium | Sally Adebamowo (U Maryland, PRIMED)
- 10-11 am Recruitment and Retention Strategies | Digna Velez Edwards (VUMC), Ingrid Holm (BCH), & Wendy Chung (Columbia)
- Workgroup breakout sessions
    - 11:15- 12:15 pm PRS & Genotyping
    - 11:15- 12:15 pm Provider Uptake & Outcomes
    - 12:55- 1:55 pm Comprehensive Risk Assessment & Return
    - 12:55- 1:55 pm Phenotyping
    - 3:10-4:10 pm EHRI
    - 3:10-4:10 pm R2/sIRB/ELSI
- Scientific Presentations
    - 1:55-2:15 pm Pediatric Asthma Risk Prediction and the Impact of PRS | Lisa Martin (CCHMC)
    - 2:15-2:35 pm Diverse Perspectives on Genetic Study Participation | Courtney L Scherr (Northwestern)
    - 2:35-2:55 pm Scanning the medical phenome to identify new diagnoses following recovery from COVID-19 in a US cohort | V. Eric Kerchberger (VUMC)
- 4:10-4:30 pm Closing remarks | Rex Chisholm (SC Chair, Northwestern)

eMERGE Day 2: Thursday, October 28th, 2021

- 9:20-9:35 am eMERGE Network overview: Priorities, goals, progress and ESP recommendations | Rex Chisholm (SC Chair, Northwestern)
- 9:35-10 am Ancestry Adjustment of PRS Population Variance | Chris Kachulis (Broad)
- 10-10:25 am Prospective cohort: PRS progress | Eimear Kenny (Mt. Sinai) & Niall Lennon (Broad)
- 10:40-11:05 am Recruitment and retention | Digna Velez Edwards (VUMC), Ingrid Holm (BCH), & Wendy Chung (Columbia)
- 11:05-11:30 am Outcomes measurement & comparison groups | Nita Limdi (UAB) & Noura Abul-Husn (Mt. Sinai)
- 12:10-12:45 pm GIRA return | Beth Karlson (MGB) & Margaret Harr (CHOP)
- 12:45-1:20 pm EHRI progress and site dependencies | Luke Rasmussen (NU) & Bob Freimuth (Mayo)
- 1:45-2:15 pm Input/Feedback from the ESP, general discussion
- 2:15-2:30 pm Closing remarks | Rex Chisholm (SC Chair, Northwestern)

eMERGE Day 1: Wednesday, October 27th, 2021

01. NHGRI program official report | Robb Rowley (NIH/NHGRI)

a. Over the last 15 months, the network has made significant accomplishments including identifying and validating 10 PRS, developing a method for calculating tranancestry PRS, and establishing outcomes and analyses plans for the GIRA.

b. The NIH ACMG Fellowship in Genomic Medicine Program Management is designed to increase the pool of professionals trained in managing research and implementing programs in genomic medicine. Up to two qualified clinicians are selected annually to acquire credentials and experience. Applications for the two year fellowship are due annually December 1.
c. The NIH also has a loan repayment program if committed to perform research for two years, $50,000 will be given per year and covers loans and federal taxes.
    i. NHGRI supports this program for genetic counselors as well.
d. There is a new R25 funding opportunity for medical students to support the development of curriculum designed to be freely available at no cost to the community. Application due dates are January 25, 2022 (2023 and 2024).
e. Groups are encouraged to apply for R01 and R21 funding opportunities. The next due dates are August 1, 2022 and March 23, 2023.

02. Announcements, opening remarks | Rex Chisholm (SC Chair, Northwestern)

a. Goals for the day: Confirm we are on track for PRS clinical implementation and identify any upcoming barriers.

03. PRS Diversity Consortium | Sally Adebamowo (U Maryland, PRIMED)

a. Goals
    i. Leverage genetic diversity to develop methods and improve the applicability of PRS across diverse populations and for a broad range of health and disease measures.
    ii. Optimize the integration of large-scale, harmonized genomic and phenotype data to facilitate collaborative analysis, dissemination of PRS-related data, and development of related resources.
b. Diversity first
    i. Prioritize the use of non-EA over EA data to ensure that scientific and clinical analysis can be broadly applied.
c. Study site contributions
    i. Bring existing cohorts to maximize sample size and genetic diversity, address challenges related to data availability, identify and harmonize health/disease measure for analysis, ancestry into analysis, identify metrics for improving PRS prediction, refine PRS based on updated data, participate in consensus approaches to developing and applying PRS, contribute to cross-consortium working groups
d. PRIMED-Cancer is leveraging diversity in cancer epidemiology cohorts and novel methods to improve polygenic risk scores.
e. CAPE study site aims to integrate data from 230,000+ individuals.
    i. Aims: Incorporate admixture in PRS. Principled biobank and cohort data harmonization. Functionally augmented admixture-PRS. Longitudinal calibration
f. CARDINAL study site will bring together 100,000 individuals in african-ancestry populations. The goal is to improve PRS predictions for diverse groups.
g. FFAIR-PRS study site is focused on south asian ancestry populations.
    i. Fine-mapped variants have enhanced transferability for assigning cause of probabilities.
    ii. Incorporating fine mapping and functional annotations improves score performance and transferability.
    iii. Integration of polygenic scores with additional risk factors is critical for deployment.
h. D-PRISM is focused on developing PRS for diabetes in people with diverse ancestry backgrounds. There are over 600,000 participants of diverse ancestry who are a part of the program.
i. EPIC-PRS is focused on blood cells immune response adaptation. The goal is to develop and apply PRS to estimate the risk of common disorders and relate biomarkers to these health disparities.

j. PREVENT aims to develop PRS for coronary heart disease and its major risk factors like Hypertension, diabetes, and obesity.
k. Consortium deliverables: Project datasets with harmonized data and create consensus PRS models: SNPs, weights, and covariates.
l. Tools and resources: Policies and standards to enable data sharing, including ELSI and data and approaches facilitating validation in clinical settings.
m. Timeline
    i. Year 1 -2: Integrate data, convene WGs, harmonize measures, agree on PRS approach.
    ii. Year 2 - 4: Collaborative PRS analyses, Refine models based on updated data.
    iii. Year 4-5: Disseminate results, further refinements based on community input.
n. Additional working groups expected include more analysis and methods-focused WGs, Phenotype-specific WGs, ELSI WG, Publications WG.
o. Possible areas of PRIMED-eMERGE collaboration
    i. Phenotype harmonization
    ii. Analysis of common phenotypes
    iii. Integration of genomics with social,environmental and clinical factors
    iv. ELSI
    v. Community of practice
p. Questions & discussion
    i. Possible ways emerge and PRIME can link up on ELSI efforts. Can have joint meetings to learn from the ELSI group in emerge
    ii. PRIME will work with other groups to disseminate the best methods for sharing access to legacy cohorts.
    iii. The PRS analyses will include risk factors across the lifespan of participants for many of the study groups, like: smoking, drinking, and physical health.
    iv. As of now no sites have a dedicated focus on testing monogenic conditions.

04. Recruitment and Retention Strategies | Digna Velez Edwards (VUMC), Ingrid Holm (BCH), & Wendy Chung (Columbia)

a. Inclusions
    i. The target recruitment population age is between the ages of 3-75.
    ii. Participants will be able to consent in either English or Spanish.
    iii. Participants will have to be able to provide a healthcare provider (HCP), and stay in the area where healthcare is being provided.
    iv. Participants must be willing to accept their GIRA results.
b. Exclusions
    i. Participants that had bone marrow transplants will be excluded from the study due to the possible complications of determining the genetics of the actual participant.
    ii. eMERGE staff and investigators will be excluded from this study.
c. The ROR for asthma and type 1 diabetes is specifically for children. The ROR for the following conditions: atrial fibrillation, breast cancer, chronic kidney disease (CKD), colorectal cancer, coronary heart disease (CHD), hypercholesterolemia, and prostate cancer are strictly for adults. The ROR for obesity and type 2 diabetes are for both children and adults. The PRS will not be returned for colorectal cancer, only monogenic findings. Breast cancer and CHD will have an integrated score returned. It is estimated that approximately 25% of the cohort will have some sort of high-risk return. BOADICEA will be used to return the integrated score for breast cancer.

d. For participants to be fully enrolled, they will need to consent, completed the baseline survey, and have submitted biosamples. The biosamples will be sent to two labs: Invitae for the monogenic risk, and the Broad for the polygenic risk.
e. Results will be sent to MeTree to assess family history risk.
f. Sites are working together to build a REDCap survey so enrollment can be more automated.
g. The sites propose creating a website for the study. The estimated cost of the website is approximately $10,000. Each site may be using different methods to recruit participants.
h. Each site will have different clinically-trained individuals (Research Assistant, Study Genetic Counselor, Study Nurse, Study Physician, Other) returning high-risk results to participants. For most sites, high-risk monogenic scores will be returned by a genetic counselor. There will be educational training sessions for individuals who will be returning results to participants. Some of the material that will be covered consist of: Level of risk by age and sex, Certainty/uncertainty of risk estimates, Screening options by age, and risk reducing options.
i. The Network is taking under consideration supporting participants after their results have been returned that may be under-resourced for their recommendations. For eMERGE education needs, there are three main stakeholders: participants, providers, and the study team. The goals of the eMERGE education sub-group is to develop new materiais, curate existing materials, and review and provide feedback on materials developed by other workgroups. The education sub-group's work has been guided by ELSI studies and patient advisory panels. There are different modes of education, the participant-facing ones being infographics, videos, written materials, site-specific websites, and a potential coordinating center website.
    i. For participant-facing information, a one page flyer, an infographic about the steps of the study, and a study introduction video have been developed and IRB-approved.Materials that will be used across sites are currently undergoing Spanish translation.Other materials that are currently being developed are retention post cards, education infographics, and local REDCap enrollment tracking. The potential participant-facing website will include a place to house education materials, access to the PDF of consents, site contact information, and a link to the site-specific R^4 pre-screen link.
    ii. The study team/provider-facing modes of education are online written materials, self-guided content with and without audio, video tutorials, webinars, EHR integrated DST, a potential coordinating center website, and site-specific websites. There has been discussion regarding what materials are needed for provider education.
j. Discussion
    i. Some of the participant-facing materials are ready to be reviewed and edited by the phenotype leads, however, there is not any phenotype-specific information that is being developed for participant-facing.
    ii. ACTION ITEM: The CC will create a Google Drive folder where sites can store their provider education materials and view other sites' materials.
    iii. Sustainability should be considered when developing a study team/provider facing education materials. Self guided content may be preferred since individuals can complete them at their own pace.
    iv. The network strongly favors implementing a participant-facing website.
    v. Mountain Park found sending postcards to participants monthly helpful for retaining participants.
    vi. The eMERGE website can possibly host the provider-facing education materials.

05. Workout breakout sessions

Notes can be found in the workgroup google docs, linked below for reference.

a. **PRS & Genotyping | PRS minutes**
b. **Provider Uptake & Outcomes | Workgroup minutes**

Action Items:

- Sofia Labrecque will consult the R4 developers and sIRB team to gather more information about options and report to the co-chairs which options are most feasible by November 19th.
- Sofia Labrecque will send out an email inviting Outcomes workgroup members to sign up for and lead the provider outcomes subgroup. The membership and input of clinical providers will be particularly important for this subgroup.

c. **Comprehensive Risk Assessment & Return | Workgroup minutes**

Action Item:

- Sophie Forman will circulate care recommendations to the phenotype leads for confirmation and updates for their conditions.

d. **Phenotyping | Workgroup minutes**

Action Items:

- Next steps include manual chart review data collection and metadata of phenotyping algorithms collection.
- The phenotyping and CARE workgroups will coordinate to confirm data extraction needed for the participant clinical data elements.
- Once the final common variable refresh is released, the CC will update the manuscript concept sheet to reflect the new data elements.

e. **EHRI | Workgroup minutes**
f. **R2/sIRB/ELSI | Workgroup minutes**

06. **Scientific Presentations**

a. **Pediatric Asthma Risk Prediction and the Impact of PRS | Lisa Martin (CCHMC)**

i. The Asthma Predictive Index (API): The API is the most widely used and validated index. It requires early wheeze and has a high specificity (96%), but relatively low sensitivity (28%). It is useful for predicting which children will not develop asthma, but it misses many children who will go on to develop asthma.

ii. Asthma in Childhood: Asthma causes wheezing, difficulty breathing and coughing and affects approximately 1 in 12 children. It also causes significant morbidity (50% of asthmatic children have attacks which can lead to ER visits or hospitalization) and negatively impacts quality of life (physically, emotionally, and educationally). Risk for asthma includes family history of atopy, early wheeze, and atopic dermatitis.

1. Major Challenge: Identifying kids who will go on to develop asthma. If we can reliably predict which kids will go on to develop asthma, then we can focus on prevention methods, but tools are needed.

iii. Conclusions from the development and testing of the Pediatric Asthma Risk Score (PARS), created with the intention of better predicting who will develop asthma (relative to the API) are listed below.

1. The PARS model, developed and tested in CCAAPS (Cincinnati Childhood Allergy and Air Pollution Study), showed improved performance in identifying at-risk children when compared to the API. This was evident in children of intermediate risk.
2. These results were replicated in multiple cohorts, which suggests the model is robust.

3. Children of intermediate risk are hardest to detect but may be the most amenable to modification of risk factors (i.e., best targets for intervention).
4. While the PARS is an improvement over the API, there is still room for improvement – inclusion of PRS should be evaluated.

iv. Discussion of presentation
1. Currently, PARS is not integrated with the PRS as part of eMERGE, though there is a goal to be able to test integration in the future.
2. eMERGE participants who are at high risk for asthma will receive a link to a website where they can calculate their PARS score in the GIRA packet and their clinical risk factors will be pulled from the EHR. Clinical risk factors were originally planned to be collected in the baseline survey, but the questions were not added before the survey was finalized and translated.

b. Diverse Perspectives on Genetic Study Participation | Courtney L Scherr (Northwestern)

i. Focus groups were held on zoom with at least one genetic counselor.
ii. Below are questions/thoughts given by the prospective participants.
iii. How well did they understand the GIRA: Most seemed to understand the gist of the presentation. Many had logistical questions about the procedures of the test. Participants tried to make sense of the GIRA; many asked about the difference between GIRA and other familiar risk assessments.
iv. What would motivate participation: Personal or family history of disease (also a hesitation; too much to handle) and genetic advancements.
v. What are some hesitations that participants might have: Doubt of the development of the science and worries about the "dark side".
vi. Concerns for data use and security: Wanted to ensure that their info would be secure. Some did not want insurance companies having access to their info. Less likely to join study if a third party was involved.
vii. Participant expectations for consent: Expected to see if they were signing to have their data shared with the desire of not being blind sided by the sharing of their data.
viii. Concerns that Latinx have about recruitment: Latinx said that prevention is not valued in their culture so recruitment can be difficult. Information from Gira might be ignored; meaning recommended life changes will not be made by the participants. Need to use trusted sources of information (few use the internet in latinx communities). Latinx individuals want to see and hear about our efforts to recruit from the community.
ix. What support do they anticipate needing if they are identified at high risk: They want to ensure efforts are made so that they fully comprehend the diagnosis. Interpersonal clinical support after diagnosis. Help coping with their new diagnosis, they do not want to be forgotten after receiving this info.
x. Results disclosure: Want to receive results from someone knowledgeable. Latinx said it is essential to have someone that speaks the patient's native tongue. Want long term support (not just a single report and meeting).
xi. What can we do with this data for social justice: Make policy recommendations that can help reduce health disparities.
xii. Take-home points from the presentation:
1. Participants want to be clearly and fully informed.
2. Participants want to be supported during ROR.
3. Participants are concerned about social implications.

xiii. Questions & discussion:
    1. Want to create more than one point of contact for participants; for example a participant facing website would be helpful. This website will need to be in spanish. Navigation of the system will need to be super clear for next steps and other support.
    2. Ideas on how to give emotional support for participants. Ideally GCs are trained in emotional support techniques for patients. Need to be able to calmly talk patients through the next steps after receiving a result.

c. Scanning the medical phenome to identify new diagnoses following recovery from COVID-19 in a US cohort | V. Eric Kerchberger (VUMC)

i. Post acute COVID-19: Post acute COVID-19 is when individuals develop persistent problems approximately four weeks after their acute illness. There are over 200 million individuals at risk for having prolonged or new symptoms after having COVID-19.

ii. Objective: The objective of this study is to assess new symptoms and medical conditions developing among COVID-19 survivors (vs. never infected), and survivors of severe COVID-19 after recovery from the acute illness period. Another objective was to access post-acute COVID-19 conditions in VUMC over regular intervals.

iii. Methods: The VUMC COVID-19 registry was used for this study. The registry captures individuals that were tested for COVID-19 at VUMC or VUMC-affiliated clinics. The diagnosis codes were extracted from the EHR and mapped to the PheWAS code system. PheWAS software was used to assess the association between COVID-19 and the entire medical phenome. The post-acute phenotypes were defined as 30 days after a patient has first tested negative after having COVID-19, or 30 days after a patient has been discharged from the hospital due to COVID-19.
    1. The phenotypes that were excluded in this study were patients that developed phenotypes prior to their COVID-19 testing, phenotypes that developed during the acute illness phase, and during a hospitalization.

iv. Study Cohort: The cohort consisted of 153,569 adults that were tested for COVID-19 between March 9, 2020 and May 1, 2021. A total of 4.859 patients were excluded from the study. 2,830 patients died within 30 days of their first COVID-19 test. 2,020 patients reported having COVID-19, but did not have a positive COVID-19 test. Nine patients had poor data quality. Approximately 16% of the cohort tested positive for COVID-19. The cohort was majority white non-hispanic (70%). The median age of the cohort was 46 years old.

v. Results: The PheWAS data method was determined to accurately capture phenotypes associated with clinical acute COVID-19. It was found the COVID-19 survivors had increased odds for 24 clinical phenotypes, 5 being circulatory related, 3 respiratory related, 3 neurological related, 2 endocrine/metabolism related, 2 digestive related, and 2 genitourinary related. It was found that survivors of hospitalized COVID-19 increased odds for 18 clinical phenotypes, 6 being respiratory related, 4 circulatory related, 2 endocrine/metabolism related, and 2 related to other symptoms such as muscle weakness/fatigue. Alopecia and loss of smell/taste was detectable by August 2020. Assessments that were done before 2021 were limited by smaller sample sizes.

vi. Limitations: The data is limited to VUMC. The sample population was limited to the American Mid-South The diagnosis codes may not fully describe the spectrum of COVID-19 symptoms. The COVID-19 variants were not prevalent in TN until mid-June 2021. False-positive COVID-19 tests and re-infections were not accounted for during this study.

07. **Closing remarks | Rex Chisholm (SC Chair, Northwestern)**

a. Some are concerned about final GIRA presentations and that PRSs are still being finalized. EHRI will report tomorrow on progress to date.

eMERGE Day 2: Thursday, October 28th, 2021

08. eMERGE Network overview: Priorities, goals, progress and ESP recommendations | Rex Chisholm (SC Chair, Northwestern)

a. ESP Meeting Goals
   i. The network is looking for advice and input from the ESP. There has been progress in ancestry PRS implementation, and the network is working diligently on the GIRA.
   ii. The GIRA format is a complex system for data generation sequence, return, and reporting.
   iii. The network is currently deciding on next steps needed to finalize outcomes analysis plans and comparison groups.

b. GIRA and Risk Management
   i. The underlying goal for eMERGE is to understand how the GIRA impacts risk management.
   ii. To understand the goal, eMERGE will complete multiple tasks including recruiting from an unbiased general population age 3-75 in English and Spanish, emphasizing underserved populations, and generating trans ancestry PRS on 10 conditions. Additionally, eMERGE will measure participant and provider outcomes six months post return. The outcomes include whether or no conversations were had with medical providers, changing lifestyles, etc.
   iii. The timeline includes first enrollment to begin in January 2022 and first target return in June 2022. The last participant enrolled is planned for January 2024 and the last GIRA return for May 2024.

c. There are 10 eMERGE conditions which will be included in the return. Colorectal Cancer will only be assessed for monogenic sequencing risk. Most conditions are adult and 4 are pediatric (Asthma, Obesity, Type I and Type II Diabetes).

d. Progress & Updates
   i. The sIRB was approved in July 2021 and 7 site reliance agreements have already been approved.
   ii. Consent part I and II surveys, participant FAQs, education materials, and recruitment and retention materials are all finalized.
   iii. All PRS scores have been transferred to the Broad and the Broad is finalizing implementation of clinical pipelines.
   iv. The GIRA and PRS report is finalized with good harmonization of text. The GIRA subgroup has initiated a knowledge base to centralize all logic and data needed for GIRA generation.
   v. Finalized consent and surveys have been built in R4 and many sites are looking into eConsenting.
   vi. EHRI initiated additional monthly meetings focused on clinical decision support.

e. The CC has worked hard to keep the network on track for recruitment. Dashboards, trackers, and timelines help with organization and decision logs with risk mitigation charts highlighting rationale for protocol decisions.

f. Goals for the next 6 months include reviewing the full GIRA models at the Steering Committee level, estimating the number of participants to receive high-risk results, and finalizing data flows from R^4 for integration into local databases and EHR, among more goals. The network is aiming for around 25% of 25,000 participants to receive high-risk results.

g. Challenges for the next 6 months include development and testing of GIRA generation software, dynamically pulling structured data into REDCap and determining integration options for reports to ensure appropriate providers are alerted to results.

09. Ancestry Adjustment of PRS Population Variance | Chris Kachulis (Broad)

a. PRS distributions differ across ancestries. The same raw PRS score may indicate high risk for an individual of one ancestry and low risk for an individual of another. This is often handled by comparing individual scores to only others of the same ancestry. This creates other issues in that there is significant diversity within ancestry groups, so the Broad takes a different approach introduced in a paper by Amit Khera.
b. The adjustment writes the mean as a linear model of an individual's N PCs and variances and alphas are fitted assuming the variance is consistent over the entire population. This approach handles individuals of mixed ancestry, accounts for diversity within ancestry groups, and it is completely genetic so there is no need for self-reported ancestry.
c. When applied to the CAD score, for example, it very nicely results in a standard normal distribution and all populations overlap with each other.
d. Sometimes variance can be population dependent. In Hypercholesterolemia (using 1000 Genomes) for example, when variance is based on an entire population, East Asian individuals will very rarely be labeled "high-risk" (defined as z-score > 95th percentile).
    i. A potential solution to this is modeling the variance as ancestry dependent in addition to the mean. When applied to Hypercholesterolemia, the issue is largely corrected.
e. The adjustment was applied to 7/10 eMERGE conditions and improvement is consistent across all conditions.
    i. A condition like Asthma which worked well before this method was applied does not change much. A condition like BMI does change.
f. A subset of the eMERGE I-III data was then used as a reference population to fit the model to a random selection of 500 Asian, 500 Black/African American, 500 White, and 500 Hispanic. This stimulated the option of using initial samples from the prospective cohort to create a reference population. The reference population could then be used to apply the adjustment to the full eMERGE I-III dataset.
    i. Analyzing how large the training data needs to be, the Broad started with 40 individuals per bucket and increased to 1000 while looking at how well it performed (10 times total). The objective function of fit can then be applied to the entire eMERGE dataset.
    ii. The suboptimal fits perform reasonably once the bucket size reaches about 300. Even the worst fit performed well once 300 were reached. Median fits perform quite well once the bucket size reaches about 300 as well.
g. Performance of different ancestries are independent of sizes of other ancestry buckets. Each bucket needs 200-300 samples and one bucket cannot be smaller.
h. There is missingness in 1000 Genome data that is on the GDA chip. Additionally, there are biases in whole genome data versus GDA data. For example, when eMERGE III data is projected into PC spaces and compared to 1000 Genome data, they do not project on top of each other.
i. If a project was using the same chip in eMERGE, it could use the eMERGE's first 300 cohort to do this adjustment and vice versa would be true if someone else had a publicly available cohort on the GDA chip, eMERGE could use that. GDA is relatively new and started as the All of Us array.
    i. **ACTION ITEM**: A dataset for normalization from All of Us consortium could be used on the same array in the same ancestry groups, it could avoid delay for eMERGE in trying to make up some numbers in various ancestry buckets. The network is reaching out to All of Us for discussion.
j. Using self-reported ancestry for the buckets, and correcting for mean and variance, could be analyzed. For CHD, this correction was tried using the 1000 Genomes and found the mean and variance correction method was not adequate. Using self-reported ancestry and computing mean and variance in the controls seemed to perform well. For African American, the top 5% was at higher risk than using the projection method.

k. The field is constantly changing and the network should be planning for an adjustment method that is broadly usable rather than planning specifically for eMERGE IV. However, for the eMERGE project, there may have to be trade-offs to gain robust trans ancestry scores.

10. Prospective cohort: PRS progress | Eimear Kenny (Mt. Sinai) & Niall Lennon (Broad)

a. The 10 selected conditions have differing numbers of SNPs and are at different stages of validation. All scores use more than one population and some are considered trans ancestry scores (the score is optimized across population groups using algorithmic methods or fine mapping approaches). For four of the 10 conditions, those scores were generated and available for use.
b. A lot of the most recent work has been around bringing cores out of the genomic research setting and putting them into a real world clinical lab setting.
c. The PRS workgroup has been using a GRID as a tool to facilitate transparency and adherence to community standards for PRS validation. This GRID includes pipelines and multiple metrics established by the PGS Catalog and ClinGen common disease working group.
d. Together with the genotyping workgroup, the PRS group has been evaluating models including PC and other non-genetic covariate adjustment. The groups are aiming for clear understanding and consistency across conditions.
e. The PRS group is creating a subgroup that has input on language on reports and supporting documents particularly around reporting metrics and population labels issues.
f. Continuing work in the PRS and genotyping workgroups includes developing flagship paper concept sheets that focus on experience of developing clinical PRS for the project.
g. Genetic ancestry is not being reported on the GIRA to participants.
h. The most recent work from the genotyping workgroup has been working with the PRS workgroup to transfer, implement, and validate scores at the Broad. Sites sent SNPs and weights files to the Broad which was implemented in the cloud-based WDL code pipeline (to make the process CLIA compliant). The implementation of a new adjustment method was used and condition specific performance was measured.
i. Validation packages need to be separately submitted to New York State which requires specific assay approvals.
j. For all conditions, validation progress has been recorded. Currently two conditions, Chronic Kidney Disease and Type I Diabetes, still have work in progress but no issues are anticipated. Type I Diabetes has a slightly different model for implementation so the code needs to be changed. CKD needs an additional test implementation of APOL 1 risk estimate.
k. From the beginning, the network was not limited to validated and published scores (several of the PRS being used are from large consortium groups that are not out yet) because of the possibility of a better PRS being developed during the study, the network was not limited. Many scores were tailored to the outcomes of interest and diversity within the cohort. The network is not in a position to change scores after they have been implemented.
l. Data can be submitted to the Michigan server and have it imputed against TopMed but TopMed data cannot be accessed to create a reference panel for our own imputation pipeline.

11. Recruitment and retention | Digna Velez Edwards (VUMC), Ingrid Holm (BCH), & Wendy Chung (Columbia)

a. Participants will be identified by several different methods including, but not limited to: Mass email and portal messaging, posters and other recruitment materials, social media.
b. ELSI has informed the study by providing input on ancestry-specific risk and the informed consent document. ELSI has also provided input on the GIRA report language and the various surveys being utilized.

c. Education subgroup: Several modes of participant education have been explored, including: video, written materials. This subgroup is also working on provider education. Several of the same modes from participant education will be utilized for the providers. Strategies on provider education and support were discussed including offering CMEs is one method to possibly have higher provider participation and just In Time materials might be helpful.
d. A clinical decision support (CDS) subgroup will be formed to develop and implement decision support for clinicians.

12. GIRA Return | Beth Karlson (MGB) & Margaret Harr (CHOP)

a. GIRA return categories: There are three categories for return: high genetic risk (in-person RoR), high family/clinical risk (RoR electronically or by mail), and not high risk (RoR electronically or by mail).
b. GIRA risk categories: The risk categories include polygenic risk scores, monogenic gene sequencing, family history, and clinical data when applicable to the given condition. These categories come together to create a high risk GIRA for the different conditions. Breast cancer will use an integrated score, and CHD will use the pooled cohort equation along with the PRS.
c. GIRA packet components: Different components of the packet have different functions. The packet was designed this way, in part, to make portions easily interchangeable/customizable according to the risk category each participant falls into in addition to customizations for other factors (e.g., age).
    i. Patient education pages in the GIRA packet will be translatable into Spanish, but as the report itself is targeted to providers the majority of the text will be provided in English.
d. GIRA format: Contents have been designed to present information in a logical and digestible manner for both the participant and the provider to review. Participant risk for monogenic, polygenic, family history, and clinical factors (when applicable) are displayed in a summary in addition to a general summary of care recommendations displayed for the provider on the first page.
e. GIRA Return presentation discussion: The GIRA logic is being built into R4 to populate the PDF reports and is scheduled to be ready for launch in June of 2022. Educational information for participants that are not high risk can be found on the summary page (sample summary page; condition-specific summary text) and on the FAQs pages.

13. Outcomes Measurement & Comparison Groups | Nita Limdi (UAB) & Noura Abul-Husn (Mt. Sinai)

a. Study highlights and overview
    i. The study is designed to be a pseudo-randomized regression discontinuity design (RDD); because some participants will change their behavior independent of the results they receive, the study is not a clear cut regression discontinuity design.
    ii. Outcomes will be measured approximately six months after the return of results (RoR).
    iii. There are three categories of outcomes: process, intermediate, and clinical. Generally, the primary outcome of interest is the adoption of the recommended interventions.
    iv. UAB completed power calculations for the study and accounted for the variability in RDD in their calculations. The Outcomes Workgroup will work closely with the PRS Workgroup to account for overlap in high risk between conditions (e.g., one participant receiving high risk results for more than one condition).
b. Outcomes Measurement and Comparison Groups presentation discussion
    i. Previous iterations of eMERGE have not had 50% male participants; there is typically an overrepresentation of females (approximately 60% female and 40% male), so the actual numbers may shift slightly.
    ii. In study planning thus far, intervention uptake and adherence has been conceptually viewed as a dichotomous outcome, though most of the phenotypes will have multiple suggested

interventions. The power calculations were completed with the operational understanding of participants implementing any of the recommended interventions.

iii. Specific interventions and the level of adoption will be considered in analysis. It was suggested that an integrated, continuous adoption score that would take multiple interventions into account may be a good approach to measuring intervention adoption.

iv. eMERGE's use of RDD is novel in that it will address 10 conditions and multiple outcomes. Typically, studies that use RDD address one condition and one outcome. There is ongoing research to determine what percentile below the high risk threshold would be the most appropriate cutoff for the comparison group for eMERGE's application of RDD; particular focus is being given to reducing bias in threshold selection for each phenotype and across phenotypes.

v. Individuals who are not high risk will not be provided information regarding interventions as part of the study. Behavior and lifestyle changes will be assessed via survey and clinical outcomes will be assessed via EHR so that the rate of intervention adoption by not high risk participants can be factored into the effect size calculations for high risk participants.

vi. Prevalent diseases can be accounted for and removed in the analysis process as deemed appropriate at a later date. Power calculations can also be redone to account for the removal of prevalent disease(s).

vii. Sites are considering having coordinators help participants complete the perceived risk and personal health history section of the baseline survey since beta testing at UAB suggested that participants had a difficult time understanding and accurately completing the section.

viii. The Network has yet to specifically discuss how it would handle discrepancies between health history information in the EHR and self-reported health history information. Preliminarily, it is unlikely that the Network would make any changes to the EHR based on self-reported health history, though the Network will discuss this further in the future.

1. ACTION ITEM: The Outcomes group should consider how to handle data discrepancies between EHR and survey outcomes and include this in the detailed analysis plan they will present to the ESP in April 2022.

ix. After assessment and discussion, the Network decided that measuring outcomes at six months post RoR would be the best design. The Network recognizes that there will inevitably be some instances in which participants have not yet had a follow-up discussion with their PCP regarding their results by the six month time point. While the post-RoR survey will be administered at approximately six months post-RoR, information from the EHR could theoretically be pulled multiple times and at time points further from RoR, though there are timing and logistical constraints.

14. EHRI progress and site dependencies | Luke Rasmussen (NU) & Bob Freimuth (Mayo)

a. Progress that has been made since the last ESP meeting was displayed, including: Network requirements for data, workflow, and EHR integration. Specifications for PRS, monogenic, and GIRA were defined.

b. Four distinct reports need consideration for generating the GIRA. These reports have eight possible artifacts that need to be considered in how/if and where/when they will be integrated at the local site.

c. The structured file format (JSON) for the PRS report has been demonstrated. Return will be pushed from Broad to the R^4 portal. Sites will pull the result from the R^4 portal.

d. The monogenic report will be in the HL7 v2 format. Results will be pushed from Invitae to the R^4 portal. Sites can pull the results from the R^4 portal and/or download from the Invitae portal.

e. The MeTree structured report format is JSON. MeTree will push results to the R^4 portal. Summary data will be incorporated into the GIRA. The full data set can also be pulled from the R^4 portal

f. The GIRA report will be in PDF and a yet to be determined structured report.

g. GIRA Integration dependencies: The GIRA content and how to use it in a CDS was discussed. A data dictionary will be common through the network but how the site implements the CDS will be site-specific. Work is currently being done on structured format, order process, result process, and measured outcomes. To meet immediate network goals, sites need to know what their available data elements are for EHR integration.
h. While vendor engagement may be beneficial, the goal should be to have generalizable standards that are broadly supported by multiple vendors.
i. GIRA integration and decision support integration will need to be ready by May/June 2022.

15. Input/Feedback from the ESP, general discussion

a. The ESP is impressed by the amount of progress made over the past six months and the project appeared to be on track.
b. The transancestry PRS adjustment presentation was very informative, and the Network should continue in that direction.
    i. There are logistical issues associated with recruiting 400 individuals in each of the four ancestry groups prior to proceeding. The ESP is in favor of finding another way to perform the transancestry PRS adjustment, which was suggested to use All of Us data.
    ii. Rex Chisholm and Niall Lennon already began contacting All of Us and the data access seemed likely to occur within the next few weeks.
    iii. If the All of Us data could not be accessed, the ESP suggested using data from a different source.
c. The ESP sought clarity on the PRS report being incorporated into the GIRA report, even though it is intended for the clinician.
    i. The Network had previously discussed extensively the question of clinicians receiving only the high risk/not high risk designation versus the clinician knowing where the participant's PRS falls. There is a resource constraint regarding returning detailed results.
    ii. The return language has been carefully crafted to avoid implying that not meeting the study cut off means that the participant is at low risk.
    iii. The post-RoR provider survey should explore the provider thoughts on this.
    iv. The decision to make the report language provider facing, instead of participant facing, was to ensure providers are receiving what they need to best advise their patients.
    v. The GIRA is being returned as a packet and includes multiple attachments, such as the Broad report and Invitae report.
        1. The Broad report language is much more technical than the GIRA summary. Participants will receive them both in the packet. The GIRA summary is more simplified and accessible.
d. eMERGE must decide on what the outcomes are and how they are going to be defined soon.
    i. There are many downstream decisions and logistics that stem from the outcomes being collected.
    ii. The ESP expressed concern regarding using dichotomous outcomes. Participants acting upon any of the care recommendations counting as adoption will lead to a high rate of adoption in control groups. The ESP encourages the Network to consider how to define endpoints.
        1. ACTION ITEM: The Provider Uptake & Outcomes workgroup must prioritize finalizing the outcomes, their measurements, and comparator groups.
        2. ACTION ITEM: The Outcomes group should define what 'adoption' of outcomes means for each of the conditions.
            a. The PI GIRA presentations and the Outcomes by Phenotype document have assisted in the process of this.

3. The ESP recognizes that this would take effort, but a combinatorial endpoint that takes into account interventions for each phenotype would be ideal.
4. A suggestion was made to move lifestyle adoption into one dichotomous variable, and have clinical outcomes be separated.

iii. Once the desired outcomes are decided, the Network must ensure that the chosen comparator group will yield that information.
1. The main comparator groups would be individuals having a high risk PRS for any condition compared to individuals who did not have high risk for any of the conditions.
a. There was previous discussion on using individuals not at high risk for condition X, but at high risk for condition Y, as a comparator group for condition X. This was decided against due to the belief that individuals who are at high risk for any condition may adopt lifestyle changes that apply to many conditions.
2. The high risk PRS is tied to a risk reducing recommendation that is placed in the GIRA, and the adoption intervention would be compared in those groups.
3. Many risk reducing interventions are standard in primary care practice. The question is if there will be a difference in individuals who have heard they now have a high PRS risk for a certain condition. The Network can look at characteristics within the high risk group that are predictors of intervention adoption.
4. ACTION ITEM: The Provider Uptake & Outcomes group should present a final analysis plan and study design at the April 2022 ESP meeting, including how data discrepancies will be handled.

iv. In type 2 diabetes, PRS is the only risk category that will trigger a high risk GIRA return. For individuals with high PRS, family history risk and clinical risk will be displayed on the report to aid the physician providing care. If the participant does not have a high risk PRS for type 2 diabetes, the family history risk and clinical risk will not be returned. The assumption is that primary care providers are already paying attention to family history and clinical factors. The study is examining the differential addition of PRS.

v. Few phenotypes are returning high risk results based on family history. Those phenotypes use care recommendations directly related to family history risk.

e. The ESP appreciated the EHRI presentation and detailed explanations.
i. The ESP encourages the Network to continue to work toward using HL7 and FHIR.
ii. The EHRI workgroup will continue to support sites as they implement local processes for ingesting data and linking to CDS. The workgroup is also working in parallel to explore how the data will be represented in the GIRA using HL7 and FHIR in the long term.
iii. Although it is not likely that a FHIR based solution will be able to be implemented for the purposes of the Network, the workgroup can still make progress toward that goal.

f. The ESP also recognized the terrific progress made in the recruitment and education materials, and the outstanding work of the PRS development group.

16. Closing remarks | Rex Chisholm (SC Chair, Northwestern)

a. The contribution that the ESP has historically and is continuously made to eMERGE is greatly appreciated and helpful.

b. The network has a lot of work to do and being how difficult the work is, the progress is a testament to the hard work everyone has put in.

c. The recruitment piece will not be trivial but hopefully work will get easier.

d. The next Steering Committee meeting will take place on February 2nd and 3rd, 2022 in Bethesda.

Action Items

Coordinating Center:

- The CC will create a Google Drive folder where sites can store their provider education materials and view other sites' materials.

Provider Uptake & Outcomes:

- Sofia Labrecque will consult the R4 developers and sIRB team to gather more information about options and report to the co-chairs which options are most feasible **by November 19th**.
- Sofia Labrecque will send out an email inviting Outcomes workgroup members to sign up for and lead the provider outcomes subgroup. The membership and input of clinical providers will be particularly important for this subgroup.
- The Outcomes group should consider how to handle data discrepancies between EHR and survey outcomes and include this in the detailed analysis plan they will present to the ESP in April 2022.
- The Provider Uptake & Outcomes workgroup must prioritize finalizing the outcomes, their measurements, and comparator groups.
- The Outcomes group should define what 'adoption' of outcomes means for each of the conditions.
- The Provider Uptake & Outcomes group should present a final analysis plan and study design at the April 2022 ESP meeting, including how data discrepancies will be handled.

Comprehensive Risk Assessment & Return:

- Sophie Forman will circulate care recommendations to the phenotype leads for confirmation and updates for their conditions.

Phenotyping:

- Next steps include manual chart review data collection and metadata of phenotyping algorithms collection.
- The phenotyping and CARE workgroups will coordinate to confirm data extraction needed for the participant clinical data elements.
- Once the final common variable refresh is released, the CC will update the manuscript concept sheet to reflect the new data elements.

PRS/Genotyping:

- A dataset for normalization from All of Us consortium could be used on the same array in the same ancestry groups, it could avoid delay for eMERGE in trying to make up some numbers in various ancestry buckets. The Network is reaching out to All of Us for discussion.

Recordings:

Recording Name	Links	Passwords
Main room 10/27 (includes PRS/Genotyping, CARE, EHRI breakout sessions):	https://zoom.us/rec/share/XhCAzOZPoz8U2W_UIcEW5zDJ7tcOvqON6uO8GnvoWVEFgmlSCiD2JH9sdFW0ehLX.5DLo6fMcYWhlC0h4	bEv%=nX3
Provider Uptake & Outcomes breakout session:	https://zoom.us/rec/share/1BbKc-OBpR_uRuGClVFeHVOp-0GvamO6lQ8KkdtUsvu5IreMSTr--Bwz4QrYrMaa.24H8HVPQ3gbL1_Df	*4C$.Gew
Phenotyping breakout session:	https://zoom.us/rec/share/3DeWQe5rEruRB6XZbWaLd-XDc9piPGPeH_UJJg07-26QwGybQUf4Bb5Ne5VsSxhf.EhiylG8M5O_G6mLR	X&z5q$^V
R2/sIRB/ELSI breakout session:	https://zoom.us/rec/share/e4bskFfPIByp1Khh2dSUpP8EuO_WXDSulqPtrVPAvjHyGFJppOPtotAAZJRHeBx7.HTfGqPCP9xisyoPc	Q7q4yC??

Main room link 10/28:	https://zoom.us/rec/share/EzXD0N0GW5W631tqLBiLkGFLYSteWAnZGSaLNT9omYzOE6Fz-7gpveMhX-8_Op3w.iAdpMJD9pPFsO5JN	KkD9I3*Q

Official ESP Recommendations

Meeting Summary
eMERGE Network- External Scientific Panel and Steering Committee
Executive Session- 10/28/2021

ESP	Dan Rader, University of Pennsylvania – Chair Kimberly Doheny, Johns Hopkins University Stanley Huff, Intermountain Healthcare Janina Jeff, Illumina Lisa Parker, University of Pittsburgh Clesson Turner, Uniformed Services University of the Health Sciences John Witte, Stanford University	**NHGRI**	**Dave Kaufman** **Teri Manolio** **Robb Rowley** **Baergen Schultz** **Rene Sterling** **Ken Wiley**

The External Scientific Panel (ESP) met with NHGRI program staff members during the executive session of the eMERGE ESP and Steering Committee (SC) hybrid meeting held on October 28, 2021. Overall, the ESP was extremely impressed with the Network's activities and progress in the last six months. The investigators' presentations helped address the ESP's questions about the purpose and intent of the PRS report, overarching plans for EHR integration, and the approach for trans-ancestry PRS development. Recognizing the tight study timeline, the ESP provided observations and recommendations about several aspects of the study to help move efforts forward. Feedback is elaborated upon in the summary below.

Ancestry Adjustment of Polygenic Risk Score (PRS) Population Variance

The ESP was impressed with the proposed trans-ancestry PRS approach but was concerned about logistical challenges created by the need to recruit 300 participants in each population descriptor before being able to return a GIRA report. The ESP was strongly in favor of using existing data to estimate the adjustment method prior to enrollment of patients. Data from *All of Us* would be ideal and it sounds like it may be available; even if not, other existing sources of data should be considered.

Recruitment and Retention

The ESP commended the Recruitment, Retention, sIRB and ELSI workgroups' efforts to date. They appreciated that site-specific ELSI projects continue to inform the research framing and approach and highlighted the well-developed education materials. The ESP had no specific recommendations related to this area.

PRS Progress and GIRA Return

The ESP understood the Network's decision to categorize risk dichotomously due to current limitations in understanding how to interpret PRS in the middle of the risk spectrum. These findings contrast with a few of the ELSI study findings that

suggest that patients and providers prefer continuous risk scores. One option the Provider Uptake and Outcomes Workgroup could consider is incorporating a research question about how a continuous PRS risk score could best be implemented. These findings could inform future research on PRS implementation. There is also an ELSI research opportunity to consider applying for outside funding to examine how people understand, and behaviorally respond to, "not at high risk," including in light of their own and familial health history.

Outcomes Measurement and Comparison Groups

Now that the list of 10 conditions has been finalized, the ESP expressed the need to focus significant attention on aligning the research questions to the necessary outcomes. The ESP stressed the importance, scientifically and ethically, of ensuring that survey instruments clearly define outcomes and gather reliable and valid data to measure the research question being proposed. The ESP recommended that the Network:

- Define how the research questions align with the corresponding outcome measures;
- Define the appropriate comparison group(s) for addressing the main research questions as well as any secondary studies;
- Decide how to address discrepancies of outcomes when they are measured by different instruments (e.g., EMR vs. survey);
- Determine how the definition of patient uptake (adoption of any risk-reduction recommendations in the GIRA) may bias and inflate the measured levels of PRS adoption;
- Determine how to separate the collection of health history and risk perception to avoid biasing a participant's perception of their risk;
- Decide what to do with the results of the beta testing of the survey instruments that suggest that revisions are needed.

The ESP recognized the immense amount of work involving the above but also stressed the time sensitivity of bringing these issues to resolution.

EHRI Progress and Site Dependencies

The ESP was impressed with the EHRI workgroup's presentation and acknowledged the challenge of developing Health Level Seven (HL7) and Fast Healthcare Interoperability Resources (FHIR) standards. Realizing it is unlikely to implement a FHIR based solution for the purposes of this Network, the ESP encouraged the EHRI workgroup to continue establishing data elements for the GIRA which can then be translated into a FHIR standard once utility is demonstrated, similar to what was done in previous phases of eMERGE.

ESP Recommendations for the Network:

1. The Network should pursue using existing data, such as All of Us data, to quickly complete estimation of the trans-ancestry PRS adjustment model.
2. The Network should consider incorporating a research question addressing patient and provider understanding of the risk information in the PRS and GIRA and how it might be improved.
3. The Provider Uptake and Outcomes workgroup should structure the participant and provider surveys and the proposed comparison groups around specific research questions and align with outcomes of interest.
4. The Network should clearly define how they will measure adoption of risk reduction recommendations. One possibility is combining risk-reducing behavior outcomes into a single measure for each condition.

5. The Network should decide how to deal with possible discrepancies between self-reported health history and health history information found in the EHR.
6. The EHIR workgroup should continue exploring ways to represent data in the GIRA using HL7 and FHIR standards.
    a. Returning results into current EHRs is challenging and will require the Network to be pragmatic in how they approach this task.
    b. Making sure that any needed LOINC and SNOMET CT codes are submitted for creation will be a great benefit to everyone that follows.
    c. The use of FHIR would be preferred but may not be feasible given the timeline. HL7 V2.X may be the best option.