****

**Summary of Steering Committee Meeting: Winter 2020**

February 19-20, 2020 Bethesda, MD

[**eMERGE Day 1: Wednesday, February 19th, 2020**](#vtbm7k2nrnr3)

* [NHGRI program official report | Robb Rowley (NIH/NHGRI)](#fxzl8n9o5445)
* [Meeting Goals *&* Announcements | Rex Chisholm (SC Chair, Northwestern)](#3rnclq6vr6ll)
* [eMERGE data migration | David Crosslin (KPW/UW)](#l1843fujgr6f)
* [Data utilization on the AnVIL |Robert Carroll (VUMC)](#iy5umbevr7cl)
* [Reclassifications | Hana Zouk (Partners/Broad) & Eric Venner (Baylor)](#ptuc5znv97sn)
* [Lessons Learned: Penetrance Panel | Dan Roden (VUMC), Iftikhar Kullo (Mayo), Wendy Chung (Columbia), Gail Jarvik (KP/UW)](#115uqjpmeltm)
* [Evaluation of Breast Cancer Polygenic Risk Scores Using eMERGE Data | Cong Liu (Columbia)](#8cquzy6kfghi)
* [PheMap: a Multi-resource Knowledgebase for High-throughput Phenotyping within Electronic Health Records | Neil Zheng (VUMC)](#cp4ut6e91wen)
* [Use of ClinVar for Variant Selection for Real-Time Genetic Diagnosis of Epilepsy | Nephi Walton (Geisinger)](#gutpod2eawe4)
* [An Implementation Science Framework to Develop a CDS Tool for Familial Hypercholesterolemia | Hana Bangash (Mayo)](#r2plo8ik4fv0)

[**eMERGE Day 2: Wednesday, February 20th, 2020**](#2r1pu75npalm)

* [Opening remarks | Robb Rowley (NIH/NHGRI)](#mh21dmakw9tg)
* [Six-month post-return Outcomes analyses | Josh Peterson (Coordinating Center), Iftikhar Kullo (Mayo), Wendy Chung (Columbia)](#2lcsp2d997jg)
* [Natural language processing phenotyping | Wei-Qi Wei (VUMC) & Chunhua Weng (Columbia)](#3gombr1onmt0)
* [Return of Results: Experiments of nature | Ingrid Holm (BCH) & Julia Wynn (Columbia)](#hn7f2cyf2mtn)
* [XML FHIR update & progress | Mullai Murugan (Baylor)](#fpc8m4lmq3ju)
* [Genomics Overall lessons | David Crosslin (KP/UW), Patrick Sleiman (CHOP), & Megan Roy-Puckelwartz (Northwestern)](#gfim7ug9xe2b)
* [Pharmacogenomics lessons from eMERGE III | Cindy Prows (CCHMC) & Laura Rasmussen-Torvik (Northwestern)](#evu9f14vk8es)
* [eMERGE III final deliverables | Rex Chisholm (SC Chair, Northwestern)](#cqdxwi5nod02)

[**ACTION ITEMS**](#h8yjox8qz48i)

**eMERGE Day 1: WEDNESDAY**

* **NHGRI program official report | Robb Rowley (NIH/NHGRI)** 
  + Dan Roden recently received the Schottenstein Prize & Oscar B. Hunter Career Award. Ken Wiley received the Harold D. West, Ph.D. Distinguished Biomedical Science Award by Meharry Medical College.
  + Josh Denny is the New Chief Executive Officer of the NIH *All of Us Research Program* and Stephanie Devaney became the Program’s Chief Operating Officer and Eric Dishman became the Program’s Chief Innovation Officer.
  + Recently, NHGRI clarified the Institute’s expectation: specifically, that all human data generated by NHGRI-supported research must be derived from specimens or cell lines for which explicit consent for future research use and broad data sharing can be documented. NHGRI set this expectation to increase transparency with researchers and participants and to promote consistent data sharing expectations across genomics studies.
    - Exceptions to this expectation will be considered on a case-by-case basis when there is a compelling scientific justification.
    - Pertains to RFAs with receipt dates after January 25, 2020 and all other NHGRI-supported research applications after January 25, 2021
  + NIH-ACMG Fellowship has expanded to include applicants beyond physicians including genetic counselors, nurse practitioners and physicians’ assistants. Applications are due annually on December 1st.
  + No-Cost Extensions (NCE)
    - Requests for No Cost Extensions (NCE) must be a formal request done via the “Prior Approval” module in the eRA Commons or via email.
    - NCE submission must include information regarding the time required to close out the grant, funds to be used, and project activities.
    - NCE submissions must be received at least 30 days before the end of the current project period
  + There will be a gap between the end of eMERGE III and when the NOGAs arrive for eMERGE IV.
  + The final eMERGE Steering Committee call will take place on June 4th, 2020 from 2:00-5:00 PM EST.
* **Meeting Goals *&* Announcements | Rex Chisholm (SC Chair, Northwestern)** 
  + This is the last in-person Steering Committee meeting for eMERGE III, and now we are finalizing work to be completed. Goals include discussing the [seven network-wide milestones](https://emerge-network.org/wp-content/uploads/2020/03/Network-Milestones.docx), finishing up manuscripts and other projects, establishing timelines for completing remaining work, and holding the lessons learned panel for penetrance.
  + The Clinical Annotation and Outcomes workgroup plans to submit phenotype-specific publications for analysis and lessons learned regarding penetrance and outcomes.
  + Specific outcomes analysis of pediatric FH is available for 20-30 participants. The group is still deciding if there is enough power for a manuscript. The ROR group will discuss barriers to collecting pediatric data in the breakout session and will share key points on day two.
  + Structural variant/CNV variant datasets were created from the imputed dataset and are available to the Network through the CC.
* **eMERGE data migration | David Crosslin (KPW/UW)**
  + David Crosslin presented an update on the eMERGE Network genetic data migration to AnVIL.
  + eMERGE Network datasets include eMERGE Phase I-III Imputed SNVs and Structural Variants (SV), eMERGEseq, PGRNseq, Exome chip, Whole genome sequencing, and Whole exome sequencing.
  + Support meta-files from the GWAS set and eMERGEseq BAMs will be removed from the UW server by March 1st. Analysis files will be on UW server until the end of eMERGE-III funding, March 31, 2020.
  + The CC has transferred local data to the AnVIL and is working with the team to link dbGaP studies to the ‘eMERGE only access’ bucket.
  + Future work may include adding the analysis scripts used to compile the datasets to GitHub to support the scientific community.
  + Downloading the data from the SFTP (UW) site is the easiest way to obtain a copy of the data.
    - The GWAS support meta-files, eMERGEseq BAMs, Exome Chip, and WGS files were removed January 31, 2020
    - Analysis sets (Multisample VCFs; eMERGEseq, GWAS, PGRNseq) will be available until the end of funding
  + The targeted launch of AnVIL is March 2020 however due to possible egress fees, eMERGE members are asked to download data from SFTP site before it is removed. Egress fees for AnVIL are the same as Google Cloud.
  + If data is available to the scientific community there would not be storage fees but if the data is for Network use only storage fees would apply. There is an estimated 3-6 month timespan to clean up data for the public facing transition of AnVIL.
  + AnVIL is hosted by Terra.Bio, and analysis can be completed in the cloud without egress fees, however there are costs of computation to consider if conducting analysis in the cloud. The analysis costs for the AnVIL are the same as for utilizing the [Terra](https://support.terra.bio/hc/en-us/articles/360029772212-Controlling-Cloud-costs-sample-use-cases) platform.
* **Data utilization on the AnVIL | Robert Carroll (VUMC)** 
  + Robert Carroll provided insight on data utilization on AnVIL and gave a demo of a Rheumatoid Arthritis polygenic risk score use case.
  + To complete polygenic risk scores analysis, 95/101 SNPs were extracted from the eMERGE imputed set.
  + Logistic regression was completed to obtain Phecodes derived from ICD-9-CM and ICD-10-CM. Data was adjusted by year of birth, sex, and genetic principal components; data was stratified by site and genetic ancestry.
  + The AnVIL dashboard has a description of the project, information on how to link data with sample information, and cloud storage function associated with the workspace (under files).
  + There are notebooks associated with Python and R. Users are able to examine the cost, time, and computational power needed under RunTime configuration. Users can also custom configure processing ability and environments.
  + Similar to other notebooks, text can be added using Markdown and analysis can be executed in cells. Notebooks can be edited, exported and downloaded as high-resolution images.
  + Files can be copied into local virtual machines to conduct analyses concurrently with others.
  + Notebooks can be shared in order to review and debug analyses and run records are saved for future review.
  + When downloading results, AnVIL will suggest the price for the egress fees. The egress warnings will help prevent investigators from downloading large files and accruing egress fees without their knowledge. The analysis is on a virtual machine, which has a set fee for a specific amount of time.
  + The [STRIDES](https://datascience.nih.gov/strides) (Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability) Initiative is encouraging the usage of cloud resources by providing credits. NHGRI is piloting these methods and interested universities can currently set up a separate agreement with STRIDES.
* **Reclassifications** 
  + **Eric Venner (Baylor)** 
    - ClinVar Variant reanalysis was completed, and the goal is to provide updated reports when classifications change. Reanalysis was accomplished by pulling the latest release from ClinVar and comparing with local data to report upgrades or downgrades using a tool called ReVU ( Reanalysis of Variants and Updater).
    - After filtering for site reporting preferences, there were 112 original VUSs that had new Pathogenic classification in ClinVar.
      * There were three upgrades issued from the initial 26 VUS or benign reports. There were 40 hours of review work completed for upgrades.
      * There were two downgrades released from the original 86 reports. There were six hours of review work completed for downgrades.
    - The lesson learned regarding FHIR implementation is that the mechanics of integration are made possible using open software.
    - Baylor is releasing a new feature for reporting pipeline, Variant Watchlist. Groups of investigators (including members outside of clinical teams) will receive automated reports on variant reclassifications of a defined list of variants. Alerts contain no PHI like a ClinVar entry. The group is requesting [feedback](http://www.hgsc.bcm.edu/watchlists) on this feature with the possible updates, including for gene regions.
  + **Hana Zouk (Partners/Broad)**
    - Variant reclassification affected 0.3% of cases (32/10,500) and represented 2.2% of reported variants (86/806). 56% were low impact (not changing reportable threshold). 44% had a higher impact (affecting reportable thresholds; 8 variants and 14 reports).
    - A comprehensive variant reanalysis was performed with the goal of identifying eMERGE III variants with new evidence that may lead to an updated classification.
    - The strategy was to compare the difference between the CSG and the ClinVar database submissions with at least 1-star status
      * Upgrade VUS or below to LP/P occurring after initial eMERGE analysis
      * Downgrade LP/P variants to VUS or below
    - LOF reanalysis was completed by applying the ClinGen framework to the LOF variants
    - Identification of newly classified LP/P variants by LMM were upgraded after initial reporting in eMERGE.
    - Out of 1995 unique eMERGE variants:
      * 381 LMM unique variants in ClinVar as LP/P assertions with 1-stars and above with 73 discrepancies (LP/P in ClinVar and VUS by CSG) with only 9 potential upgrades
      * 609 LMM unique variants in ClinVar as VUS assertions with 1-stars and above, there were 4 discrepancies (VUS in ClinVar and P/LP by CSG), there were no variants for possible downgrades.
      * Of the 588 LOF variants, 4 variants for potential upgrade.
      * 366 LP/P in LMM database. 7 VUS in ClinVar on initial analysis that have now been upgraded in ClinVar.
      * There were two sites that prompted reassessment from functional assays that led to one upgrade.
      * 29 variants had the potential for an upgrade upon primary review, and 27 VUSs were upgraded to LP/P in 18 genes and 35 participants
        + Reasons for variant upgrades include new guidelines, internal laboratory data, new evidence from the literature, and both internal data and literature evidence.
      * The next step is to issue reports for upgrades as well as reclassification alerts.
    - Sites were asked if they would want updates after eMERGEIII ends. Five sites asked for updated reports from one year to no defined time. Partners Broad will discuss with KPW/UW concerning their request for continuous support. ClinVar created functionality to receive automated variant reports.
    - In the next Clin Gen grant, there will be a request for an Open-source app to site on EHR systems for groups to receive reporting.
* **Lessons Learned: Penetrance Panel** 
  + Dan Roden (VUMC)
    - The focus was on penetrance of Mendelian arrhythmia disease gene variants.
    - Variants of interest are 2631 non-synonymous variants, including 111 P/LP variants in 175 subjects.
    - P/LP variants called by eMERGE CSGs for the four long QT genes
    - Analysis relationship to ICD/CPT codes
      * ICD/CPT codes for 14 arrhythmia phenotypes were curated, similar phenotypes were merged.
      * Gene based analysis: Analyzed penetrance of all P/LP’s or ultra-rare VUS’s.
      * Individual variant-based analysis: examined penetrance of all ultra-rare variants.
      * Logistic regression or burden test to determine odds ratio and significance.
    - Challenges
      * 4/8 genes did not have pathogenicity calls.
      * It is challenging to decipher variants in alternate splice isoforms
      * There are likely differences across sites in patterns of ICD/CPT code usage.
      * Some arrhythmia phenotypes were not well-documented/ascertained by ICD codes.
    - The next steps will be updating the ECG algorithm to find normal ECG dating back prior to eMERGE I.
    - The EHR can be used to identify long QT, not necessarily penetrance of the phenotype of sudden cardiac death.
    - Penetrance is measured by the length of QT interval and if the participant has sudden death. KCNQ1 and KCNH2 are around 25% penetrance. VUMC is using in vitro data to predict penetrance, which is complex, and EHR will not be the only tool.
    - Penetrance should be analysed in context across the lifespan, it is not necessarily a yes or no, but a rage depending on age.
  + Iftikhar Kullo (Mayo)
    - Familial Hypercholesterolemia (FH)
      * There are around 203 P/LP FH genes found in the eMERGE Network, but penetrance focused on 145 individuals, 40 of whom were adolescents.
      * Penetrance estimates are based on the LDL level, using the max LDL in EHR.
      * LDL ascertained at the time of statin use lowered the max LDLR.
      * Analyzed variants and LDL value, the highest LDL level for Indel and Splice variants, in the middle for nonsense and lowest for CNV and missense pathogenic variants.
      * P/LP variants in FH genes are highly penetrant dependent on the max LDL threshold.
      * Expressivity varies with gene and by type of variants in a gene (LDLR).
      * CHD risk is consistent with previously reported ORs. Pack-year cumulative exposure to LDL-C is associated with CHD risk, confirms preliminary results by adding missing data, and receives Marshfield data.
    - Aortopathy
      * Marfan Syndrome (FBN1)
        + All six individuals had P/LP variants in FBN1 had an existing clinical diagnosis of Marfan Syndrome before RoR.
        + Only one individual had clinical genetic testing before eMERGE III RoR, two individuals had an aneurysm + dissection before ROR, one individual had aneurysm only, and one individual had dissection only.
        + Most individuals had a family history of FBN1, and there were two individuals with an unknown history.
      * FTAAD (ACTA2)
        + There were seven eMERGEseq (Geisinger) participants that had P/LP ACTA2 variants, and six participants had the same variant (p.Arg118Gln), four participants began aortic surveillance (1 with ectasia), three participants had a family history of dissection, aneurysm, and ICA. None of the participants had a clinical diagnosis or prior genetic testing.
      * Vascular Ehlers-Danlos (COL3A1)
        + Two eMERGEseq participants had P/LP variants in COL3A1, both of which had undergone genetic counseling.
        + A 68-year-old participant had features suggestive of vascular Ehlers-Danlos (aortic root ectasia, ‘intestinal torsion,’ easily bruisable).
        + Another 66-year-old participant had a daughter with SCAD in her 30’s but did not complete further evaluation.
      * Lessons Learned
        + Some centers consider SLCA10 penetrance as penetrable.
        + The next steps are to solve the missing data vs. unknown problem
        + z scores are needed to display variation in aortic diameters, and index post-ROR treatments and aortic diameters.
  + Wendy Chung (Columbia) -Breast Cancer
    - There were 387 (1.55%) out of 24,956 participants with at least one P/LP for breast cancer gene
    - Penetrance by sex and gene results were expected, with BRCA1(60), BRCA2 (84), and CHEK2, PALB2, ATM (Ataxia-Telangiectasia Mutated) for women and mainly BRCA2 for men.
    - Eighty of the participants (1/3 ) knew of their genetic results prior and purged from the analysis.
    - Curves shifted to lower penetrance after removing the prior diagnosis. Estimates are still higher than the literature.
      * The eMERGE breast cancer phenotype algorithm was used to flag positive variants.
      * The case/control data from the breast cancer algorithm showed more positive diagnoses of breast cancer than observed on the outcomes forms. It is unknown why this discrepancy exists but may be due to the preliminary outcomes form data, or the timing of extraction from the medical record.
    - There may need to be a sensor for older ages as data points are not sufficient.
    - For many, the penetrance for BRCA1 is similar to literature, suggesting the estimates are not accurate due to the number of people. There is little data on CHEK2 penetrance, and PALB2 results are lower than expected. ATM results were smaller than estimated than literature
    - Key: Need **OUTCOME** forms completed to finish breast cancer penetrance analysis (slide 13)
    - Discussion points
      * Penetrance inflated but the rank and order of eMERGE is probably correct
      * Lesson learned
        + In future need to consider excluding individuals that have had previous testing
        + Designed to minimize ascertainment bias
        + Plan to prospectively capture diagnosis to give a true penetrance
  + Gail Jarvik (KP/UW) - Colorectal Cancer
    - Review of penetrance data with plans to present at SC meeting and perform analysis for: Arrhythmias, Colorectal cancer, aortic dilatation/aortopathy, FH and Breast cancer. Plan to perform analysis but not present at SC included is cardiomyopathy. There was not enough data to perform penetrance analysis for Ehlers-Danlos syndrome and chronic kidney disease.
    - Individuals ascertained for CRC were removed from the data for Lynch Syndrome and the MUTYH gene.
      * The study was interested in APC 1307K common in Ashkenazi.
        + Wide variation in APC 1307K results among labs ranging from P/LP and VUS’s.
      * Medical record phenotypes were used to to add disease association observational evidence for ClinGen variants
* **Evaluation of Breast Cancer Polygenic Risk Scores Using eMERGE Data | Cong Liu (Columbia)**
  + The full dataset included 3,943 breast cancer cases, there were 24 PRS scores for each participant using different models and methods. Phenotyping was done by ICD codes, algorithms, and outcomes forms.
  + There were 6 PRS models, including UK Biobank, US Latinas, and Latin American women and Women’s health initiative (WHI) and three from the breast cancer association consortium.
  + For every PRS model, there is a European association, and for Latino, there are only a few models.
  + The best model available was from the UK biobank due to more SNPs being included.
  + All GWAS statistics were from the breast cancer association consortium.
  + Controls are all females without ICD codes for breast cancer. This may contribute to lower outcome rates than if controls are used for future events.
  + The odds ratio compared to the original dataset showed a reduced outcome rate. The outcome rate of PRS5218 was lower in women with a family history. PRS performs better in European subjects.
  + There are different ways to define case and control using the phenotyping algorithm but it is still in development.
  + PRS was significantly associated with the age of the first Breast Cancer diagnosis in Europeans.
  + Lessons learned:
    - Though the trend was consistent, the PRSs did not perform as well as in the original studies.
    - The PRS performed better in subjects without a family history of disease.
    - The PRS performed better in Estrogen Responsive (ER)-positive subjects
    - UKBB derived model (PRS\_5218) transferred better in the eMERGE dataset.
    - There was a decline in predictive power for the Latino population.
    - No heterogeneity found among healthcare provider organizations.
    - There was no evidence that current PRS models can predict Breast Cancer recurrences.
* **PheMap: a Multi-resource Knowledgebase for High-throughput Phenotyping within Electronic Health Records | Neil Zheng (VUMC)**
  + EHRs have greatly facilitated scientific discovery but are designed primarily for clinical care, making the development of algorithms to extract phenotypes from EHRs challenging and time-consuming.
  + PheMap leverages independent online resources by extracting articles describing disease phenotypes from MedicineNet, Mayo Clinic, WikiDoc, Wikipedia, and MedlinePlus. These articles are combined into 942 ‘phenotype’ documents by using NLP to extract medical concepts.
    - Over 90% of phenotypes were covered by at least two or more resources.
    - Term Frequency-Inverse Document Frequency (TF-IDF) was used to quantify concepts.
  + To compare to eMERGE phenotype workgroup algorithms Type 2 Diabetes, Hypothyroidism and Dementia algorithms were selected because they were readily implementable at Vanderbilt.
    - Consistency between eMERGE case-control status and PheMap was over 97% for all three phenotypes.
    - eMERGE algorithm will leave a lot of cases unclassified that PheMap classifiest. There were a number of unclassified patients that will be investigated through chart reviews. Clinical review will help capture additional patients.
  + PheMap can streamline the phenotyping process and facilitate phenotyping portability through the OMOP Common Data Model.
  + PheMap can effectively distinguish cases from controls and can provide a quantitative trait for genetic and phenotypic association studies.
  + The PheMap knowledgebase and phenotyping scripts are freely available at: <https://www.vumc.org/cpm/phemap>.
* **Use of ClinVar for Variant Selection for Real-Time Genetic Diagnosis of Epilepsy | Nephi Walton (Geisinger)**
  + Nephi presented Geisinger's efforts to effectively integrate sequencing data in clinical care. The feasibility study assessed the ability of ClinVar to classify variants in patients with epilepsy of known genetic etiology where each variant has been identified and vetted by at least two clinical specialists and determined to be causative.
  + ClinVar is a freely accessible, public archive of reports of evidence-based relationships among human variation and phenotypes. ClinVar allows submission of variants and supporting information about their pathogenicity from any registered submitter.
    - The variant classifications are given a star rating according to a classification schema that assesses the reliability and evidence provided. The ratings range from zero to four stars with zero-star being no assertion criteria and four-star variants meeting all criteria with practice guidelines available.
  + To conduct the study, a cohort of patients were selected with known or suspected epilepsy with a genetic basis. The ClinVar Classification was assessed for each variant. NLP was performed on all patient notes to detect Human Phenotype Ontology (HPO) triggers that could activate a real-time diagnosis panel.
    - Genes were defined as actionable if epilepsy management was based on gene (medications, diet) or had surveillance guidelines.
    - NLP Triggers, the time point that the NLP system could capture terms that would map to HPO concept triggers to make a diagnosis, were compared to the time a patient received the actual genetic diagnosis.
    - If ClinVar classification was based on the report of the patient being evaluated at Geisinger then the classification was adjusted to the value prior to the patient report.
  + Exome sequencing yields 45% diagnoses in pediatric populations.
  + Real time diagnoses improve management of the disease. There are up to 1500 genes involved, and determining threshold based on suspected pathogenicity is difficult. ClinVar was used to help determine when to return results.
  + A Wall Street Journal article “A Genetic Test Led Seven Women in One Family to Have Major Survey. Then the Odds Changed”, discusses the high impact of false positives and false negatives in Hereditary Breast and Ovarian Cancer compared to Epilepsy tested at Geisinger.
    - An empirical study of genomic malpractice cases in the United States through 2016 found most cases were not based on misinterpretation of genetic variants but were related to failure to perform or act on genetic testing. Study suggested 57% of the genomic medical malpractice cases could have been avoided if genetic testing had been performed when the patient first presented with the condition.
  + **An Implementation Science Framework to Develop a CDS Tool for Familial Hypercholesterolemia | Hana Bangash (Mayo)**
  + The uptake of cascade testing for Familial Hypercholesterolemia is poor as current EHRs are not equipped for genomic medicine. Digital tools could improve awareness, early detection, treatment, and cascade testing.
  + After using an implementation science framework to evaluate the clinical decision support (CDS) for FH, 13 provider interviews were conducted to gauge provider preferences regarding a CDS tool for FH.
  + Framework analysis was applied to interview transcripts to identify key themes, usability testing was conducted to identify recommendations on CDS tool development, and post-interview surveys were used to understand contextual factors that may impact CDS implementation.
    - Some key themes identified from provider interviews were understanding and awareness of FH, usual clinical workflow, physician preferences and value of CDS tools, physician perspectives on patient needs and values, and dissemination and implementation.
    - CDS usability recommendations included format and placement, content, timing and frequency, and prioritization.
  + 84% of providers agreed that the tool would help with early diagnosis of FH. 92% of providers agreed that the tool would help identify and refer patients to have their FH managed. 30% of providers did not think the tool would increase the time needed with the patient.
  + Implementation barriers encountered were multiple committee approvals for CDS and order set, aligning CDS with pre-existing knowledge resources, differing workflows in primary care and specialty clinics, lack of institutional guidelines on CDS implementation, transitioning to new EHR, and EHR limitations.
  + In summary, a majority of healthcare providers were in favor of a CDS tool and order set for FH. Implementation science framework was useful in developing the CDS tool. The number of alerts triggered over a two-month period indicates a high burden of FH and the potential for CDS to increase detection and optimize treatment.
  + Future work for the team are a clinical trial to assess outcomes after deploying CDS, computable representation of genomic data that can trigger CDS, scale development and deployment of genomic CDS in the current generation of EHRs, and apps for genomic medicine that can be layered on top of a standard, interoperable EHR with a crowdsourced effort.

**eMERGE Day 2: THURSDAY**

* **Opening remarks | Robb Rowley (NIH/NHGRI)**
  + There will be no formal March quarterly report for the Network. If sites have any updates or have any other programmatic concerns, they should reach out to NHGRI program staff directly.
* **Six-month post-return Outcomes analyses |Josh Peterson (Coordinating Center), Iftikhar Kullo (Mayo), & Wendy Chung (Columbia)**
  + Josh Peterson (Coordinating Center)
    - eMERGE outcomes assessment has been completed on 15 phenotype specific forms with 4,211 data fields.
    - Out of the 1,518 results received, 1,006 were returned and can be used for outcomes analysis.
    - Several manuscript concept sheets have been submitted for phenotype specific outcomes analysis, and the group is currently prioritizing final efforts to assist with these manuscripts.
      * Outcomes data has been collected via phenotype specific 6-month and 12-month outcomes forms, and a discrepancy resolution google doc. The group is considering an appendix outcomes form to collect missing data.
      * Manuscript concept sheets have been submitted for Arrhythmia, Breast Cancer, Cardiomyopathy, Colorectal Cancer/Polyposis, Adult and Pediatric Familial Hypercholesterolemia, and 22Q Duplication/Deletion.
    - Josh Peterson is leading an overarching outcomes paper, NT296 Impact of Targeted Next Generation Sequencing on Personalized Screening Practices across a Large Cohort, focused on positive and negative results.
      * The project aims are to answer if eMERGEseq return of results for unselected participants impact testing and procedures associated with monogenic risk, if return of negative eMERGEseq reports impact test and procedure frequency compared to unreturned negative tests, and analyze the cost associated to testing based on Medicare fee schedule.
      * Approximately half of the age at event for the ICD and CPT data centrally collected at the CC occurs before the date of return of results, so it is not possible to determine if RoR altered procedures or diagnoses through the EHR. The group discussed refreshing ICD and CPT data from sites to improve the ability to determine changes to clinical procedure post RoR electronically.
      * The project will focus on the top five phenotypes with the most data: Breast Cancer (men and women), Colorectal Cancer/Polyposis, Familial Hypercholesterolemia, Cardiomyopathy, and Arrhythmia.
    - The Outcomes workgroup has decided there will be a Data Freeze 3 with a date aimed for mid-March.
  + Iftikhar Kullo (Mayo)
    - Iftikhar presented a 6-month outcomes analysis for familial hypercholesterolemia variants on 128 adults.
    - The analysis hypothesis is that the return of positive FH genetic testing results will result in increased awareness, treatment, and control of FH. The project ascertained process outcomes consequent to return to results at 6-month post-ROR across the eMERGE Network for new FH diagnosis, tests ordered, medication and procedures.
    - Out of 24,526 reports issued, there were 203 P/LP FH variants and outcomes forms that have been completed for 128 adults (99/128 LDLR variant, 26/128 APOB variant, 3/128 PCSK9 variant) with 6 month outcome forms.
    - The majority of participants have a prior diagnosis for hypercholesterolemia (111/115), however only 22/80 had a prior familial hypercholesterolemia diagnosis prior to ROR which is common.
    - Most participants have a family history of elevated cholesterol and CHD. Genetic testing for FH is not often utilized with less than 10% of participants having prior genetic testing.
    - Procedures received after return of results include, 73 participants with lipid profile, 9 participants with lipoprotein(a), 41 participants with electrocardiogram, 11 participants with stress echocardiogram, 6 participants with echocardiogram, 5 participants with CT scan for coronary calcium, 6 participants with coronary angiography.
    - Key findings so far are a double in expected prevalence of adult FH, low genetic testing and prior clinical diagnoses, 50% statin treatment, and a significant number of new drugs initiated.
    - Ascertainment of outcomes can be challenging. The current outcomes dataset has missing and unknown values. There is a need to distinguish an outcome related to return of result or standard care. An addendum outcomes form may assist with additional analyses related to the initiation of statin post-return of result and changes in LDL-C value.
    - Geisinger is responsible for pediatric FH analyses and has identified adult ages in the pediatric FH forms. It has been suggested to separate cases by age.
      * Mayo used the CC collected age at ROR dataset to confirm the ages of adult FH cases.
  + Wendy Chung (Columbia)
    - Wendy provided an update on breast cancer outcomes analyses completed at Columbia.
    - 183 female participants carried at least one P/LP in one of the breast cancer genes and received results.
      * Almost half of the results (80/183) were previously known, especially BRCA1 and BRCA2, so the clinical impact is lower.
    - Excluding previously known variants there are still genes with a high impact that are remaining.
    - The average age of female participants was around 57 years which is expected given that 80/183 already received the information.
    - Results were returned by the research coordinator and genetic counselor.
    - 10 women had a prior breast cancer diagnosis and 11 women received breast surgeries before return of results. Only one woman was diagnosed with breast cancer after the return of results.
    - At least one screening or prevention took place post-ROR such as biopsy, MRI, ultrasound, and tamoxifen (medication to reduce risk), however there was no breast cancer risk reducing surgeries post-ROR.
    - No increase in mammogram or prophylactic mastectomy surgeries post-ROR. These procedures would be commonly recommended for BRCA1/2 gene however at 6-month post-ROR none of these happened.
    - There are missing and incomplete outcomes forms and the group is working with sites to resolve issues.
    - It would have been beneficial to sequence women with an unknown breast cancer finding to serve as control during outcomes analysis.
    - The team plans to review if participants had prior genetic testing but were not tested for the new variants identified.
* **Natural language processing phenotyping |Wei-Qi Wei (VUMC) & Chunhua Weng (Columbia)**
  + The goals of the NLP group were to develop, validate, implement NLP add-ons for five high-priority phenotypes, evaluate the benefit of additional NLP components, identify and summarize the challenges in a “lessons learned” paper, and recommend best practices.
  + The group selected five phenotypes:
    - * Chronic Rhinosinusitis (CRS)
      * Long QT-Arrhythmia
      * Lupus
      * Asthma/COPD
      * Familial Hypercholesterolemia (FH)
  + The goal was to improve PPV, enrich phenotypes, and improve sensitivity. The group determined what pipeline each NLP would use, including cTAKES and MetaMap.
  + CRS, Lupus, ACO, and FH are under validation and Long QT-Arrhythmia was released for implementation.
  + The group learned that preparation is time-consuming and resource-intensive due to the conversion of code, clearance of intellectual property of real codes, implementation of cTAKES/Metamap and pre/post-processing packages, location and acquisition of desired notes, and data reformatting.
  + Other lessons learned for second validation include customization of programming language, semantic knowledge resource, and internal parameters; simplification by removing sanity check and machine learning steps; and establishing effective communication between groups.
  + The February action plans for primary sites are identifying the notes, major concepts, time periods for acquisition, NLP pipeline, package, and semantic knowledge resource, and notes structure needed.
  + The NLP timeline until June 2020 includes collection and sharing of preparatory guidelines for NLP implementation, the release of NLP phenotypes, implementation of five NLP phenotypes, and submit NLP lessons learned manuscript.
* **Return of Results: Experiments of nature | Ingrid Holm (BCH) & Julia Wynn (Columbia)** 
  + Participant surveys address participant-centered outcomes and can provide lessons learned that can be applied to the next phase of eMERGE for greater participant impact. The goal of the participant survey design is to increase power to assess participant impact by collecting the same participant survey data across the Network.
  + Participant surveys were coordinated across sites by sites conducting surveys, however as the survey was not harmonized prior to dissemination the group is having difficulty pooling the results, reducing sample size depending on how the questions were disseminated at each site.
  + CSER completed participant surveys and analysis across sites and helped create a foundation for eMERGE’s process.
  + The participant surveys were created on REDCap at the CC and the survey data dictionary was sent to sites to be imported into local REDCap for use.
  + Efforts to coordinate the participant surveys from sites have been challenging due to the different populations, site priorities and research focus, and survey timepoints with some sites combining 1-month and 6-12 month surveys. Some sites had already sent out their baseline survey before questions were confirmed. Implementation of participant surveys was distributed and not centrally housed like the penetrance and outcomes forms.
  + Data analysis challenges include that demographic variables differ across sites with some sites administering surveys before variables confirmed, response choices differ by sites, and multiple data sets for each time point and multiple IDs not consistently linked to the eMERGE ID.
  + Reconciling data across sites has taken over 6 months and is still in progress, and some data may not be reconcilable. Data from parents of a child might not be able to be combined with data from other adults participants due to the differences in surveys from pediatric and adult sites.
  + Lessons learned for assessment participant impact:
    - Use of the Healthcare Provider R01 model to help standardize what is collected
    - Don’t send out surveys until they have been finalized by all the sites and use the same survey for all the sites
    - Design surveys to address specific research questions with a statistical plan in place
    - Develop a plan to integrate data from pediatric sites
    - Centralize data collection
    - Have a statistician involved from the beginning
* **XML FHIR update & progress |Mullai Murugan (Baylor)**
  + The FHIR Specification & Pilot Implementation project had three goals:
    - Develop computable and standardized clinical reporting specification for eMERGE with HL7 FHIR
    - Create a POC implementation pilot to generate eMERGE reports with FHIR
    - Identify the feasibility of FHIR enabled ingestion & CDS with EHRs
  + When collaborating with the HL7 Clinical Genomics Workgroup, conversations revolved around harmonization and data interpretation.
  + Baylor completed the pilot implementation plan of generating the FHIR reports by converting raw data to FHIR format reports.
  + The rationale for Pilot Design included demonstrating the validity and utility of FHIR resources, high reuse of existing infrastructure, evaluating new infrastructure, and maturing institutional FHIR and genomics experience.
    - The goal is to complete a PGx report junction (portal dashboard for providers) from the FHIR reports.
  + EHR Model
    - The initial focus was on FHIR spec items for ROR team CDS, the stand-alone app approach aligned with Precision Medicine Analytics Platform (PMAP) at JHU, secure, role-based, access to Microsoft Azure FHIR server, and care coordination roles.
    - Future considerations involve patient engagement, clinical integration, telehealth integration, and population analytics.
  + Project Status: As of now, specification is complete, and documentation is pending. Regarding the pilot, success criteria had been identified, the Alpha version of FHIR report generation, EHR integration, and CDS implementation is in progress.
  + Key Takeaways
    - Create an eMERGE specification using FHIR which gives insight into what HL7 is doing with genomics and in general the community.
    - Demonstrated the use of the FHIR specification into an established and new EHR integration pipeline at JHU.
    - Proposed potential pathways for future adoption.
    - A baseline for ongoing work has been established.
    - HL7 group is an area that needs to grow more, but there is talk of how the FHIR workgroup could be used.
* **Genomics Overall lessons | David Crosslin (KP/UW), Patrick Sleiman (CHOP), & Megan Roy-Puckelwartz (Northwestern)** 
  + Genomics workgroup achievements have balanced collaborative discovery-based research with implementing genomic technology and results into clinical findings.
  + MCS NT357, *Lessons from the eMERGE Network: Balancing genomic discovery and implementation science*, led Jodell Jackson (CC/VUMC) and David Crosslin (CC/UW) describes the goals and lessons from this past phase.
  + Collecting genetic and EMR data centrally will maximize output and relieve repeated site efforts.
  + Lessons learned include genetic fingerprint linking phenotype to genetic data, automated eMERGEID and local ID assignment, master list of genetic and phenotype data, reproducible analysis pipeline and reporting, and more cross-fertilization individually among working groups.
  + The NIH is working to develop stand-alone opportunities investigators. Some IRBs include existing post-approval reporting (PARs) related to existing datasets; this will be used collaboratively by NIH. The goal is to understand and shape current data (rather than simply create or generate new data).
    - The purpose of the PAR application is to provide the Principal Investigator with a mechanism to submit information relevant to the rights/welfare of research participants in a timely fashion to the IRB.
* **Pharmacogenomics lessons from eMERGE III | Cindy Prows (CCHMC) & Laura Rasmussen-Torvik (Northwestern)**
  + The PGx workgroup has had collaborations with the CPIC group through individual and group phone calls
    - The CPIC group has also played a prominent role in the eMERGE III June 2019 in-person meeting
  + Some eMERGE sites are collaborating with IGNITE.
  + SPHINX: The goal was to determine the feasibility of SPHINX, primarily through complex prescription information. Issues included outcomes information requires precise timing of prescriptions compared to outcomes data, frequency is often limited, even in large datasets, no PGx phenotypes selected for site-wide implementation, and centrally collected updated drug information limited to data collected for standard variables. Future opportunities with expanded prescription and drug response variables in SPHINX
  + The PGx workgroup has published [NT155](https://emerge-network.org/wp-content/uploads/2015/04/NT155_Herr-Clinical-Decision-Support-for-Pharmacogenomics.doc) as well as had contributions to various workgroup papers.
  + The PGx workgroup has also participated in the brain trust which is a method to disseminate lessons learned when they involve a specific EHR system.
  + Geisinger PGx CDS: There were 339 total alerts, 126 of which were actionable, there were 45 unique patients and 66 individual events.
  + The project was discussed as a multi-site project, but there was limited bandwidth for network-wide participation.
  + PGx Challenges
    - Translation of rare SNPs to haplotypes can be complicated by issues of cis and trans
    - Some genotyping chips do not contain all data needed for full guideline implementation, particularly in non-white racial and ethnic groups.
    - CYP2D6 is critical for many medications and associated guidelines.
  + PGx: Moving Forward
    - It is essential to allow sufficient time and effort for implementation, even at sites with prior PGx implementation experience.
    - After deployment, sufficient time should be allowed for impact to be demonstrated.
    - PGx phenotypes should be identified early on.
    - Limitations of current PGx guidelines should be considered, particularly concerning non-white populations.
* **eMERGE III final deliverables | Rex Chisholm (SC Chair, Northwestern)**
  + Workgroups provided an update on progress with Network Milestones and timeline for completion before the end of eMERGE III.
  + **Milestone 1**: Conduct penetrance analysis in conditions with sufficient data and assess impact on clinical outcomes.
    - The Clinical Annotation workgroup has created penetrance forms for 10 phenotypes; however, data is insufficient for four phenotypes/disease areas.
    - Penetrance analyses have focused on the six phenotypes with sufficient data including Arrhythmia, Aortopathy, Cardiomyopathy, Familial Hypercholesterolemia, Breast Cancer (men and women), and Colorectal Cancer/Polyposis.
    - For Arrhythmia penetrance, there are three projects focused on procedure codes, electrocardiogram QT intervals, and Brugada syndrome. FH penetrance remaining work includes collecting additional data or correcting incorrect data. Breast Cancer penetrance analysis is in the process of redoing analysis with indications for referrals. Aortopathy penetrance analysis is currently calculating z scores from metrics.
    - Ascertainment bias is an important factor in conducting these analyses. For individuals ascertained, penetrance could not be conducted. Penetrance is also hard to estimate in alleles with low rates.
    - Variant classification can be conflated with penetrance, and it is important to be distinct.
  + **Milestone 2**: Determine the impact of return of genetic results (RoR) on patients’ immediate outcomes, 6 months and available 12 months after RoR for variants with sufficient prevalence and data, which includes identifying modification of clinical care and outcomes related to processes of care, utility, and patients’ psychosocial factors.
    - The Outcomes workgroup has planned a third data freeze in late March for a full dataset to clean and review prior to the closeout call.
    - 12-month outcomes analysis will prioritize high impact and high sampled phenotypes.
    - The Return of Result workgroup and Participant Survey subgroup have coordinated across sites to complete baseline, 1-month, and 6-12 month post-disclosure surveys. Site survey data dictionaries are being reconciled for common questions and the data has been merged at the CC; completion goal date is April 2020.
    - Projects in process are NT349, Family communication following return of positive results (target submission June 2020), NT373, Psychological impact on participants of receiving positive genomic results in eMERGE III (target submission June 2020), and NT363, Participants perceived clinical and personal utility of receiving positive genomic results in eMERGE III (target submission July 2020).
    - The Healthcare Provider Supplement ends April 2021. The group has completed the data collection survey and is currently conducting interviews. Manuscripts will begin September 2020 and submitted by March 2021.
  + **Milestone 3**: Improve and/or standardize genomic clinical decision support (CDS) for return of clinically relevant genetic or incidental results directly to physicians.
    - eMERGE is the first and only Network capable of linking a set of structured clinical genetic result delivery that links heterogeneous organizations in a manner that enables CDS.
    - The EHRI workgroup has developed eMERGE XML result transfer format and open sourced via GitHub, published how the network was created to provide a basis for other efforts, developed draft FHIR genetic result specification and documentation, and maintained a culture of open collaboration where all sites share lessons learned relative to ROR and CDS infrastructure.
    - The group goals for June 2020 are to submit EHRI subgroup manuscripts (NT213 and NT319), submit case series and editorial on genomic medicine implementation manuscript, and pilot two implementations of eMERGE FHIR genetics specifications.
  + **Milestone 4**: Develop a natural language processing (NLP) component for a maximum of five high-priority phenotypes, agreed upon by the phenotyping workgroup, the Steering Committee, and NHGRI.
    - The Phenotyping workgroup goals are to manage 25 eMERGE-III phenotype development and implementation, and develop NLP components for five high-priority phenotypes.
    - As of February 2020, the group has 25 phenotypes developed and validated, 23 phenotypes implemented, 20 released to the public, and two phenotypes (Breast Cancer and Type 2 Diabetes) shared with the *All of Us Research Program*.
    - Remaining work includes implementation of two remaining algorithms by February 29, 2020; collection of preparatory guidelines for NLP implementation by February 29, 2020; release of NLP phenotypes by March 31, 2020, implementation of five NLP phenotypes by May 31, 2020, and submission of NLP-based lessons learned manuscripts by June 30, 2020.
  + **Milestone 5**: Explore the challenges involved in identifying at-risk family members and informing them of their potential risk as well as collect the responses of the family members.
    - The Familial Implications of ROR outcomes form will be used to analyze the expense of having additional follow up care after RoR, the related issue of cascade screening and incorporating into the families.
    - There is the variability of cascade uptake between the Tier 1 conditions.
    - There is better ovarian syndrome cascade testing compared to cases of Lynch Syndrome and FH a paper could be created to help describe differences in prevalence.
    - 6-12 months may still not be enough time for how to go forward for some of the families of these participants; cascade testing may be seen much later on in the records.
  + **Milestone 6**: Estimate the institutional impact of ROR.
    - There are several Return of Result workgroup manuscripts aimed at estimating the institutional impact of ROR and planned to be submitted between April 2020 and June 2020.
    - The overall outcomes project, NT296 will use a medical fee schedule to help address the associated costs of follow up.
    - NT296 hopes to identify how the return of genomic information prompts changes in rates of clinical procedures.
    - Health behavior changes are difficult to implement in real-world populations. Most of these are screening tests for these diseases, for example , colonoscopy began several decades ago and uptake was slow.
  + **Milestone 7**: Disseminate lessons learned on various aspects of genomic medicine implementation by activities such as publishing articles that propose the key elements for effectively returning genomic results to providers and patients and comparing the impact different methods of RoR have on patient and physician care across all sites.
    - There are several manuscripts in progress across workgroups and many focused on lessons learned.
  + **Other workgroups**:
    - The PGx workgroup has tracked the return of PGx results during eMERGE-3, established a relationship with CPIC by sharing lessons learned, explored potential collaborations with IGNITE, examined feasibility of additional eMERGE-3 PGx outcomes or SPHINX PGx projects, and contributed lessons learned in concert with other workgroups.
      * The PGx workgroup plans to publish NT335 on star allele PGx nomenclature and provide guidance for any PGx related work in eMERGE-4.
    - The Genomics workgroup has provided guidance to the eMERGE CC regarding genetic data activities in order to produce four large multiple use discovery-based datasets. Co-chair Patrick Sleiman has led supplement-funded efforts to link Geocoding data to eMERGE participants. Focus groups were established for SPHINX PheWAS data integration to generate and integrate pilot PheWAS data. Assistance was given to VUMC for implementation of alternate methods of assessing penetrance and risk-scores using Network wide data.
      * The Genomics workgroup conducted SV/CNV calling on the eMERGE I data, but supplemental funding would be needed for future analysis of eMERGE III data.
    - Workgroups have successfully described the opportunities and challenges for pediatric eMERGE sites through the collection of outcomes data at pediatric sites, collection of participant survey data and cascade screening. Manuscript NT300 is analyzing the approaches established at pediatric sites (CCHMC and CHOP) and in the Babyseq project (BCH) for return of adult-onset only conditions.
      * Pediatric eMERGE sites plan to develop a manuscript on pediatric results by analyzing cascade testing and outcomes data to address challenges with integrating data from pediatric sites with adult sites.
  + The eMERGE III closeout call has been scheduled for Thursday, June 4th 2:00 pm EST-5:00 pm EST. The goal of the call is for groups to provide concern lessons learned and summarize work conducted over the last several years and outlined as final deliverables.
    - CC will continue in a similar capacity through June and will host the closeout call.

**ACTION ITEMS**

**Coordinating Center**

* The CC will collaborate with the AnVIL team to notify the Network when the data are available for Network use, and disseminate access instructions.
* The CC will work with Baylor to confirm all data on DNANexus have been transferred to the CC and as applicable to the AnVIL before removing data from DNANexus.
* The CC will work with the sites to refresh the ICD and CPT data on eMERGEseq in order to facilitate outcomes analyses and remaining manuscripts.

**Network**

* All workgroups should complete remaining work outlined in the final deliverables [**slides**](https://emerge-network.org/wp-content/uploads/2020/03/Chisholm_Final_Deliverables_20200220-3.pptx) and report final progress to the steering committee during the June 4th, 2020 teleconference.
* The Network should publish ‘in progress’ MCS and contact the CC with manuscript updates, including delays or withdrawals.

**Phenotyping**

* Sites should let the CC and Phenotyping co-chairs know if they will not be able to implement NLP algorithms after March 31st, 2020.
* The Phenotyping group should release all five NLP algorithms for implementation by March 31st, 2020.

**Outcomes**

* Sites should incorporate final 6-months outcomes data into the REDCap forms and the CC will initiate a final data freeze March 20, 2020.