NHLBI Data/Materials Sharing

database of Genotypes and Phenotypes (dbGaP), GWAS, microRNA, gene expression profiling

Trans-Omics for Precision Medicine (TOPMed); WGS, metabolic profiles, protein and RNA expression, DNA methylation

BioData Catalyst

Storage, Toolspace, Access and analytics for big data Empowerment

BioLogic specimen & data repositories
INformation Coordinating Center (BioLINCC); biospecimens & phenotypic data
Background – Data Repository


- Purpose:
  In order to take full advantage of NHLBI supported clinical trials and epidemiologic studies and maximize their research value, data should be made available, under appropriate terms and conditions, to the largest possible number of qualified investigators in a timely manner.
Background BioLINCC

BioLINCC established in 2009 to coordinate activities of Data Repository and Biorepository
Current controlled access portfolio: 50 Observational studies (778,732 participants), 157 trials (460,816 participants)
Data Repository Metrics

- Utilization Metrics
- Outcome Metrics
- Effort Metrics
  - Number of interactions and time to complete data access request
  - Support encounters
  - Monitoring
Utilization Metrics: Completed data access requests
Utilization Metrics: Cumulative data access requests for trial and observational study data (thru, 2019)
Utilization Metrics: Observed / Expected Utilization by study type

Observed Data Requests (% of Total)/ Expected Data Requests (% of Portfolio)

- General Population OS
- Race/Ethnic Specific OS
- Respiratory Distress CT
- Primary/Sec. Prevention CT
- Heart Failure CT
- Disorder Population OS
- Cardiovascular CT
- Other OS
- Other CT
- Other Lung CT
- Emergency Medicine CT
- Sickle Cell CT
- Asthma CT
- Transplant/Transfusion CT
Utilization Metrics: Trends in Utilization by Study Type

Average number of data access requests per study per year

- General Population OS
- Respiratory Distress CT
- Heart Failure CT
- Race or Ethnicity Population OS
- Primary/Sec. Prevention CT
- Cardiovascular CT
- Other OS
- Disorder Population OS
- Emergency Medicine CT
- Other CT
- Sickle Cell CT
- Other Lung CT
- Transplant/Transfusion CT
- Asthma CT

2013-2015 vs 2016-2019
Utilization Metrics: Primary reason for data access request

1. New Question
2. Meta Analysis
3. Statistical Methods
4. Clinical Trial Methods
5. Comparison Group
6. Clinical Prediction, Risk Prediction
7. Other (Pilot, Simulation)
Utilization Metrics: Time from availability of study to first access request

CT: 60.8%  
OS: 74.8%

CT: 83.5%  
OS: 91.6%

CT: Median 270 days.  
OS: Median 205 days  
Log-rank p=0.296
Outcome Metrics: Publications

Number of publications per calendar year

Count of publications

0 20 40 60 80 100 120 140 160
Outcome Metrics: Time to “Incident” publication by type of study (through 2019)

1179 publications
596 Included data from 1 or more trials
69 Trials 1+ publications
506 investigators have published at least once. Mean number of publications per investigator publishing=3.0

Log-rank p=0.858
Outcome Metrics: Citation Percentile (Top N%) publications using repository trial data, observational data and 10% random sample of all NHLBI supported articles (Articles thru May 2015)

Kruskal-Wallis Test p=0.333
Outcome Metrics: Workforce training

- 23% of completed data access requests indicate primary user has 0-5 years of research experience

- QVR search of NIH grant applications found only two training applications (1 K01, 1 K99) mentioned BioLINCC
Messages from the metrics

- Demand for secondary use of data from clinical studies continues to steadily increase
- Growing demand for data from clinical trials
- Metrics can suggest gap areas
- Data from repositories fulfill a range of purposes
- Unclear if publication rates are low
- Citation metrics suggest quality of publications utilizing repository data are similar to publications supported directly by the Institute
- Need to better assess role of repositories in training new investigators