Phenotypes Environmental Scan

From the NIH Collaboratory Phenotypes, Data Standards, and Data Quality Core

Available at: https://www.nihcollaboratory.org/Pages/Knowledge-Repository.aspx

Background: The Phenotypes, Data Standards, and Data Quality Core of the NIH Health Care Systems Research Collaboratory is continually surveying for efforts related to electronic health records (EHR)-based phenotyping to inform work in this area and prevent duplication of effort. This document contains a catalog of phenotype-related efforts identified through the Core’s environmental scan, presented in alphabetical order. Further information on EHR-based phenotyping can be found in the Living Textbook.

Comments: This inventory is ongoing. Clarifications and suggestions for additional phenotyping initiatives or resources should be directed to Michelle.Smerek@dm.duke.edu.

Groups Active in Developing EHR-Based Phenotype Definitions

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Description and Activity in Phenotype Space</th>
</tr>
</thead>
</table>
| Database of Genotypes and Phenotypes (dbGaP) | • Developed by the National Center for Biotechnology Information (NCBI)  
• Available via NCBI’s dbGaP Entrez interface  
• Contains information on various genome-wide association studies (GWAS)  
• Authorized users can access archived GWAS data for use in new exploratory research or cross-study validation |
| Electronic Medical Records and Genomics (eMERGE) Network | • National consortium organized by the National Human Genome Research Institute  
• Combines DNA biorepositories with EHR systems for large-scale, high-throughput genetic research with goal of returning results to patients  
• Phenotype algorithms used to define case and control cohorts for GWAS  
• Published 13 electronic phenotype algorithms; >12 additional algorithms under investigation  
• Includes the Phenotype KnowledgeBase initiative (PheKB.org), a collaborative repository for creating, validating, and sharing electronic phenotype algorithms and their performance characteristics |
| HMO Research Network (HMORN) | • Includes >1,400 scientists and research staff from innovative healthcare systems  
• Seeks to improve health through research using the resources and capabilities of a learning healthcare system  
• Members have expertise in a variety of disciplines, including epidemiology, economics, disparities, outcomes and quality assessment, trials, and genomics |

Prepared by: Rachel Richesson, Anne Heath, and Michelle Smerek
Reviewed by: The Phenotypes, Data Standards, and Data Quality Core of the NIH Collaboratory
Version: 1.0, last updated August 11, 2014
| **Innovation in Medical Evidence Development and Surveillance (IMEDS)** | • Public-private partnership that aims to identify the most reliable methods for analyzing large volumes of data from heterogeneous sources to improve patient care  
• Employs approaches from epidemiology, statistics, and computer science, among others, to find ways for medical researchers to gain knowledge from assessing health databases  
• Goals include allowing the medical research community to do more studies in less time, using fewer resources, while achieving more consistent results |
| --- | --- |
| **Mini-Sentinel** | • Pilot project sponsored by the U.S. Food and Drug Administration  
• Uses EHR data to monitor safety of regulated medical products  
• Data partners in its distributed data network execute standardized computer programs within their institutions, then share aggregate results |
| **National Quality Forum (NQF)** | • A nonprofit, nonpartisan, public service organization committed to transforming the U.S. healthcare system to be safe, equitable, and of the highest value  
• Reviews, endorses, and recommends use of standardized healthcare performance measures  
• Its Quality Positioning System (QPS) helps users find endorsed measures and is searchable by measure title or number, condition, care setting, or measure steward |
| **PhenoTips** | • Open-source software tool for collecting and analyzing phenotypic information for patients with genetic disorders  
• Consists of Web-based user interface coupled with back-end database where phenotypic information is represented using standardized Human Phenotype Ontology  
• User interface designed to mirror clinical workflow to facilitate recording of observations during patient encounters |
| **Phenotype Discoverer (PhenoDisco)** | • Web-based information retrieval system for dbGaP  
• Designed to improve phenotype searching of dbGaP using text processing tools that standardize phenotype variables and study metadata |
| **Phenotype Portal** | • Tool funded by the SHARPn Project  
• Uses informatics-based phenotyping to enable clinicians and investigators to identify patient cohorts through EHR data  
• Identifiable cohorts facilitate clinical trial enrollment and outcomes research and inform clinical decision support  
• First platform for generating and executing Meaningful Use standards-based phenotype algorithms for use across multiple institutions and investigators |
| **SHARPn Project** | • Funded by the Office of the National Coordinator  
• Collaboration among 16 academic and industry partners committed to developing/distributing software to support secondary use of EHR data  
• Areas of focus include high-throughput phenotyping, data normalization, natural language processing, and data quality metrics  
• SHARPn uses the term phenotype to mean the algorithmic recognition of any cohort within an EHR for a defined purpose |

**Prepared by:** Rachel Richesson, Anne Heath, and Michelle Smerek  
**Reviewed by:** The Phenotypes, Data Standards, and Data Quality Core of the NIH Collaboratory  
**Version:** 1.0, last updated August 11, 2014
• Their view of phenotyping includes criteria for patient inclusion in clinical trials, outcomes research, and observational studies; numerators and denominators for clinical quality metrics; and triggers for clinical decision support rules

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Description and Activity in Phenotype Space</th>
</tr>
</thead>
</table>
| National Library of Medicine Value Set Authority Center (VSAC) | • Provides downloadable access to all official versions of vocabulary value sets contained in the 2014 Clinical Quality Measures  
• Each value set consists of the numerical values (codes) and human-readable names (terms), drawn from standard vocabularies such as SNOMED CT, RxNorm, LOINC, and ICD-10-CM, which are used to define clinical concepts used in clinical quality measures (e.g., patients with diabetes, clinical visit) |
| Office of the National Coordinator (ONC) | • Principal federal entity charged with coordination of nationwide efforts to implement and use the most advanced health information technology and the electronic exchange of health information  
• Funded the SHARPn project, which developed the Phenotype Portal (www.phenotypeportal.org) |

Prepared by: Rachel Richesson, Anne Heath, and Michelle Smerek
Reviewed by: The Phenotypes, Data Standards, and Data Quality Core of the NIH Collaboratory
Version: 1.0, last updated August 11, 2014