Practical Development and Implementation of EHR Phenotypes

NIH Collaboratory Grand Rounds
Friday, November 15, 2013
The Southeastern Diabetes Initiative (SEDI)
Setting the Context

- Population health interventions to accomplish the triple aim in people with diabetes
- Understanding the health of a community and burden of disease

Phenotyping methods and implementation to support project objectives

Foundation of electronic health record (EHR) data from healthcare delivery in 4 counties
Risk Prediction and Intervention: The Need for Clinical Risk Factors and Outcomes
Why Computable Phenotypes?

• Correct “disease” identification
• Several downstream implications
  – Estimation of incidence
  – Study design: include and exclude
  – Identification of “risk factors”
  – Effect estimation
  – Who (how) to treat and who to spare
  – Bias due to incorrect disease specification

Attribution: Paramita Saha Chaudhuri, PhD
Phenotyping and discovery.

True patient state

Recording process

Raw EHR data

High-throughput phenotyping

Phenotype

Discovery

Knowledge
- Classify
- Predict
- Understand
- Intervene

Health care process model

Hripcsak G, and Albers D J J Am Med Inform Assoc 2013;20:117-121
SEDI Medical-Social Risk Algorithm Drives Intervention

• Different intensities of intervention
  • High-intensity clinical teams vs. lower-intensity community-based teams

• Different modes of intervention
  • Patient basis, neighborhood basis, community basis

• Targeted intervention
  • Stratifying patients based on risk, both at patient and neighborhood levels
The Diabetes Phenotype Comparison
Problem Statement

• EHR-driven computable phenotypes exist and are an important source of knowledge
  – Electronic Medical Records and Genomics (eMERGE) Network and Phenotype Knowledge Base
  – Entities such as the Center for Medicare & Medicaid Services
  – Many others

• How should we recognize, document, implement, and validate authoritative source phenotypes?

• How should we evaluate the best fit and utility of phenotypes, especially applied to population health management?
Problem Statement

• Which patients in a 5-year EHR dataset have diabetes?
A comparison of phenotype definitions for diabetes mellitus

Rachel L Richesson,1 Shelley A Rusincovitch,2 Douglas Wixted,3 Bryan C Batch,4 Mark N Feinglos,4 Marie Lynn Miranda,5 W Ed Hammond,6,2 Robert M Califf,3,7 Susan E Spratt1

ABSTRACT
This study compares the yield and characteristics of diabetes cohorts identified using heterogeneous phenotype definitions. Materials and methods Inclusion criteria from seven diabetes phenotype definitions were translated into query algorithms and applied to a population (n=173,503) of adult patients from Duke University Health System. The numbers of patients meeting criteria for each definition and component (diagnosis, diabetes-associated medications, and laboratory results) were reported. Results Three phenotype definitions based heavily on ICD-9-CM codes identified 9%-11% of the patient population. The electronic medical records and genomics, NCIC Registration, and diabetes-associated medications definitions, which have restricted or no ICD-9-CM criteria, identified the smallest proportions of patients (7%). The populations. Furthermore, standard phenotype definitions can streamline the development of patient registries from healthcare data, and enable consistent inclusion criteria to support regional surveillance and the identification of rare disease complications. An understanding of the populations generated from various phenotype definitions will inform standard methods for identifying diabetes cohorts, facilitate the rapid generation of patient registries and research databases with uniform sampling criteria, and enable comparative and aggregate analysis. This descriptive study presents and compares the size and characteristics of patient populations retrieved using different phenotype definitions adopted from registries and research networks, and can help inform intervention programs in our current federal reporting standards.

Table 1 Data domain criteria used in selected phenotype definitions

<table>
<thead>
<tr>
<th>Phenotype definitions</th>
<th>ICD-9-CM (250.2) and 250.22 (excludes type 1 specific codes)</th>
<th>Expanded ICD-9-CM Codes (250.2, 250.22, 250.3, 250.4, 250.6)</th>
<th>Fasting glucose</th>
<th>Random glucose</th>
<th>Abnormal QT</th>
<th>Diabetes-associated medications*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richesson</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Rusincovitch</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Wixted</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Batch</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Feinglos</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Miranda</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Hammond</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
<tr>
<td>Califf</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
<td>△</td>
</tr>
</tbody>
</table>

*Medications vary by phenotype definition and are listed here in the supplementary appendix (available online). The RASCAL phenotype definition is the only one that includes the case scenarios with varying combinations of criteria. Any instance of type 1 specific codes (x:250.2, 250.22): results in the exclusion of the patient.

## Individual Cohort Yields

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>DUHS reference population</th>
<th>ICD-9-CM 250.xx Codes</th>
<th>CMS CCW (full ICD-9 set)</th>
<th>NYC A1c Registry</th>
<th>Diabetes-associated medications</th>
<th>DDC phenotype</th>
<th>SUPREME-DM</th>
<th>eMERGE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose for phenotype</strong></td>
<td></td>
<td>Billing</td>
<td>Health services research</td>
<td>Care management</td>
<td>*T2DM preferred</td>
<td>Community-wide intervention</td>
<td>*T2DM preferred</td>
<td>All</td>
</tr>
<tr>
<td><strong>Type of diabetes targeted</strong></td>
<td></td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>*T2DM preferred</td>
<td>All</td>
<td>T2DM</td>
</tr>
<tr>
<td>Age—year† (mean±SD)†</td>
<td>41.7±17.5</td>
<td>56.1±15.8</td>
<td>56.8±15.5</td>
<td>56.2±15.1</td>
<td>54.1±15.3</td>
<td>55.6±16.3</td>
<td>56.6±15.9</td>
<td>57.3±15.4</td>
</tr>
<tr>
<td>Female sex: # and (%)</td>
<td>99,695 (57%)</td>
<td>10,644 (56%)</td>
<td>9,185 (56%)</td>
<td>6,812 (56%)</td>
<td>6,933 (56%)</td>
<td>12,603 (57%)</td>
<td>10,681 (56%)</td>
<td>6524 (56%)</td>
</tr>
<tr>
<td>No of encounters§ (mean±SD)†</td>
<td>20±33.5</td>
<td>46±57.1</td>
<td>49±58.6</td>
<td>54±59.3</td>
<td>54±60.4</td>
<td>46±56.3</td>
<td>48±57.9</td>
<td>45±52.5</td>
</tr>
<tr>
<td>Length of time (in days) between first and last patient encounter (mean±SD)†</td>
<td>861±67.5</td>
<td>1252±587.6</td>
<td>1295±558.4</td>
<td>1365±524.5</td>
<td>1394±500.5</td>
<td>1224±395.9</td>
<td>1237±576.1</td>
<td>1258±579.4</td>
</tr>
<tr>
<td>Total patients identified</td>
<td>173,503</td>
<td>18,893</td>
<td>12,182</td>
<td>11,800</td>
<td>22,050</td>
<td>18,958</td>
<td>11,620</td>
<td></td>
</tr>
<tr>
<td>% Reference population identified</td>
<td>n/a</td>
<td>11%</td>
<td>9%</td>
<td>7%</td>
<td>7%</td>
<td>13%</td>
<td>11%</td>
<td>7%</td>
</tr>
</tbody>
</table>

*Project focus or intent is for T2DM populations, but phenotype does not aggressively eliminate T1DM patients.
†Patients with indications of T1DM are specifically excluded.
‡Age at the beginning of the observation period, 1 January 2007.
§Within observation period, 1 January 2007–31 December 2011.
Table 1 Data domain criteria used in selected phenotype definitions

<table>
<thead>
<tr>
<th>Phenotype definitions:</th>
<th>ICD-9-CM 250.xx</th>
<th>ICD-9-CM 250.x0 and 250.x2 (excludes type 1 specific codes)</th>
<th>Expanded ICD-9-CM Codes (249.xx, 357.2, 362.0x, 366.41)</th>
<th>HbA1c</th>
<th>Fasting glucose</th>
<th>Random glucose</th>
<th>Abnormal OGTT</th>
<th>Diabetes-associated medications*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-9-CM 250.xx</td>
<td>✔️</td>
<td>✲*|</td>
<td>|</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMS CCW</td>
<td>▲*|</td>
<td>|</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYC A1c Registry</td>
<td>✲*|</td>
<td>|</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes-associated medications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDC</td>
<td>▲*|</td>
<td>|</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUPREME-DM</td>
<td>▲*|</td>
<td>|</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eMERGE†</td>
<td>✔️</td>
<td>✲*|</td>
<td>|</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Medications vary by phenotype definition and are listed for each in the supplementary appendix (available online only).

†The eMERGE phenotype definition consists of five case scenarios with varying combinations of criteria. Any instance of type 1 specific codes (i.e., 250.x1, 250.x3) results in the exclusion of the patient.

- ✔️ = Sole criteria.
- ▲ = Optional criteria, one of many.
- * = Distinction made between inpatient and outpatient context.
- \|\| = Distinction made for multiple instances and/or time points.

CMS CCW, Centers for Medicare and Medicaid Services Chronic Condition Data Warehouse; DDC, Durham Diabetes Coalition; eMERGE, electronic medical records and genomics; HbA1c, hemoglobin A1c; ICD-9-CM, International Classification of Disease, revision 9, clinical modification; NYC, New York City; OGTT, oral glucose tolerance test; SUPREME-DM, Surveillance, Prevention, and Management of Diabetes Mellitus.
Simple Phenotype Criteria Example: ICD-9-CM Diagnosis Category 250.xx

Source:
ICD-9-CM diagnosis code 250 with any degree of specificity in the fourth and fifth decimal precision (250.xx).

Definition:
Adult Durham Population patients who meet ONE OR MORE of the following criteria during a DukeMed encounter between 2007-2011:
- One or more instances of the ICD-9-CM diagnosis code 250.xx (see table 1) for any type of encounter (inpatient, outpatient, ED)
Complex Phenotype Criteria Example: SUPREME-DM Phenotype

**Definition:**
Adult Durham Population patients who meet **ONE OR MORE** of the following criteria during a DukeMed encounter between 2007-2011:

- One or more instances of the specified ICD-9-CM diagnosis codes (see table 7) on an **inpatient** encounter
- OR 2 or more instances of the specified ICD-9-CM diagnosis codes (see table 7) on **outpatient** encounters on separate days
- OR 1 or more instances of active stand-alone medication (see table 8) reported during outpatient medication reconciliation
- OR 1 or more Oral Glucose Tolerance Test (OGTT) 2-hour 75g result $\geq 200$ mg/dl where there is NO DIAGNOSIS CODE on the same encounter indicating pregnancy (V22, V23)
- OR 2 or more hemoglobin A1c results $\geq 6.5\%$ on 2 different days within 730 day span
- OR 2 or more fasting glucose results $\geq 126$ mg/dl on 2 different days within 730 day span
- OR 2 or more random glucose results $\geq 200$ mg on 2 different days within 730 day span
- OR within a 730 day span on 2 different days:
  - Fasting glucose results $\geq 126$ mg/dl
  - AND Random glucose results $\geq 200$ mg
- OR within a 730 day span (can be same day):
  - Hemoglobin A1c results $\geq 6.5\%$
  - AND Fasting glucose results $\geq 126$ mg/dl
- OR within a 730 day span (can be same day):
  - Hemoglobin A1c results $\geq 6.5\%$
  - AND Random glucose results $\geq 200$ mg
Very Complex Phenotype Criteria Example:

eMERGE (NW) Phenotype

Definition:
Adult Durham Population patients who meet ONE OR MORE of the following criteria during a DukeMed encounter between 2007-2011:

- **Case 1:**
  - Zero instances of T1 diagnosis codes (see Table 9) on any type of encounter (inpatient, outpatient, ED)
  - AND 1 or more instances of T2 diagnosis codes (see Table 10) on any type of encounter (inpatient, outpatient, ED)
  - AND has BOTH T1 meds and T2 meds (see Table 11 and Table 12)
  - AND i2 med date is PRIOR TO i1 med date

- **OR Case 2:**
  - Zero instances of T1 diagnosis codes (see Table 9) on any type of encounter (inpatient, outpatient, ED)
  - AND 1 or more instances of T2 diagnosis codes (see Table 10) on any type of encounter (inpatient, outpatient, ED)
  - AND has NO instances of T1 meds (see Table 11)
  - AND has one or more T2 meds (see Table 12)

- **OR Case 3:**
  - Zero instances of T1 diagnosis codes (see Table 9) on any type of encounter (inpatient, outpatient, ED)
  - AND 1 or more instances of T2 diagnosis codes (see Table 10) on any type of encounter (inpatient, outpatient, ED)
  - AND has NO instances of T1 meds (see Table 11)
  - AND has NO or more T2 meds (see Table 12)
  - AND has at least one abnormal lab:
    - Hemoglobin A1c result >= 6.5%
    - Fasting glucose result >= 125 mg/dl
    - Random glucose result >= 200 mg/dl

- **OR Case 4:**
  - Zero instances of T1 diagnosis codes (see Table 9) on any type of encounter (inpatient, outpatient, ED)
  - AND 0 instances of T2 diagnosis codes (see Table 10) on any type of encounter (inpatient, outpatient, ED)
  - AND has 1 or more instances of T2 meds (see Table 12)
  - AND has at least one abnormal lab:
    - Hemoglobin A1c result >= 6.5%
    - OR Fasting glucose result >= 125 mg/dl
    - OR Random glucose result >= 200 mg/dl

- **OR Case 5:**
  - Zero instances of T1 diagnosis codes (see Table 9) on any type of encounter (inpatient, outpatient, ED)
  - AND 2 or more instances of T2 diagnosis codes meds on at least TWO SEPARATE dates on any type of encounter (inpatient, outpatient, ED)
  - AND has 1 or more instances of T2 meds (see Table 11)
  - AND has 0 instances of T2 meds (see Table 12)
Phenotypes Development: A Pragmatic Approach
Phenotype development and validation.

Newton K M et al. J Am Med Inform Assoc 2013;20:e147-e154
Process Diagram

Identify candidate phenotype (condition or event)
Examples: diabetes, stroke, myocardial infarction

Analyze existing EHR definitions to choose 5-7 authoritative computable phenotypes
Authoritative sources may include professional societies, Joint Commission, CMS, AHRQ, etc

Implement authoritative computable phenotypes with Duke EHR data
Communicate to keep focus in synch

Develop statistical analysis plan (SAP) and data collection form (DCF)
SAP includes sampling strategy (eg, superset)

Determine “gold standard” clinical definition/source
The definition needs enough specificity for unambiguous chart review and adjudication (eg, professional societies)

Develop protocol for chart review
Includes chart review protocol, IRB approvals, reviewer recruitment

REDCap Database Go-Live

Perform chart review to create “gold standard” cohort
Each chart is reviewed twice, with adjudication performed for inter-reviewer discrepancies

REDCap Database Lock

Analyze results
Includes sensitivity/specificity of individual authoritative phenotypes against “gold standard” cohort, reviewer concordance analysis, etc

Evaluate fit/utility for SEDI project

As appropriate: Modify authoritative phenotype to better meet SEDI objectives

Rinse & Repeat

Attribution: Center for Predictive Medicine

Blue = CPM-led activity
Green = Clinician-led activity
Blend = Joint effort
## Recognizing Authoritative Sources

### Table 1: Primary Phenotype Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Classifications Software (CCS), also known as AHRQ Bundles</td>
<td>Only based upon diagnosis codes, but very large listing of conditions; this is the basis for most early SEDI variables. <a href="http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp">http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp</a></td>
</tr>
<tr>
<td>CMS Chronic Conditions Warehouse (CCW)</td>
<td>Only based upon diagnosis codes and procedure codes; clinical review to date has felt that inclusion logic can be overly broad. <a href="https://www.ccwdata.org/web/guest/condition-categories">https://www.ccwdata.org/web/guest/condition-categories</a></td>
</tr>
<tr>
<td>Mini-Sentinel</td>
<td>Exhaustively researched definitions, but limited number of phenotypes represented. <a href="http://www.minisentinel.org/assessments/diagnostic_and_medical_procedures/default.aspx">http://www.minisentinel.org/assessments/diagnostic_and_medical_procedures/default.aspx</a></td>
</tr>
<tr>
<td>gMERGE Network and PhenKB phenotypes library</td>
<td>Probably the most well-recognized phenotyping source at present, but limited number of phenotypes represented; should be carefully evaluated because core mission of genomic studies can result in exclusionary logic inappropriate for the SEDI population health focus. <a href="http://www.phenkb.org/phenotypes">http://www.phenkb.org/phenotypes</a></td>
</tr>
<tr>
<td>QualityNet (joint effort of CMS and Joint Commission)</td>
<td>Separates measures between inpatient basis and outpatient basis. Go to the &quot;specifications manual&quot; option; the appendices contain specific listings of ICD-9 code tables, medication tables, and CPT codes. This is one of the only CPT code groupings that we’ve seen so far (CPT license is very restrictive), but QualityNet only includes for outpatient context. <a href="https://www.qualitynet.org/">https://www.qualitynet.org/</a></td>
</tr>
<tr>
<td>Professional society guidelines</td>
<td>These are an important source for definitions of abnormal laboratory results and specific ranges, which are often not represented in other definitions. Examples: American Diabetes Association, National Kidney Foundation, American College of Cardiology</td>
</tr>
<tr>
<td>Major and well-recognized clinical trials and registries using EHR data to identify cohorts</td>
<td>Clinical and expert guidance can be important for identification of these pivotal trials, another potential technique might be to limit results to high-impact journals via a PubMed search.</td>
</tr>
</tbody>
</table>

### Table 2: Secondary Phenotype Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Commission</td>
<td>The CMS/Joint Summit QualityNet is generally the better source, not using the Joint Commission directly. This organization evaluates hospital adherence with federal regulations, and publishes a specifications manual for inpatient quality measures. Appendix A.1 lists the definitions for specific conditions, mostly based upon ICD-9. A limitation is that these definitions are centered on inpatient admissions, and may not be applicable in an outpatient setting. <a href="http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx">http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx</a></td>
</tr>
<tr>
<td>World Health Organization (WHO) Global Burden of Disease</td>
<td>In general, this may be useful for mental health, but probably not helpful for most clinical condition phenotypes. The Global Burden of Disease classifications include both ICD-9 and ICD-10 diagnosis code groupings. See “cause-specific documentation” for individual conditions (e.g., cerebrovascular disease, diabetes mellitus, etc). The diagnostic codes are not granular (e.g., it just lists 230 for diabetes mellitus), due to global application, and the clinical conditions are very broad. May be somewhat out of date. It appears that the classifications date back to 2000, the last formal GBD update appears to have been 2004, although this is difficult to ascertain from their website. However, there are a lot of mental health classifications, which may be useful. <a href="http://www.who.int/healthinfo/global_burden_disease/data_sources_methodology/index.html">http://www.who.int/healthinfo/global_burden_disease/data_sources_methodology/index.html</a></td>
</tr>
</tbody>
</table>

**Attribution:** Center for Predictive Medicine
# Evaluating Existing Definitions

## Phenotype Overview: Acute Myocardial Infarction (research by Maria V. Grau-Sepulveda)

**Clinical Definition Source:** Joint ESC/ACCF/AHA/WHF Task Force for the Universal Definition of MI

### Table 1: Authoritative Phenotype Comparison

<table>
<thead>
<tr>
<th>Source</th>
<th>Evaluation of Prevalence vs Incidence</th>
<th>EHR Data Subject Areas</th>
<th>Phenotype Comments</th>
<th>Phenotype Encounter Basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRQ Bundles (Clinical Classifications Software)³</td>
<td>Prevalence</td>
<td>ICD-9 Diagnoses</td>
<td>Broad definition&lt;br&gt;• AMI diagnosis codes:&lt;br&gt;  o initial episode&lt;br&gt;  o subsequent episode&lt;br&gt;  o unspecified episode</td>
<td>Any encounter</td>
</tr>
<tr>
<td>CMS Chronic Conditions Warehouse²</td>
<td>Incidence</td>
<td>ICD-9 Diagnoses Encounter Basis</td>
<td>Only AMI initial episode codes</td>
<td>Inpatient basis, first/second diagnosis code</td>
</tr>
<tr>
<td>Mini-Sentinel #1 (AMI/Anti-Diabetic Agents)³</td>
<td>Incidence</td>
<td>ICD-9 Diagnoses Encounter Basis Death Data</td>
<td>AMI initial/unspecified episode codes</td>
<td>Inpatient basis, first diagnosis code&lt;br&gt; Also includes death w/ one day of ED visit with ischemic disease codes</td>
</tr>
<tr>
<td>Mini-Sentinel #2 (Validation of AMI Cases)³</td>
<td>Incidence</td>
<td>ICD-9 Diagnoses Encounter Basis</td>
<td>AMI initial/unspecified episode codes Do not include death criteria</td>
<td>Inpatient basis, first diagnosis code</td>
</tr>
<tr>
<td>CMS/Joint Summits QualityNet (Yale models for AMI and HF)⁴ Joint Commission identification of AMI⁷</td>
<td>Incidence</td>
<td>ICD-9 Diagnoses Encounter Basis</td>
<td>AMI initial/unspecified episode codes</td>
<td>Inpatient basis, first diagnosis code</td>
</tr>
</tbody>
</table>

---

³ [http://www.hcup-us.ahrq.gov/toolssoftware/ccs/AppendixASingleIX.txt](http://www.hcup-us.ahrq.gov/toolssoftware/ccs/AppendixASingleIX.txt)
⁶ [http://mini-sentinel.org/work_products/Validation_HealthOutcomes/Mini-Sentinel-Validation-of-AMI-Cases.pdf](http://mini-sentinel.org/work_products/Validation_HealthOutcomes/Mini-Sentinel-Validation-of-AMI-Cases.pdf)

**Attribution:** Maria Grau-Sepulveda, MD, MPH
SEDI Core Data Domains

1. Patient Demographics
2. Encounters
3. Diagnoses
4. Procedures
5. Lab Results
6. Vital Signs
7. Medications
8. Social History

# Healthcare Workflows and EHR Data

## Data Reflective of Biomedical Phenomena:
- Laboratory result values
- Vital sign measures
- Direct physiological measures (such as EKG, pulmonary function tests, etc)
- Pathology specimens
- Images

## Data Reflective of Diagnostic Processes:
- Diagnosis codes (includes professional billing, technical billing, medical coding)
- Problem lists
- Clinical narrative related to diagnosis (including pathology and imaging reports)

## Data Reflective of Behavior, Functioning, or Experience of Symptoms:
- Patient-reported outcomes
- Social and family history
- Other instruments addressed to patient

## Data Reflective of Treatment Decisions:
- Provider orders (including medications)
- Procedure codes
- Procedure reports (such as surgery reports)
- Clinical narrative relating to treatment plans

---

Patient-centered context, but mediated by provider decisions of diagnostic testing and exposure to health system

Healthcare-centric context, but mediated by billing processes, medical coding conventions, and healthcare EHR system platform

---

A work in progress with Greg Simon, Michelle Smerek, and Rachel Richesson
Contact Information

Shelley A. Rusincovitch, Senior Informatics Analyst
Health Intelligence and Research Services
Duke Health Technology Solutions (DHTS)

Phone: 919-668-5954 / shelley.rusincovitch@duke.edu
Supplemental Slides
Selected SEDI References


Center for Predictive Medicine

Paramita Saha Chaudhuri, PhD
Maria Grau-Sepulveda, MD, MPH
Eric Laber, PhD
Nick Meyer
Ben Neely, MS
Charlotte Nelson, MS
Michael Pencina, PhD
Shelley Rusincovitch