Can we trust the Phenotypes?

Shawn Murphy MD, Ph.D.
Use Phenotyping Algorithms to define cohorts of treatment-resistant and treatment-responsive depression

Initially:
AUC = 0.54
Finally:
AUC = 0.87

<table>
<thead>
<tr>
<th>Clinical Status</th>
<th>Model</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Precision</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressed</td>
<td>Billing Codes</td>
<td>0.95</td>
<td>0.09 (0.03)</td>
<td>0.57 (0.14)</td>
<td>0.54 (0.02)</td>
</tr>
<tr>
<td>Depressed</td>
<td>NLP</td>
<td>0.95</td>
<td>0.42 (0.05)</td>
<td>0.78 (0.02)</td>
<td>0.88 (0.02)</td>
</tr>
<tr>
<td>Depressed</td>
<td>NLP + Billing Codes</td>
<td>0.95</td>
<td>0.39 (0.06)</td>
<td>0.78 (0.02)</td>
<td>0.87 (0.02)</td>
</tr>
<tr>
<td>Well</td>
<td>Billing Codes</td>
<td>0.95</td>
<td>0.06 (0.02)</td>
<td>0.26 (0.27)</td>
<td>0.55 (0.03)</td>
</tr>
<tr>
<td>Well</td>
<td>NLP</td>
<td>0.95</td>
<td>0.37 (0.06)</td>
<td>0.86 (0.02)</td>
<td>0.85 (0.02)</td>
</tr>
<tr>
<td>Well</td>
<td>NLP + Billing Codes</td>
<td>0.95</td>
<td>0.39 (0.07)</td>
<td>0.85 (0.02)</td>
<td>0.86 (0.02)</td>
</tr>
</tbody>
</table>
White matter abnormalities associated with treatment-resistant depression

- Scans collected as part of routine clinical care
- Diffusion tensor imaging in 150 pts
- Age-related decline in white matter integrity increases with treatment resistant depression

Medial fornix shows strongest effect

Rapid investigation of QTc prolongation

- **FDA warning 2011 for Celexa**
  Safety Announcement: [8-24-2011] ”should no longer be used at doses greater than 40 mg per day because it can cause abnormal changes in the electrical activity of the heart.”

- **But, did NOT include Lexapro (which is active ingredient of Celexa [s-enantiomer])**

- **Shown to be true with RPDR-derived data set with >38,000 EKGs obtained within 14 – 90 day window after medication initiated**

<table>
<thead>
<tr>
<th>Anti-depressant</th>
<th>Adjusted model†</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SSRI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citalopram (Celexa)</td>
<td>2.85</td>
<td>0.004</td>
</tr>
<tr>
<td>Escitalopram (Lexapro)</td>
<td>3.80</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Fluoxetine (Prozac)</td>
<td>1.44</td>
<td>0.150</td>
</tr>
<tr>
<td>Paroxetine (Paxil)</td>
<td>0.07</td>
<td>0.943</td>
</tr>
<tr>
<td>Sertraline (Zoloft)</td>
<td>0.87</td>
<td>0.383</td>
</tr>
<tr>
<td><strong>Other anti-depressants</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amitriptyline</td>
<td>4.10</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Bupropion</td>
<td>-2.15</td>
<td>0.032</td>
</tr>
<tr>
<td>Duloxetine</td>
<td>0.60</td>
<td>0.547</td>
</tr>
<tr>
<td>Mirtazapine</td>
<td>-1.46</td>
<td>0.145</td>
</tr>
<tr>
<td>Nortriptyline</td>
<td>1.23</td>
<td>0.219</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>1.15</td>
<td>0.251</td>
</tr>
<tr>
<td><strong>previously known prolonger</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methadone</td>
<td>5.32</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

† Adjusted for age, gender, race, type of insurance, history of major depression, history of myocardial infarction and Charlson comorbidity score

Roy Perlis MD, MSc and team
QT interval and antidepressant use: a cross sectional study of electronic health records

BMJ 2013;346:f288 doi: 10.1136/bmj.f288

Victor M Castro team lead, Caitlin C Clements clinical research coordinator, Shawn N Murphy associate professor of neurology, Vivian S Gainer team lead, Maurizio Fava Slater Family professor of psychiatry, Jeffrey B Weilburg assistant professor of psychiatry, Jane L Erb assistant professor of psychiatry, Susanne E Churchill executive director, i2b2 National Center for Biomedical Computing, Isaac S Kohane director, i2b2 National Center for Biomedical Computing, Dan V Iosifescu associate professor of psychiatry, Jordan W Smoller associate professor of psychiatry, Roy H Perlis associate professor of psychiatry.

Partners EMR

~4 M

Antidepressant / methadone datamart

241,308

Adult patients prescribed an anti-depressant or methadone and no prior history of Torsade de pointes

202,911

patients had no EKG in the exposure window EXCLUDED (Table S1)

38,397

Main Study Cohort

patients prescribed 1+ type of antidepressant or methadone with a follow-up EKG 14-90 days after prescription

467

Paired Dose Cohort

Patients prescribed escalating doses of the same antidepressant (or methadone) with a follow-up EKG for each prescription
Use Case: QT interval and antidepressant use

Mean (SD) corrected QT (QTc) interval recorded on electrocardiogram 14–90 days after prescription of antidepressant or methadone, by drug dose.

* Dose a significant predictor of QTc in fully adjusted linear models at α=0.05
† QTc at specified dose is significantly different from that at prior dose in fully adjusted linear models at α=0.05
**Aim**: To develop electronic techniques for ACCURATELY identifying clinical conditions (phenotypes) in patient populations using EHR data
1 - Define the phenotype of interest. What are we looking for (e.g. a disease diagnosis, medication history, treatment response)? For example, we may be looking for a set of patients with Bipolar Disease. (It is helpful to have an idea of the population prevalence for the phenotype of interest.)
2 - Create a phenotype filter. Using coded data from the EHR, delineate the inclusion and exclusion criteria required to capture the population with the phenotype of interest. It is generally recommended to ‘cast a wide net’, that is, to create a superset from which the true phenotype population will be discerned, rather than starting with overly specific criteria that may result in inadvertent exclusion of valid patients.
Orange Zone for Phenotype Filter

Number of codes supporting the phenotype

Phenotype is Probably No

Phenotype is Probably Yes

EXISTING CODES FOR A PATIENT

Number of codes overall for Patient

against 0 supporting
Bipolar Disorder

- 328 filter positive patients (3.12%)
- Estimated prevalence: 0.78% (global), 38.20% (filter positive)

<table>
<thead>
<tr>
<th>Label</th>
<th>Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>264</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
3 - Create a gold standard training set. Conduct chart reviews on a subset of the patients obtained in step 2 to determine whether the subjects have the phenotype of interest. During the chart review, each patient is labeled as ‘Yes’ or ‘No’ in regard to whether or not they are positive for the phenotype. The number of chart reviews required is based on the prevalence of the phenotype in the population.
4 - Create a comprehensive list of features (concepts/variables) that describe the phenotype of interest.
Combination Variables

A calculated variable based on two or more data points. Examples:

- **Disease ICD-9 codes / Total ICD-9 codes**
  - Measure of disease-specific care patient is receiving within the health system. Patients may be receiving care for bipolar disease at the institution but their PCP is outside the system.

- **Total healthcare facts / Observation years**
  - Number of facts per year of care received. Compare patients with a long inpatient visit admitted for cerebral aneurysms vs. patient with lengthy longitudinal neurological care.

- **Distinct ICD-9 codes / Total encounters ( = patient_dxenct)**
  - Measure of diversity of care patient is receiving in the health system.

- **Charlson scores**
  - Comorbidity measure computed using weighted ICD-9 codes and adjusted by age.
5 - Use NLP to extract the relevant features from the set of patient notes.
Rheumatoid arthritis is an autoimmune disease that results in an inflammatory disorder affecting many tissues and organs, but primarily the joints. It can be a disabling and painful condition, which can result in limited mobility if not adequately treated.

The disease process involves an inflammatory response of the capsule around the joints (synovitis). The synovium is a synovial fluid layer that encompasses fibrous tissue (pannus) in the synovium. The pathology of the disease often leads to the destruction of articular cartilage and ankylosis (fusion) of the bones, producing diffuse inflammation in the lungs, the membranes of the lung (pleura), and white of the eye (sclera). Although the cause of RA is unknown, autoimmunity plays a big part, and RA is a systemic autoimmune disease made on the basis of symptoms, physical exam, and serological testing.
6 - Create a data analysis file for all the patients in the superset where the columns are all the features selected in step 4, both in their coded and NLP forms. The variables may exist as counts or simply as binary variables (e.g. 0=does not exist, 1=exists).
Bipolar Disorder

Features selected

patient_dxeincl
BD_COD_DX_Bipolardisorder
BD_COD_MED_MoodStabilizer
BD_NLP_antipsychotic
BD_NLP_bipolaraffectivedisorder
BD_NLP_lithium
BD_NLP_mania
BD_NLP_moodstabilizingagent
7 - Develop the classification algorithm. Using the data analysis file and the training set from step 5, assess the frequency of each variable. Remove variables with low prevalence. Apply adaptive LASSO penalized logistic regression to identify highly predictive variables for the algorithm.
Train classification algorithms

1. Sometimes over 300 words/phrases (features) are identified using chart review

2. Important features were selected for model using adaptive LASSO shrinkage

Tianxi Cai PhD and team
Bipolar Disorder

- Training within filter positive with original features

patient_dxenct                  -1.051
BD_COD_DX_Bipolardisorder        0.736
BD_NLP_antipsychotic            -0.366
BD_NLP_bipolaraffectivedisorder  1.723
Once the best model is selected, apply the algorithm to all subjects in the superset and assign each subject a probability of having the phenotype that is between 0 and 1. Select a threshold based on the desired specificity level.
Bipolar Disorder

- Training within filter positive with original features
- AUC = 0.963 (0.919 using ICD alone)

<table>
<thead>
<tr>
<th>cutoff</th>
<th>pos.rate</th>
<th>pos.num</th>
<th>FPR</th>
<th>TPR</th>
<th>PPV</th>
<th>NPV</th>
<th>NPV.all</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.944</td>
<td>0.130</td>
<td>43</td>
<td>0.000</td>
<td>0.320</td>
<td>0.994</td>
<td>0.707</td>
<td>0.992</td>
</tr>
<tr>
<td>0.905</td>
<td>0.164</td>
<td>54</td>
<td>0.008</td>
<td>0.476</td>
<td>0.973</td>
<td>0.758</td>
<td>0.994</td>
</tr>
<tr>
<td>0.783</td>
<td>0.242</td>
<td>79</td>
<td>0.017</td>
<td>0.609</td>
<td>0.941</td>
<td>0.812</td>
<td>0.996</td>
</tr>
<tr>
<td>0.670</td>
<td>0.296</td>
<td>97</td>
<td>0.045</td>
<td>0.712</td>
<td>0.875</td>
<td>0.859</td>
<td>0.997</td>
</tr>
<tr>
<td>0.649</td>
<td>0.310</td>
<td>102</td>
<td>0.058</td>
<td>0.747</td>
<td>0.858</td>
<td>0.873</td>
<td>0.997</td>
</tr>
<tr>
<td>0.631</td>
<td>0.324</td>
<td>106</td>
<td>0.064</td>
<td>0.775</td>
<td>0.859</td>
<td>0.884</td>
<td>0.998</td>
</tr>
<tr>
<td>0.608</td>
<td>0.344</td>
<td>113</td>
<td>0.078</td>
<td>0.799</td>
<td>0.848</td>
<td>0.894</td>
<td>0.998</td>
</tr>
</tbody>
</table>

Prevalence 0.008 (global), 0.382 (filter positive)
9 - Validation of the algorithm. Conduct chart reviews of a set of subjects classified as having the phenotype of interest, along with a random set of subjects from the superset to ascertain the performance of the algorithm.
Can We Trust the Phenotypes?

Validation Study (N = 185)

- Evaluate case and control algorithms compared to gold standard of diagnostic interview by expert clinician
- Recruit cases and controls as defined by informatics algorithm
- Interview by clinicians blinded to ascertainment group

Jordan Smoller MD, ScD and team
Prenatal antidepressant exposure is associated with risk for attention-deficit hyperactivity disorder but not autism spectrum disorder in a large health system.

Clements CC1, Castro VM2, Blumenthal SR1, Rosenfield HR1, Murphy SN3, Fava M4, Erb JL5, Churchill SE6, Kaimal AJ7. 


Incident user cohort study of risk for gastrointestinal bleed and stroke in individuals with major depressive disorder treated with antidepressants.

Castro VM1, Gallagher PJ, Clements CC, Murphy SN, Gainer VS, Fava M, Weilburg JB, Churchill SE, Kohane IS, Iosifescu DV, Smoller JW, Perlis RH.


An electronic health records study of long-term weight gain following antidepressant use.

O'Dushlaine C1, Ripke S1, Ruderfer DM2, Hamilton SP3, Fava M4, Iosifescu DV2, Kohane IS5, Churchill SE6, Castro VM6, Clements CC4, Blumenthal SR4, Murphy SN6, Smoller JW7, Perlis RH8.


Rare copy number variation in treatment-resistant major depressive disorder.

QT interval and antidepressant use: a cross sectional study of electronic health records.


Antidepressant response in patients with major depression exposed to NSAIDs: a pharmacovigilance study.

Gallagher PJ1, Castro V, Fava M, Weilburg JB, Murphy SN, Gainer VS, Churchill SE, Kohane IS, Iosifescu DV, Smoller JW, Perlis RH.
Validated Phenotypes Important for Basic Researchers
Collaborators

- I2b2 and SMART
  - Isaac Kohane
  - Susanne Churchill
  - Michael Mendis
  - Lori Phillips
  - Jeff Klann
  - Janice Donahue
  - Griffin Weber
  - William Simons (SHRINE)
  - Doug McFadden (SHRINE)
  - Ken Mandl (SMART)
  - Josh Mandel (SMART)

- Medical Imaging (mi2b2)
  - Christopher Herrick
  - David Wang
  - Bill Wang

- I2b2 Driving Biology Projects
  - Vivian Gainer
  - Victor Castro
  - Andrew Cagan
  - Jordon Smoller
  - Roy Perlis
  - Dan Iosifesco
  - Scott Weiss
  - Elizabeth Karlson
  - Katherine Liao
  - Ashwin Ananthakrishnan
  - Tianxi Cai
  - Sheng Yu
  - Stanley Shaw
  - Zongqi Xia