Using AI and Natural Language Processing to Uncover Population Social Determinants of Health Factors
Session #183, February 13, 2019

Varun Gupta, IT Director, Analytics and Data Management, Mt. Sinai Medical Center
Trish Birch, SVP and Global Consulting Leader, Cognizant Healthcare
Conflict of Interest

Varun Gupta
Has no real or apparent conflicts of interest to report.

Patricia Birch, MBA
Has no real or apparent conflicts of interest to report.
Agenda

• Welcome & Introduction
• Learning Objectives
• Using AI and Natural Language Processing to Uncover Population Social Determinants of Health (SDoH) Factors
• Improving Health Outcomes with SDoH
• Q&A
**Learning Objectives**

Discuss how social determinants of health help providers improve clinical models, understand disease progression and identify phenotypes.

Apply methods for cost effectively and accurately extracting social determinants of health (SDoH) data from rich content in unstructured clinical notes.

Describe best practices for defining and creating an initial ontology set.

Explain process of training natural language processing algorithms to accurately identify SDoH in patient record notes.
Using Natural Language Processing to Uncover Population SDoH Factors
About Mount Sinai

8 Hospitals
1 Medical School
9 Ambulatory Surgical Centers
190 Remote Clinical & Administrative Sites
38,000 Employees
7,800 Physicians
6,340 Faculty Members
3300+ Beds
138 Operating Rooms
3+ million Outpatient Visits
500+ thousand ER Visits
130+ thousand Inpatient Visits
425,000 IT Service Desk Calls
Moving the Needle: Population Health and Analytics

Enabling IT Initiatives
- Establish interoperability across clinical systems
- Onboard applications and technologies to drive collaboration
- Enable ‘Data as a Service’

Fee for Service
Value Based Services

Acute Care ( $$$ )
Primary Care ( $$ )
Community Services ( $ )

#HIMSS19
Natural Language Processing in Healthcare

Social Determinants Identification
- Notes (Progress, admission, procedure, consultation notes); discharge summaries

Clinical Trials Recruitment
- Identification of patient cohorts

Improved Clinical Models
- Disease progression, phenotypes identification

Revenue Cycle Management
- Identify labs, procedures, medications, etc.

Computer-Assisted Coding/ Risk Adjustments
- Billing for correct services

Determining Risk Scores
- Smoker who quits → reduced risk of heart failure

Unstructured
- Behavior (smoking, substance abuse etc.)
- Education
- Medications
- Discharge Education
- Economics
- Safety
- Support
- Cultural
- Lab Results
- Demographics
- Prior Admissions
- Clinical (current diagnostics, family history)

Structured
Technology Framework

On-Premise

- Fast Data Ingestion
- Prioritized Use Cases
- Extended team model
- Unified Data Desk

On – Cloud

- Robust data ingestion
- FHIR-based API platform
- Hadoop data lake
- Hybrid cloud-based architecture on Azure
- Unified data desk and extended data teams

- API Connect
- Big Data Ecosystem
- Advance Analytics
- Salesforce

- Enable data as a service
- Expand data mgmt. capability
- Analytics ecosystem
- Enterprise CRM Expansion
### NLP Process – Categorization

<table>
<thead>
<tr>
<th>Economic Stability</th>
<th>Care Giving Responsibility</th>
<th>Legal</th>
<th>Support System</th>
<th>Physical Activity</th>
<th>Fall Safety</th>
<th>Special Health needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Housing, Income</td>
<td>• Patient’s care giving responsibilities are having adverse effect on Patient’s health</td>
<td>• Legal issues like divorce, child custody, alimony etc.</td>
<td>• Formal support: health home etc.</td>
<td>• How active patient is and does patient have avenues to stay active</td>
<td>• Includes, Identification of measures (not) taken for fall safety</td>
<td>• Special health needs like vision, hearing, and Urinary incontinence</td>
</tr>
<tr>
<td>• Food and Nutrition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Lack of education or Current education level causing any challenges</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthcare System</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Medication, Prevention care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Insurance, Language barriers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical Environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Transportation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Concerns on overall safety and well being</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Depression/anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Substance abuse and Sleep relates problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Legend:**
- Phase 1: Economic Stability
- Phase 2: Care Giving Responsibility
- Phase 3: Legal
- Phase 2: Support System
- Phase 2: Physical Activity
- Phase 2: Fall Safety
- Phase 2: Special Health needs
NLP Workflow

Clinical Notes Extraction → Duplicate Removal → Collation of Notes → Corpus Creation

Garbage Filtering → Spell Check → Synonym Replacement

Lemmatization → Bag of Words → Document Term Matrix → Document Classification
Model Outcomes

All Medicaid notes added to the database during Dec ’17 to Jul’18 were used for the model execution. Separate sample outputs for each model were manually validated by clinical SMEs to drive model parameters.

- **Economic Stability**
  - 97% Accuracy
    - Sensitivity: 96%
    - Specificity: 99%

- **Education**
  - 93% Accuracy
    - Sensitivity: 97%
    - Specificity: 88%

- **Healthcare System**
  - 89% Accuracy
    - Sensitivity: 91%
    - Specificity: 85%

- **Physical Environment**
  - 86% Accuracy
    - Sensitivity: 98%
    - Specificity: 75%
#### Capturing Insights

**Time Period**

Aug 2016 to Oct 2018

**Medicaid**

7,240,320 Encounters
226,368 Patients

<table>
<thead>
<tr>
<th>SDoH Category</th>
<th>Patient Count</th>
<th>% of Patients with SDoH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic Stability</td>
<td>35,187</td>
<td>16%</td>
</tr>
<tr>
<td>Healthcare</td>
<td>35,781</td>
<td>16%</td>
</tr>
<tr>
<td>Education</td>
<td>8,770</td>
<td>4%</td>
</tr>
<tr>
<td>Physical Environment</td>
<td>18,685</td>
<td>9%</td>
</tr>
<tr>
<td><strong>Total Patients with SDoH</strong></td>
<td><strong>70,541</strong></td>
<td></td>
</tr>
</tbody>
</table>

31% of total patients who had an encounter Aug 2016-July 2018 were found to have at least one SDoH factor.
Next Steps to Scale

1. Economic Stability
2. Education
3. Physical environment
4. Health system
5. Behavioral Health
6. Caregiving responsibilities
7. Legal, Support system
8. Physical activity

Data Scientists

EMR / Data Repositories

Extraction of Clinical notes & Social determinants survey data

Analyze, Model, Train and Deploy

NLP Model & SDH Repository at patient level

Load libraries

Industry standard SDOH libraries

Unified Medical Language system

Data Access

Care Gives / Population Health Managers

Next Steps to Scale
Challenges

- Creating an ontology unique to Mt. Sinai, but universal enough to effectively capture SDoH
- Testing the accuracy of the algorithm’s output was labor-intensive.
- Integrating findings into clinical workflows and systems
- Integrating the generated outcome back to the individual patient EMR

Best Practices

- Analyze your available data
- Identify opportunities: process fit, most impactful SDoH
- Start simply, but be comprehensive
- Develop a roadmap
Improving Health Outcomes with SDoH
Healthcare to Social care

What Impacts Health Outcomes?

- Biology (Genetics), 30%
- Physical Environment, 5%
- Social Circumstances, 15%
- Individual Behavior, 40%
- Medicare Care, 10%

Why should you care?
- Higher utilization
- Increased presence of chronic disease
- Higher spend

---

1 Maricopa County
2 n = 305,381
Opportunity Analysis: Impact on Payers & Providers

Health Plans

• High Cost of Care
• Unpredictable Actuarial Pool
• Member Satisfaction

Health Systems

• Increased cost of care and financial risk
• Loss of Clinical time and resources
• Alignment of new mission: treating whole person and community
Today’s SDoH Tools are Inefficient

Lack of Integration Across the Care Journey

- Manual Documentation
- Lack of Integration across the Care Journey
- Manual Entry into EMR
Lack of Integration Across Care Journey

<table>
<thead>
<tr>
<th>Inefficient Integration</th>
<th>Upon identification of a SDoH, there’s no efficient way to communicate and integrate this information across the care continuum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referral Resources</td>
<td>Poor management of referral resources leads to difficulty identifying which resources will be most effective and supportive for a patient’s needs</td>
</tr>
<tr>
<td>Overburdened System</td>
<td>Care Managers and Patient Navigators are often responsible for a handful of patients at once</td>
</tr>
</tbody>
</table>
Digital Tools can Uncover and Integrate SDoH into Today’s Care Continuum

- Leverages pre-existing data
- Creates efficiency during care touchpoints
- Allows for automatic integration into the care continuum
SDoH: Whole Person Care

Medical Excellence

Mental Health & Wellness

Social Care
Caring for the Whole Person Requires Transformation at all 3 Levels

What can be done to better identify and integrate SDoH into a patient’s care journey across all 3 levels?

Patient  Health System  Community
SDoH Transformation Takeaways

1. Solutions must be efficient
2. Integration across the care journey is needed
3. A comprehensive approach is required
Questions

Varun Gupta
IT Director, Analytics And Data Management, Mt. Sinai Medical Center
https://www.linkedin.com/in/varungupta7/

Patricia (Trish) Birch
Senior Vice President and Global Consulting Practice Leader, Cognizant
Patricia.Birch@Cognizant.com
Appendix
## Sample Data

<table>
<thead>
<tr>
<th>Enc. ID - Anonymized</th>
<th>Note Text – Subset</th>
<th>Model Tenet</th>
<th>Model Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>xxxxxx85919</td>
<td>Physical Examination  ED Assessment/Plan ASSESSMENT: …single, on SSI, domiciled at Main Chance Drop-In Center, with PMH HTN, with multiple reported prior psych hosp., no reported SAs or NSSIBs, reported hx of violence in the context of incarceration (2012-2015, convicted of felony) robbery, denies violent offense, reports completed parole…..</td>
<td>Identification confirmation</td>
<td>Social Context tagged (the model successfully tagged ‘context of incarceration’ and ‘convicted of felony’).</td>
</tr>
<tr>
<td>xxxxxx62414</td>
<td>SW will also provide patient with contact info to legal aid to help with her court case with her custody battle for her children. Has a prior 8B admission in the past with DX of Borderline Personality Disorder, recent admit to Kings County for detox in October 2017, in ongoing custody battle with mother over her 3 children.</td>
<td>Identification Confirmation</td>
<td>Legal Context tagged (the model successfully tagged ‘ongoing custody battle’, ‘legal aid’).</td>
</tr>
<tr>
<td>xxxxxxx36126</td>
<td>AED SW spoke to patient’s son, (xxx-xxx-xxx), who reported significant caregiver burnout, stating that patients HHA at Big Apple Homecare, has not been coming, and that it has been difficult for (xxx-xxx-xxx) and patient’s daughter, (xxx-xxx-xxx), to supervise the patient at home.</td>
<td>Non detection-Negation</td>
<td>No caregiving issue (the model successfully avoided tagging ‘caregiver burnout’ since it is patient’s son who is the caregiver).</td>
</tr>
<tr>
<td>xxxxxx78681</td>
<td>Problem: Social Isolation/Impaired Social Skills Goal: Patient will begin to interact with others Outcome: Progressing ….. Patient has Medicaid metropolis, Patient has been to ETPH rehab in Bronx, patient feels depressed due to ETOH and lack of work and recent divorce.</td>
<td>Identification Confirmation</td>
<td>Social Context and Legal tagged (the model successfully tagged ‘social isolation’ and ‘recent divorce’)</td>
</tr>
</tbody>
</table>