Using Natural Language Processing to Identify Vulnerable Patients with Psychosocial Needs
Sumeet Singh1, MBBS, Nidhi Pai1, Christie Pizzimenti, PhD1, Ana Quiñones, PhD, MS2 David A. Dorr, MD, MS1

1Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR; 2Department of Family Medicine, Oregon Health & Science University, Portland, OR

Introduction
Approximately 75% of adults age 65 years and older live with multimorbidity, defined as the concurrent presence of two or more chronic diseases and account for a disproportionate share of health care expenditures. Our prior work on patients with multimorbidity shows decreased emergent health care utilization and disease exacerbation when care teams incorporate documented psychosocial needs into tailored health care plans, acknowledging the utility of psychosocial vital signs in electronic health records (EHR). Bejan et al. used lexical association and word2vec to mine social concepts, such as homelessness, from a large EHR repository. Here, we replicate this concept in a local EHR repository and extend this work to other psychosocial concepts—chronic stress, financial insecurity, and social isolation.

Methods
We extracted and standardized EHR notes from 120 primary care and specialty clinics in Oregon, representing 3.23 million patients. From this, a subset of 2016 and 2017 data were extracted for a total of 80,000 patients. Using lexical association and expert input, we identified seeds most likely to be used in narrative notes representing four psychosocial vital signs—chronic stress, social isolation, financial insecurity, and homelessness.

A modified bag of words approach was applied to identify notes with a minimum one seed word, in patients 65 years of age and older with multimorbidity from the aforementioned subset. These notes were then vectorized using word2vec, with higher weights assigned near selected seeds in notes. Negex was used for negation. Terms with higher average weights were examined to determine potential new seeds. The relevance of identified patients to psychosocial vital sign of interest was assessed by the standard term frequency–inverse document frequency (TF-IDF) weighted cosine metric. To assess the precision of retrieved patients and to facilitate query reformulation, research staff and faculty completed a manual annotation process. This validation was completed in the top 100 patients based on relevance score for each concept. While the main corpus of notes included 80,000 patients, a range of 14,507 to 609 patients from all four concepts met the inclusion criteria for validation—patients 65 years of age and older with multimorbidity whose notes contained minimum one seed word.

Results
NLP identified 12-18% of patients with psychosocial issues compared to structured data (ICD-10 diagnosis codes) which showed that less than 0.2%. As illustrated in Table 1, NLP matched substantial numbers of patients moderate to strong precision (.78–.94). The average score across all four concepts ranged from 3.93–4.64.

Discussion
Using NLP to identify vulnerable older adults with multimorbidity and high psychosocial needs prior to disease exacerbation is a lower cost, higher efficiency method with multiple implications for health care policy, clinical care, and refinement of clinical guidelines.

Table 1. NLP approach to extract subset of Psychosocial Vital Signs

<table>
<thead>
<tr>
<th>Topic</th>
<th>Matches (Total N = 80,748)</th>
<th>IPF range</th>
<th>Relevance (Mean 95% CI of top 600)</th>
<th>Patient precision (Patients relevant / patients reviewed)</th>
<th>Average score (0–5, 5 highly likely), min-max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chronic stress</td>
<td>14,507 (18.0%)</td>
<td>2.03 - 7.61</td>
<td>51.35 (48.4-54.6)</td>
<td>0.90 (90/100)</td>
<td>4.38</td>
</tr>
<tr>
<td>Social isolation</td>
<td>609 (8.5%)</td>
<td>5.94 - 10.90</td>
<td>14.75 (12.3-18.2)</td>
<td>0.80 (80/100)</td>
<td>4.02</td>
</tr>
<tr>
<td>Financial insecurity</td>
<td>6,945 (8.6%)</td>
<td>2.43 - 10.90</td>
<td>22.70 (19.8-26.4)</td>
<td>0.78 (78/100)</td>
<td>3.93</td>
</tr>
<tr>
<td>Homelessness</td>
<td>4,449 (5.5%)</td>
<td>3.06 - 10.90</td>
<td>64.95 (48.9-82.9)</td>
<td>0.93 (93/100)</td>
<td>4.64</td>
</tr>
</tbody>
</table>

+ Note: 2017 seed for Social Isolation was companionship/friendship
** Abbreviations: a&dh/ads: Accessibility and Disability Services, aps: Adult Protective Services, afh: Adult Foster Home

References: