Problem Statement

- Temporal information is important to understand self-reported depression diagnosis statements in online forums.
- We provide a dataset and benchmarks for temporal annotation of self-reported diagnosis statements based on RSDD.
- Dataset available with DUA at http://ir.cs.georgetown.edu/resources/

Annotation Scheme

- **Self-reported diagnosis statement**
  Where the user states his or her diagnosis
  
  *Back when I was 14, I was diagnosed with…*
  *He diagnosed me with depression, but I…*

- **Diagnosis time**
  Temporal expression indicating when the diagnosis was made
  
  *Back when I was 14, I was diagnosed with…*
  *I was diagnosed with depression in 2008.*

- **Diagnosis in doubt**
  Does the user believe the diagnosis?
  
  *He diagnosed me with depression, but I never really believed it.*

- **Diagnosis recency**
  How recent was the diagnosis?
  
  - Very recently (up to 2 months ago)
  - More than 2 months, up to 1 year
  - More than 1 year, up to 4 years
  - More than 3 years ago
  - Unspecified
  - Unspecified, but not recent

- **Condition state**
  Is the condition current?
  
  - Current (and “probably” current)
  - Unknown
  - Past (and “probably” past)

- **False positive**
  Indication of whether the diagnosis statement is real
  
  *My friend told me, “I was diagnosed with depression.”*

Dataset Characteristics

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>false positive diagnosis in doubt</td>
<td>25 of 598</td>
</tr>
<tr>
<td>diagnosis in doubt</td>
<td>16 of remaining 573</td>
</tr>
<tr>
<td>condition state</td>
<td>current (254), prob. current (64), unknown (225), prob. past (29), past (26)</td>
</tr>
<tr>
<td>diagnosis recency</td>
<td>unspecified (232), unspecified, but past (176), recent (27), &gt;2m-ly (3), &gt;3-ly (7)</td>
</tr>
<tr>
<td>time expression</td>
<td>explicit (144), inferable from age (104), non-inferable (47), na (398)</td>
</tr>
</tbody>
</table>

598 posts were annotated. See above for distribution of attributes.

Baselines

**Diagnosis recency**

- Modified SUTime for this task by adding rules for age expressions, and using heuristic to pick the date expression closest to the diagnosis statement.