Cyberchondria
Web content can lead to escalations from concerns about typically-benign medical symptoms to searches on frightening disorders. We aim to predict such escalations based on structure and content of Web pages. Search engines could use escalation likelihood as a ranking feature or present this information to users.

Identifying Query Escalations
- Anonymized health-related Web search sessions mined from toolbar logs
- Identified transitions from queries about symptoms to queries for related, serious/benign conditions

Next-query escalation: Following visit to page $P$, escalation in next query

Defined non-escalation: Follow-up query is benign as explicitly defined

Predicting Query Escalations
- Built classifier to predict if query after page $P$ contains escalation / non-escalation using features of $P$:
  - structural (e.g., serious illness precedes benign explanation)
  - title and URL (e.g., title has serious illness)
  - first-person testimonials (e.g., contains phrases such as “I felt” and “I am worried”)
  - reliability (e.g., page is externally verified by healthonnet.org)
  - commercial intent (e.g., contains advertisements), and
  - general (e.g., length in words, total KBs)

Probing Individual Features
E.g., Escalation likelihood appears related to order in which serious and benign conditions appear on preceding Web page $P$

<table>
<thead>
<tr>
<th>Query outcome</th>
<th>Order of presentation</th>
<th>True Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escalation</td>
<td>Serious first</td>
<td>68.6%</td>
</tr>
<tr>
<td>Non-escalation</td>
<td>Benign first</td>
<td>31.4%</td>
</tr>
</tbody>
</table>

Predicting Escalations Using All Features
Best prediction accuracy: 73.4% (marginal = 50%)

Predictive page features:
1. serious illness precedes benign explanation,
2. serious illness or benign explanation appears in title or near start of page,
3. page from forum, and
4. page externally verified