Studies of the **Onset & Persistence** of Medical Concerns in Search Logs

Ryen White and Eric Horvitz
Microsoft Research, Redmond
{ryenw,horvitz}@microsoft.com

Motivation

- People use the Web to diagnose and understand medical conditions
- 80% of U.S. Web users perform medical search (Pew, 2011)

- Studies to date have focused on the **impact of the Web** on medical search
  - Including **cyberchondria** (White & Horvitz, TOIS 2009) – Escalations in concerns within a single session (e.g., [headache] → [brain tumor])

- Search behavior **over time** reflects in-world events such as professional diagnosis, etc. **[Our Focus]**

Our Focus: Logs as a Lens on Health

- We explore **how medical concerns emerge over time** and **how concerns persist post onset**

- Long-term needed to understand medical search
  - **60% of medical sessions start directly with a condition query**
    - Searchers come to search engines with a condition in mind!
  - 95% of these sessions had medical query in prior session(s)

- Long-term helps us understand medical trajectories
  - Predict emerging concerns, personalize search, guide healthcare use

- **Note:** Long-term may be influenced by external events (e.g., diagnosis) – Web not always **cause**

Contributions

- Large-scale log analysis to:

1. Characterize how health-related concerns emerge in search activity over time
2. Analyze persistence of concerns post onset, including interruptions to future search tasks
3. Build models to predict condition onset, escalations in concerns, and future interruptions

- **Will focus on #1 and #2 in this talk**

Data

- Anonymized toolbar logs from millions of users
- 3 month period from Jan – Mar 2011
- Extracted search sessions (30 min timeout)
  - Included queries from Google, Yahoo!, Bing

- Assigned topical labels from ODP to URLs
- Content-based classifier
- Identify which URLs are medical

Long-Term Profile Generation

- Identify searchers with evidence of health seeking intent
- Searchers who visited at least one URL in *Health* category
- Removed outlier users

- 170k users, 25M sessions, 1B+ URLs
- 150 search sessions / user, spanning avg. of 76 days / user

- Automatically labeled queries based on:
  - **Symptoms:** Merck medical dictionary + synonyms from click graph
    - E.g., headache
  - **Benign explanations:** ICD10 list + synonyms from click graph
    - E.g., caffeine withdrawal
  - **Serious illnesses:** ICD10 list + synonyms from click graph
    - E.g., brain tumor

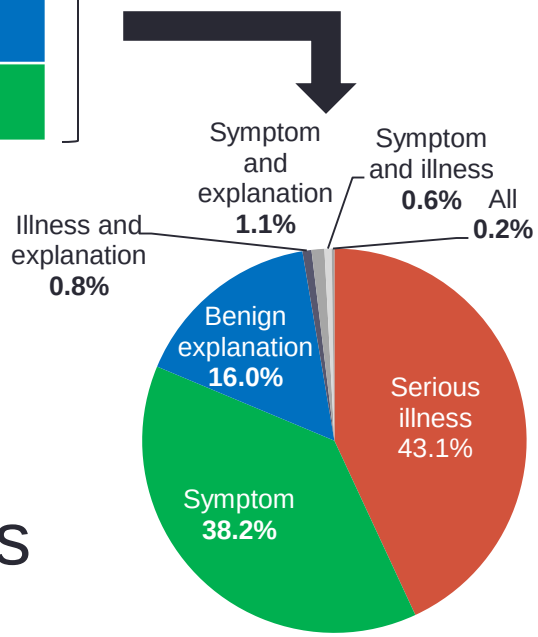
What are some key characteristics of the long-term health-search profiles?

Medical Search Over Time

Statistic (n=169,513)	Mean	SD
# medical queries (based on clicks)	36.67	60.29
% queries medical	3.4%	4.3%
# medical sessions	37.04	32.05
# medical queries (wordlist method)	10.31	17.90
# queries with serious illnesses	4.60	8.68
# queries with benign explanations	1.86	5.81
# queries with symptoms	4.15	11.39

ODP classes of SERP clicks

Dictionary lookup

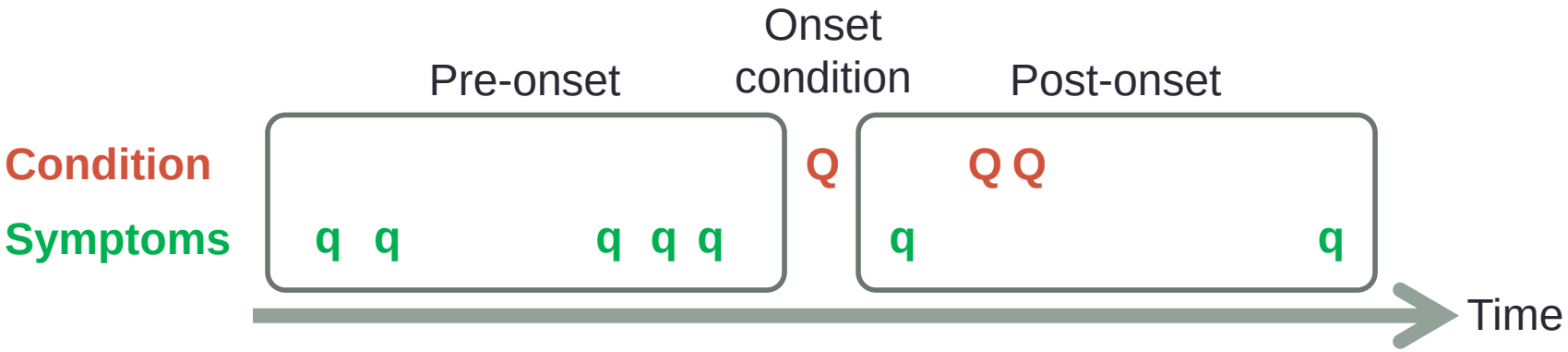


- 3% of queries are medical related
- Users search more often for symptoms and serious illnesses than benign, common explanations

What are characteristics of how health-related concerns emerge in activity over time?

Condition Onset

- Explore search behavior before the onset of the first occurrence of a condition – randomly chosen



- Pre-onset history can contain other conditions
  - 83% searched for at most one other condition prior to onset
  - 61% searched for no conditions pre-onset
- Also extracted behavior after the condition onset
- **Overall:** 159K pre-onset histories, one per user

Pre-Onset Behavior

- Extracted features of search behavior before onset
- We computed statistics of search behavior, e.g.,
  - 20 medical search sessions over 49 days before condition
  - 3% of user's time online was spent viewing medical content
- Additionally, we wanted to check if onset condition and the symptoms searched previously were related
- Extracted symptom-onset pairs, compared with known
  - E.g., [headache] → {migraine, caffeine withdrawal, brain tumor}
- **79.5% of prior symptoms were related to onset**
  - **Emergence of conditions appears to extend back over time**

Condition Severity

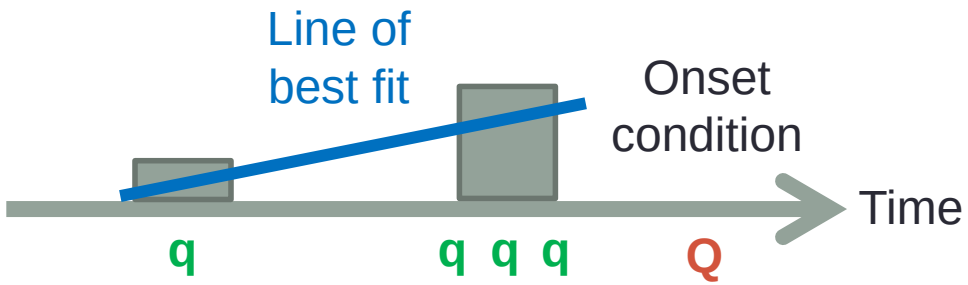
- Study the relationship between pre-onset and severity
- Many search / Web usage behaviors are similar, but …

	Onset condition is …	
Feature	**Serious (n=117389)**	**Benign (n=34424)**
% medical queries	3.0% (4.0%)	4.0% (4.2%)
# medical queries	5.30 (12.12)	7.69 (15.48)
# symptom queries	1.81 (6.43)	2.34 (7.58)
# condition queries	1.74 (5.36)	2.27 (6.36)

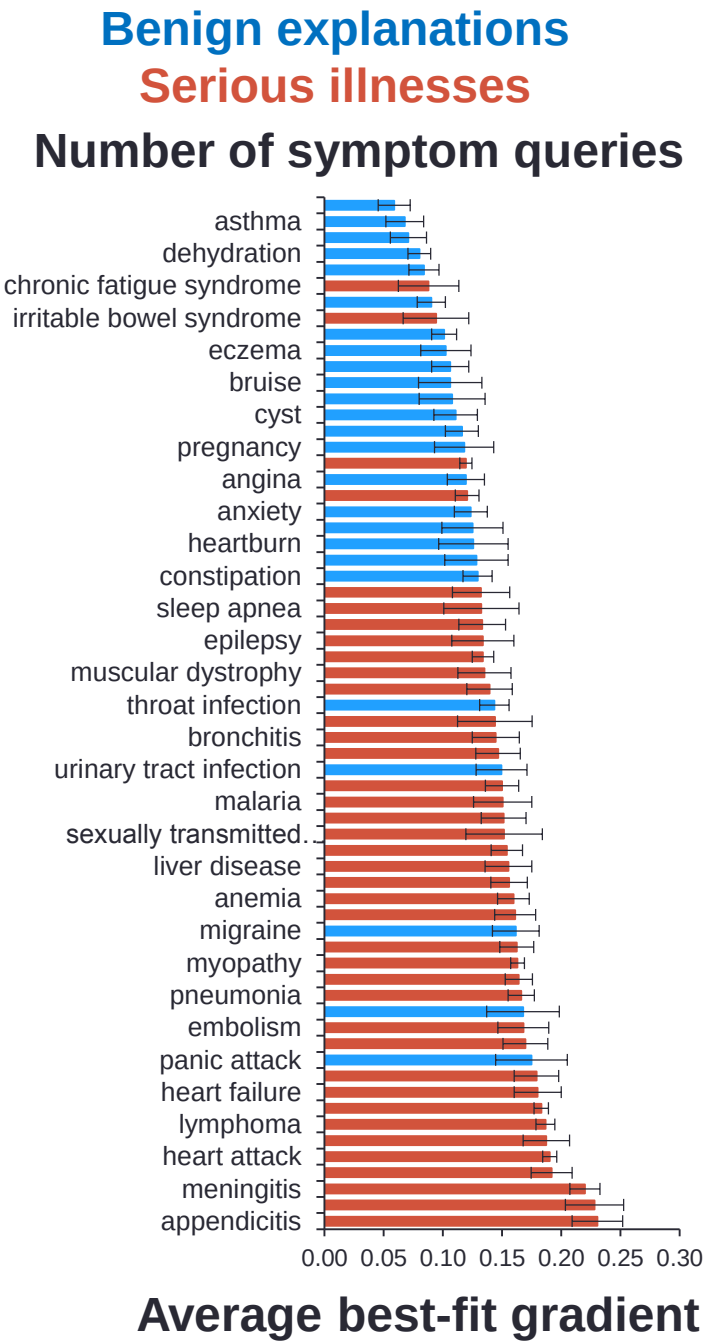
- Onset condition = benign, more medical searching
- Web **may** be contributing as a causal influence
  - Users may be better informed and less likely to escalate
  - Narrow lens of log analysis means that we cannot be sure **why**

Feature Trends

- Compute gradient of line of best fit of the features over time before the condition onset



- Observations:
  - Condition acuity increases top-to-bottom
  - Benign explanations on bottom are acute
    - e.g., food poisoning and panic attacks
  - Serious illnesses on top are chronic
    - e.g., chronic fatigue syndrome



How do searches for the condition persist following evidence of initial concern onset?

Post-Onset Search Behavior

- Computed the same features for search behavior **after the onset condition** rather than before

Feature	% or Avg (SD)	% change from pre-onset
% URLs medical	4.9%	+88.5 ↑
% queries medical	4.2%	+31.3 ↑
% online time on medical pages	8.0%	+247.8 ↑
# unique symptoms	0.50 (1.01)	−20.6 ↓
Symptom persistence (days)	2.46 (3.42)	−27.9 ↓
# unique	1.04 (1.29)	+40.5 ↑
Condition	7.57 (12.49)	+25.3 ↑

Most people searched for zero conditions pre-onset

- Medical search increases, symptom searching decreases, and condition searching increases (more focused)

Interruption: Effect of Health Concerns

- Suspension of ongoing search activity
- Given a search session, an interruption comprises at least one search for the onset condition sandwiched between two series of 1+ non-medical searches, e.g.,

- We show that interruption happens **<u>often</u>**:
  - 30.3% of users were interrupted in this way
  - 2.39 interruptions / user, totaling 2 hours and 16 mins. searching
  - 6% of future search sessions interrupted, on 2.68 total days

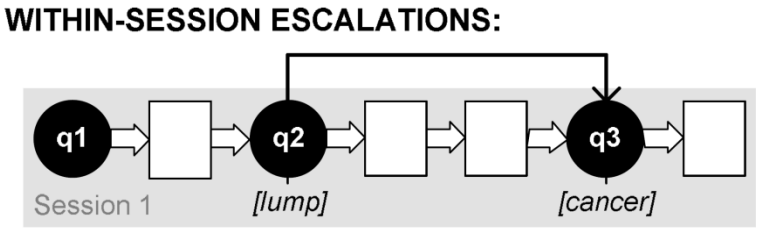
Summary and Takeaways

Summary of Presented Findings

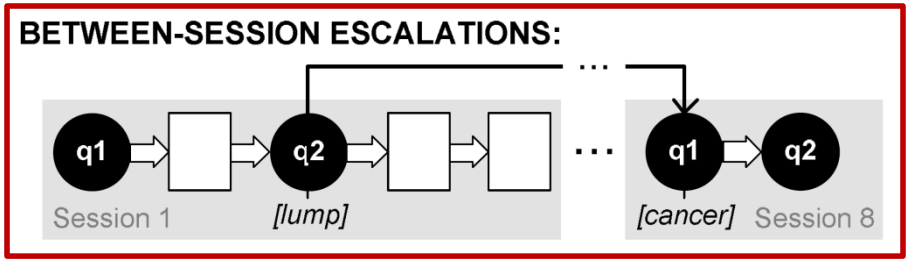
- 80% of pre-onset symptoms relate to onset condition
  - True even though the search queries are weeks apart
- Medical search increases near to onset
  - Related to acuity of condition
- Onset type is related to pre-onset behavior
  - e.g., Benign explanations follow more medical searching
- Medical search behavior post-onset ≠ pre-onset
  - e.g., Symptom search drops, condition search increases
- Post-onset interruptions occur frequently

Other Things (see paper for more info)

- Between-session escalations in medical concerns



(White & Horvitz, 2009)



(New in this paper)

- Focus-of-attention
  - Users narrow focus over time – focus on particular conditions of interest
- Periodicities in medical search
  - Medical search behavior is clustered, not spread out over time
- Accurate prediction of onset, escalation, interruption

Implications & Limitations

- Onset as cue to use previous clicks for related symptoms
- **Use combination of symptoms searched over time to <u>estimate the onset condition</u> (early warning signs!)**
  - Observed over time: [twitching], [twitching eye], [twitch], …
  - Unobserved (yet!): [ALS] [Lou Gherig's Disease]
- Given accurate prediction of:
  - **Onset**: Adjust search experience based on severity or prevalence
  - **Escalation**: Provide links to broader content (benign explanations)
  - **Interruptions**: Provide support for more effective task switching
- Limitations of log analysis:
  - Machines may be shared (not all queries from same user)
  - Limited definitions of concepts (e.g., interruptions)

Conclusions

- Log-based longitudinal study of online health search
- Characterized aspects of that behavior before condition onset – relationship with condition type, escalations
- Studied post-onset behavior and found differences in search behavior, and evidence of interruptions
- Also built predictive models of onset, escalations, impact
- Search engines could mine long-term health seeking behavior to provide better medical-search support