Mining and Modeling Online Health Search

Ryen W. White
Microsoft Research
Redmond, WA USA
ryenw@microsoft.com

In collaboration with many, including
Eric Horvitz, Ahmed Hassan, Robert West, Adam Fourney, Michael Paul,
Marc Cartright, Rave Harpaz, Nina Mishra, et al.

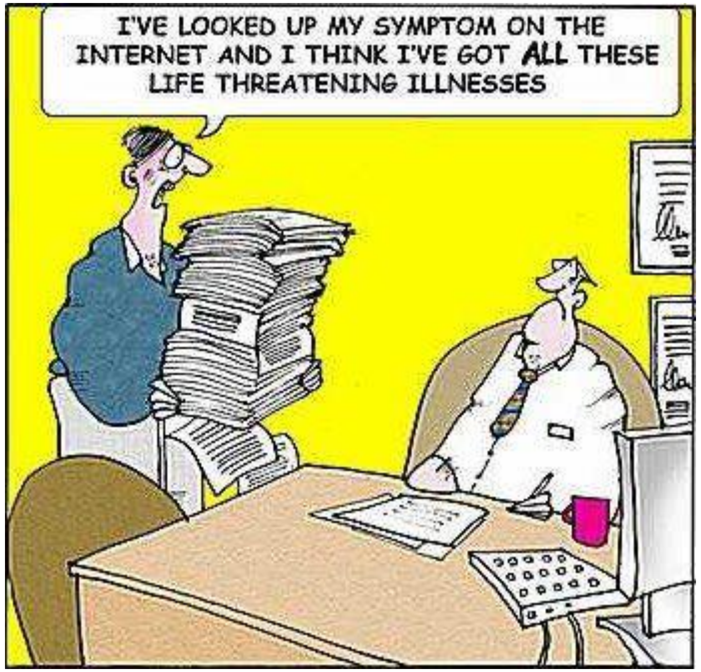
Outline

- Online Health Search
  - Short- and Long-term health searching
  - “Cyberchondria”
  - “Web to World” transitions
- Searcher and Content Biases
- Mining Health Search Data
  - 3 applications: Nutrition Tracking, Pregnancy Prediction, Detection of Drug Interactions and Adverse Drug Reactions
- Opportunities and Challenges

Part I: Online Health Search

Health Seeking

- Healthcare websites for worried (un)well
  - Provide valuable information, address concerns, etc.
- 80% U.S. adults use search engines to find medical info (Pew, 2011)
  - Majority don't verify quality (validity, date, etc.)
- Problem: Search engines for diagnostic reasoning
  - Link to pages with potentially-alarming content
  - More written about serious than benign explanations
  - Ranking algorithms use click logs;
    ignore likelihoods, reinforce alarming pages



Biased Health Content in Web Search

- Web search suffers from and amplifies biases of judgment
  - *Base-rate neglect*
  - *Availability bias*
  - *Confirmation bias*

→ *Cyberchondria!!*

(White and Horvitz, TOIS 2009)

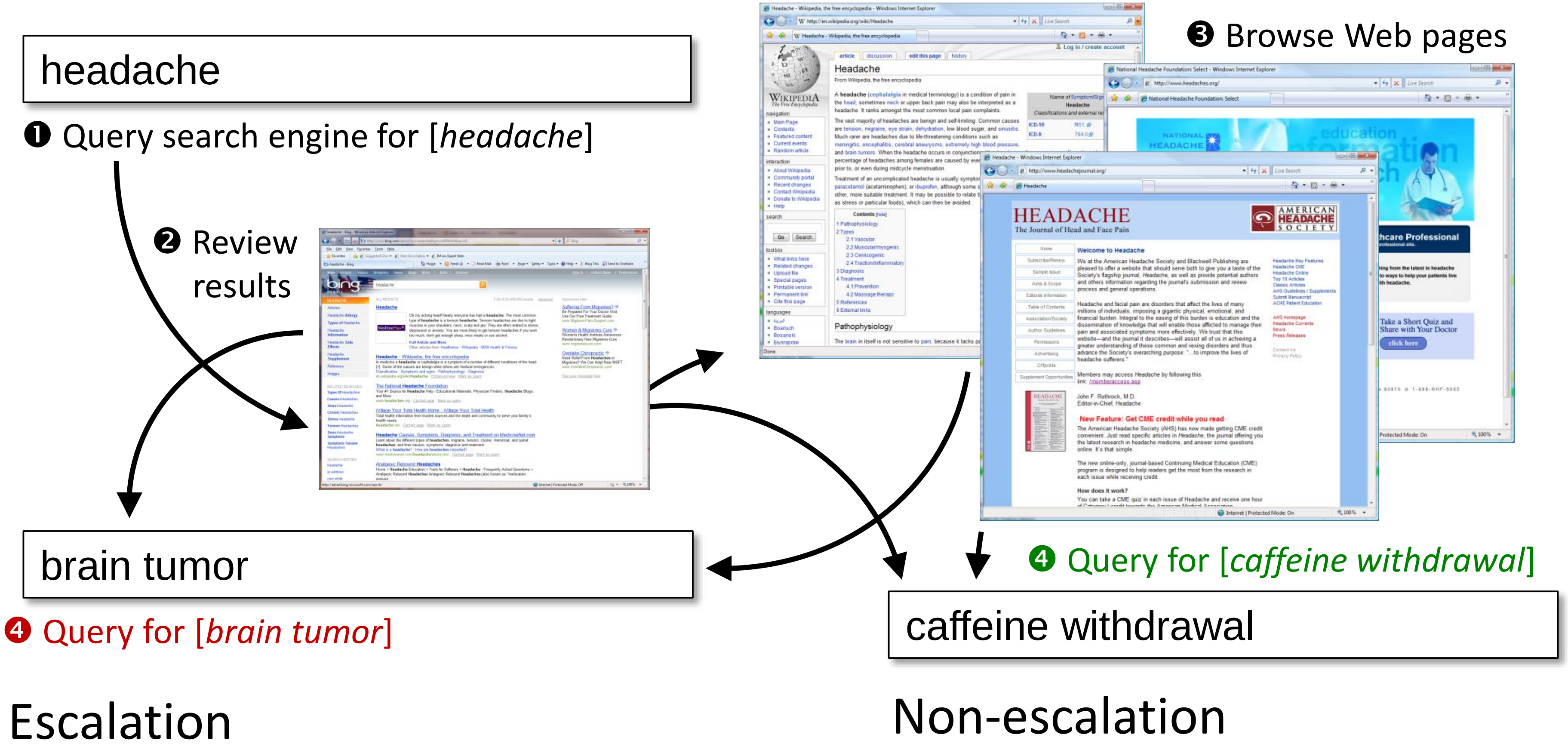
% pages with co-occurrence of symptom and cause (circa 2007)

Symptom	Cause	Web Crawl		Top 100 from Web Search	Top 100 from Domain Search (MSN Health)
headache	caffeine withdrawal	29%		26%	25%
	tension	68%		48%	75%
	brain tumor	3%	→	26%	0%
muscle twitches	benign fasciculation	53%		12%	34%
	muscle strain	40%		38%	66%
	ALS	7%	→	50%	0%
chest pain	indigestion	28%		35%	38%
	heartburn	57%		28%	52%
	heart attack	15%	→	37%	10%

Co-occurrence of symptom & serious condition most common in Web search

Cyberchondria: Escalation in Health Concerns

- Users transition from common symptoms to rare, but serious diseases
  - *e.g.*, {headache, nausea, dizziness} → malignant brain tumor



Survey to Understand Self-Diagnosis Online

- Survey of experiences with Web use for self diagnosis
- Self-report data from 500+ volunteers within Microsoft

- Web content increases anxiety (40% people), reduces anxiety (50% people)
  - Web can help, but can also cause distress, especially for those pre-disposed to anxiety
  - Key marginalizations (e.g., self-reported hypochondria) revealed larger effects

- Web helps patients understand conditions before and after diagnosis

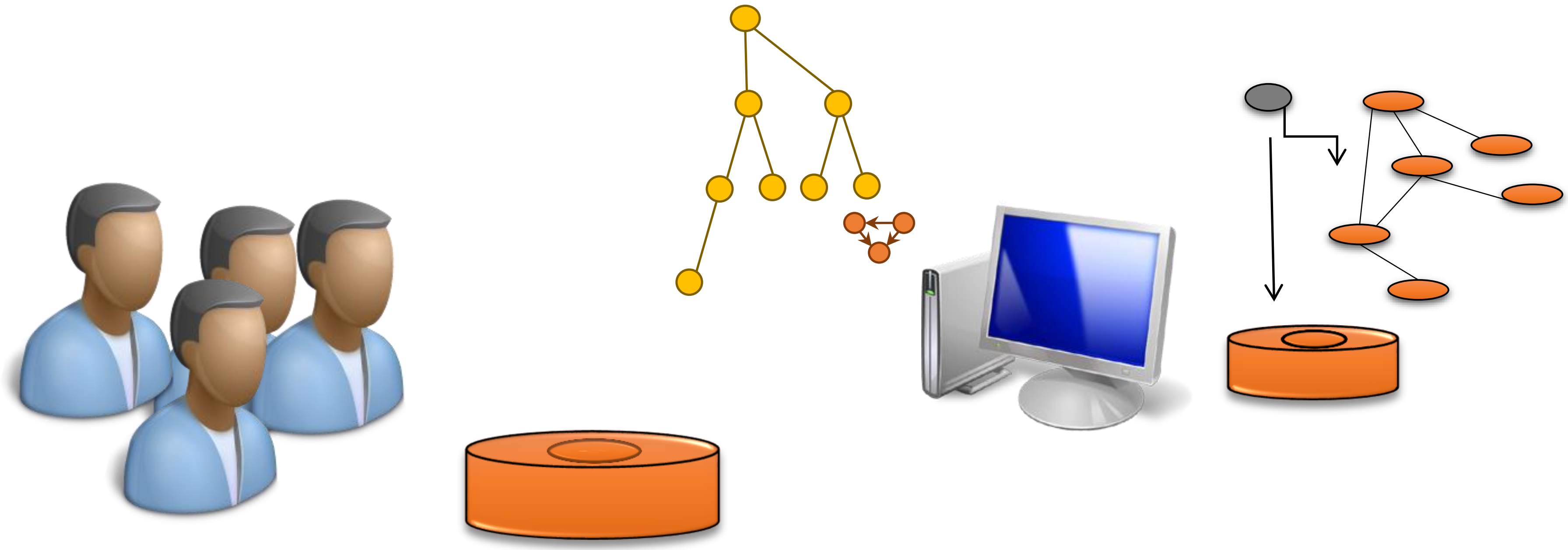
- **Escalation reported to occur *frequently* for 20% of respondents**

(White and Horvitz, AMIA 2009; TOIS 2009)

Beyond Self-Reports: Mining Health Search Activity

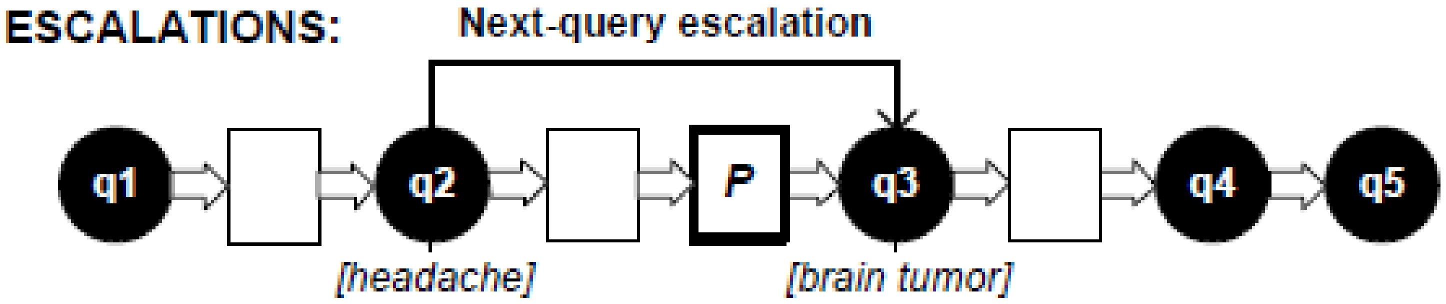
- Mining insights from large-scale logs
  - Query sequences & page accesses
  - Content distribution & dynamics
  - Insights, predictive models, services

Search engine log analysis shows:
→ Given symptom query, transition to serious condition occurs 2x as often as transition to benign
→ … even though, benign condition is often a lot more likely



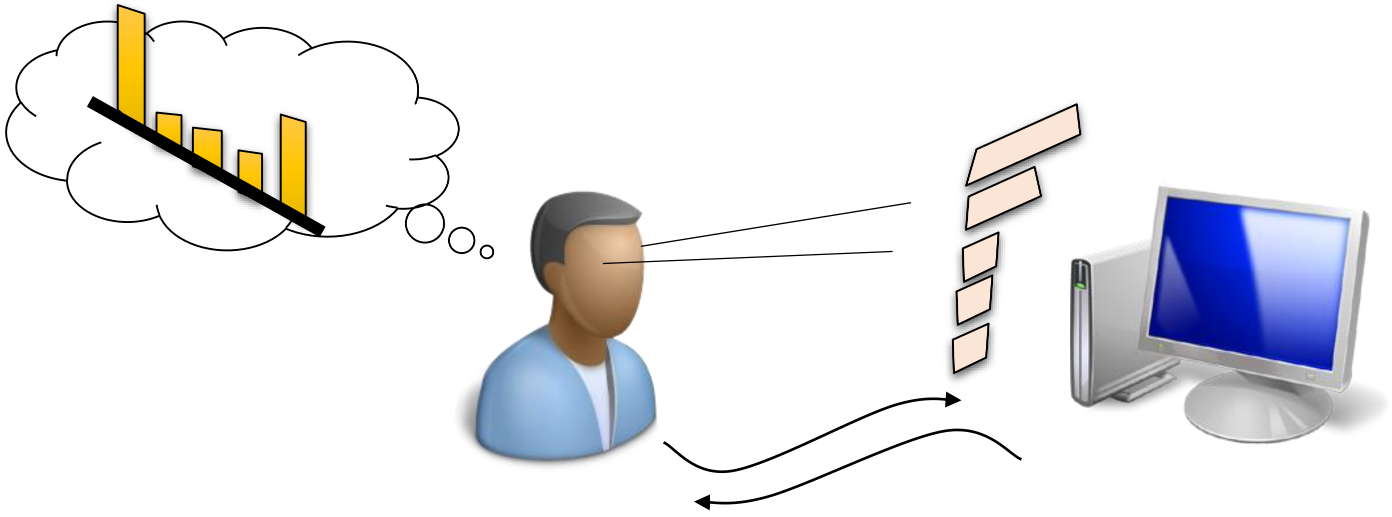
Predicting Escalations From Page Content

- Predict transition from common symptoms to rare, serious illnesses *based on features* of pages being viewed



Negatives = non-escalations
Query, click THEN no escalation, end session, etc.

Model accuracy = 73.4%
Baseline accuracy = 50%



	Order of Presentation on Page	
Query Outcome	Serious first	Benign first
Escalation	68.6%	33.4%
Non-escalation	31.4%	66.6%

(White and Horvitz, SIGIR 2010)

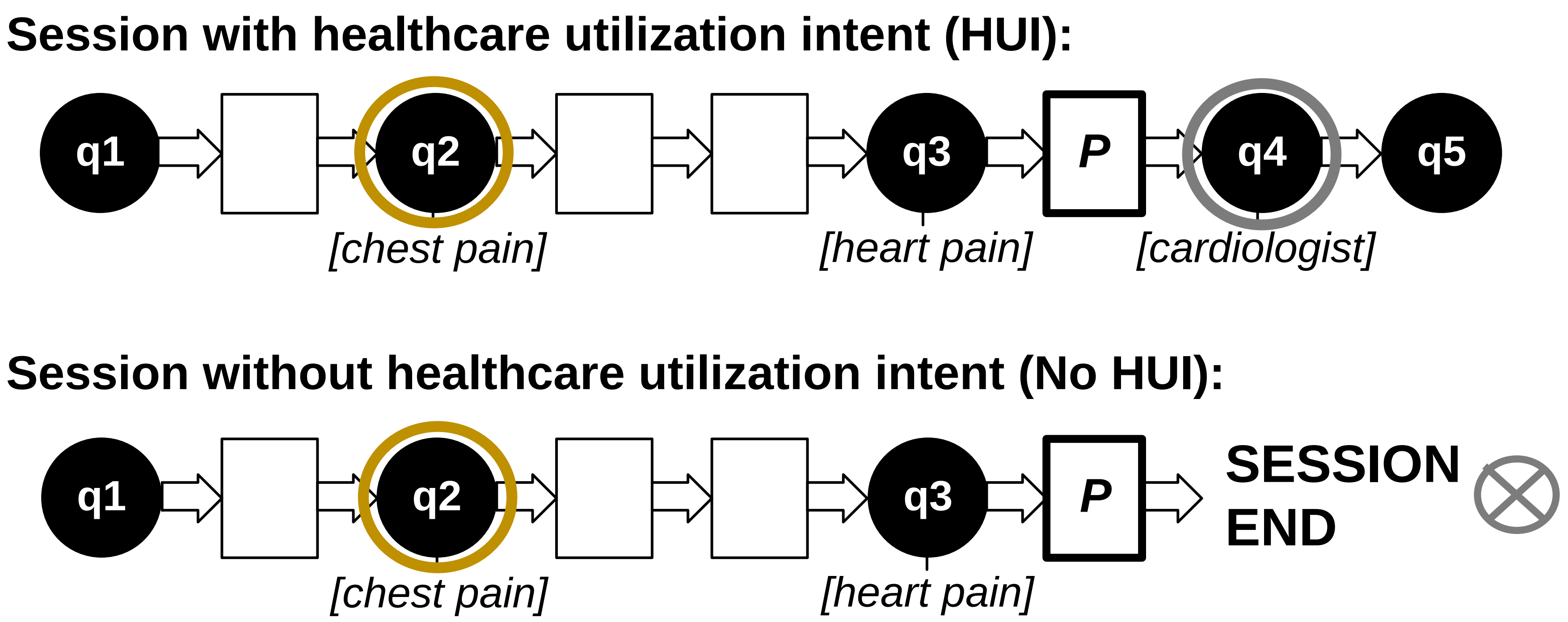
Web to World

- From our prior survey, **23.7% of respondents were put over the threshold to seek professional medical advice by Web content**
- Pursuit of in-world healthcare resources:
  - Healthcare Utilization Intention (HUI)
  - *E.g., [neurologist in seattle, wa], [evergreen hospital], [urgent care clinic]*
- Automated detection:
  - Appropriate medical specialty for the symptom (e.g., *neurologist* for symptom: muscle twitches)
  - Medical resource (e.g., *hospital*, *physician*)
  - Five-digit US zipcode, US city and state name pair (e.g., *Redmond, Washington*)

(White and Horvitz, AMIA 2010)

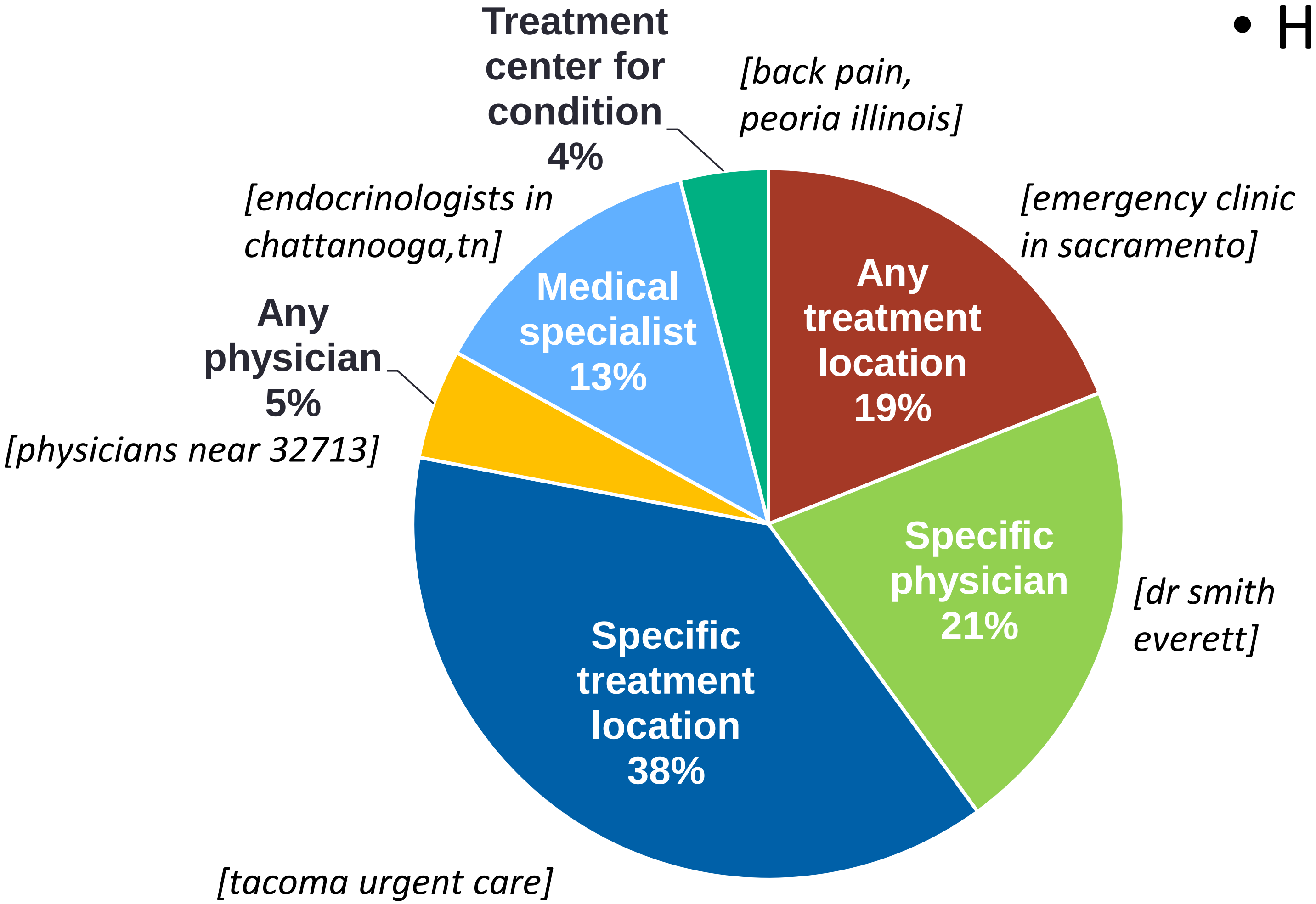
Studying Web to World

- Characterize and predict transitions to HUI in search logs



- Other methods for W2W identification, e.g.,
  - **Visitation** (via GPS tracking) (West et al., SIGIR 2013 – predicting geographic destinations)
  - **Call medical facility** (Mishra et al., SIGIR 2014 – time-critical search)

Characterizing Resource Pursuits



- HUI queries toward end of sessions
  - 36% of sessions, HUI was **last query**
  - Mean: HUIs occur at 75% of session

Predicting HUIs

- Prediction task

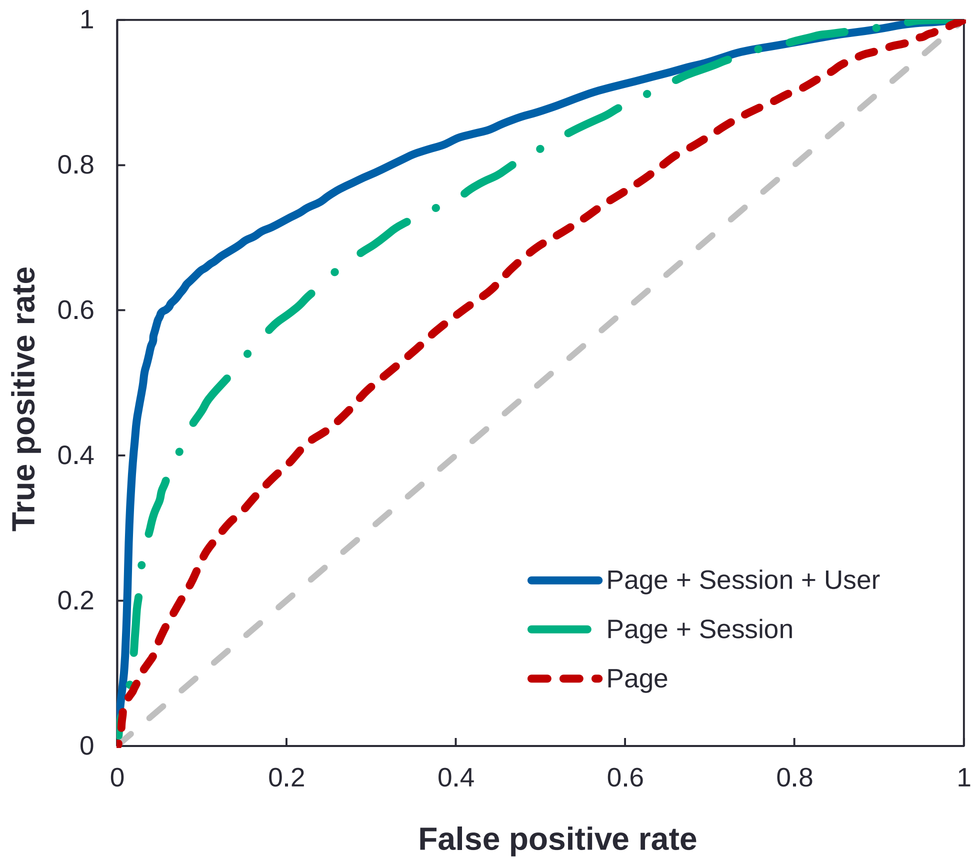
  Probability that user will next issue an initial HUI query given currently viewing page p.

- Three classes of features
  - **Page:** Structure & content of current page.
  - **Session:** Attributes of search interaction in current session.
  - **User:** Aspects of users' historic medical search interactions from the beginning of log data to start of current session.

Features	HUI	No HUI
SeriousBeforeBenign (Page)	59%	48%
IsWebForum (Page)	14%	9%
NumQueries (Session)	4.9	2.9
AvgQueryLength (Session)	4.5	4.1
NumUniqueSymptoms (User)	3.6	2.2
NumResourceQueries (User)	5.5	2.0

Logistic regression with five-fold CV

- Accuracy:
  - Page features = 59.3%
  - Page + session = 68.9%
  - **- Page + session + user = 77.7%**



Long-term Health Searching

- Logs can provide lens on 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**
  - 95% of these sessions had medical query in prior session(s)

Many searchers come to search engine with a condition in mind!

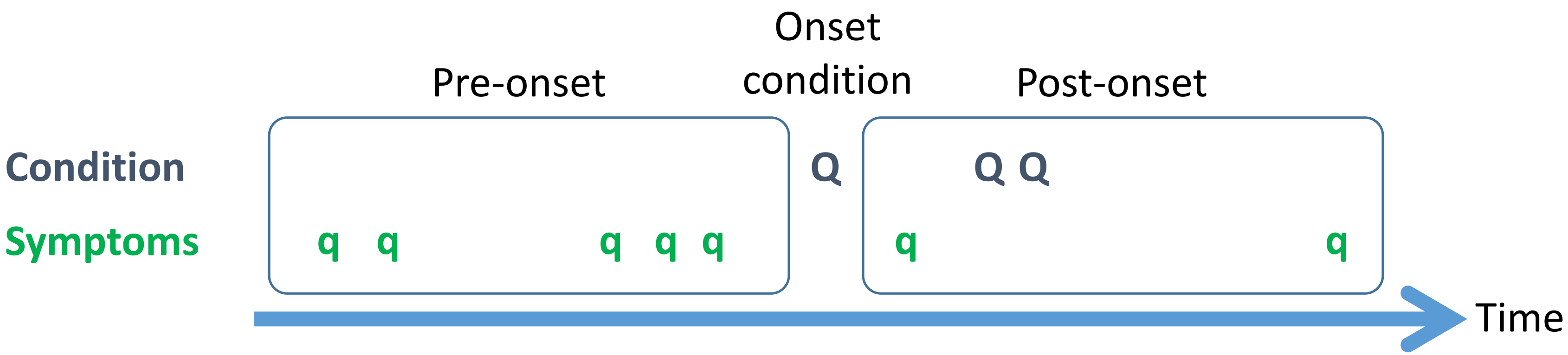
- Long-term helps us understand medical trajectories
  - Predict emerging concerns, personalize search, guide healthcare use

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

(White and Horvitz, SIGIR 2012)

Condition Onset

- Explore search behavior before first occurrence of a condition



- 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

- **79.5% of prior symptoms were related to onset**
  - **<u>Emergence of conditions extends back over time</u>**

Post-Onset Search Behavior

- Changes in search behavior **<u>after the first onset condition query</u>**:

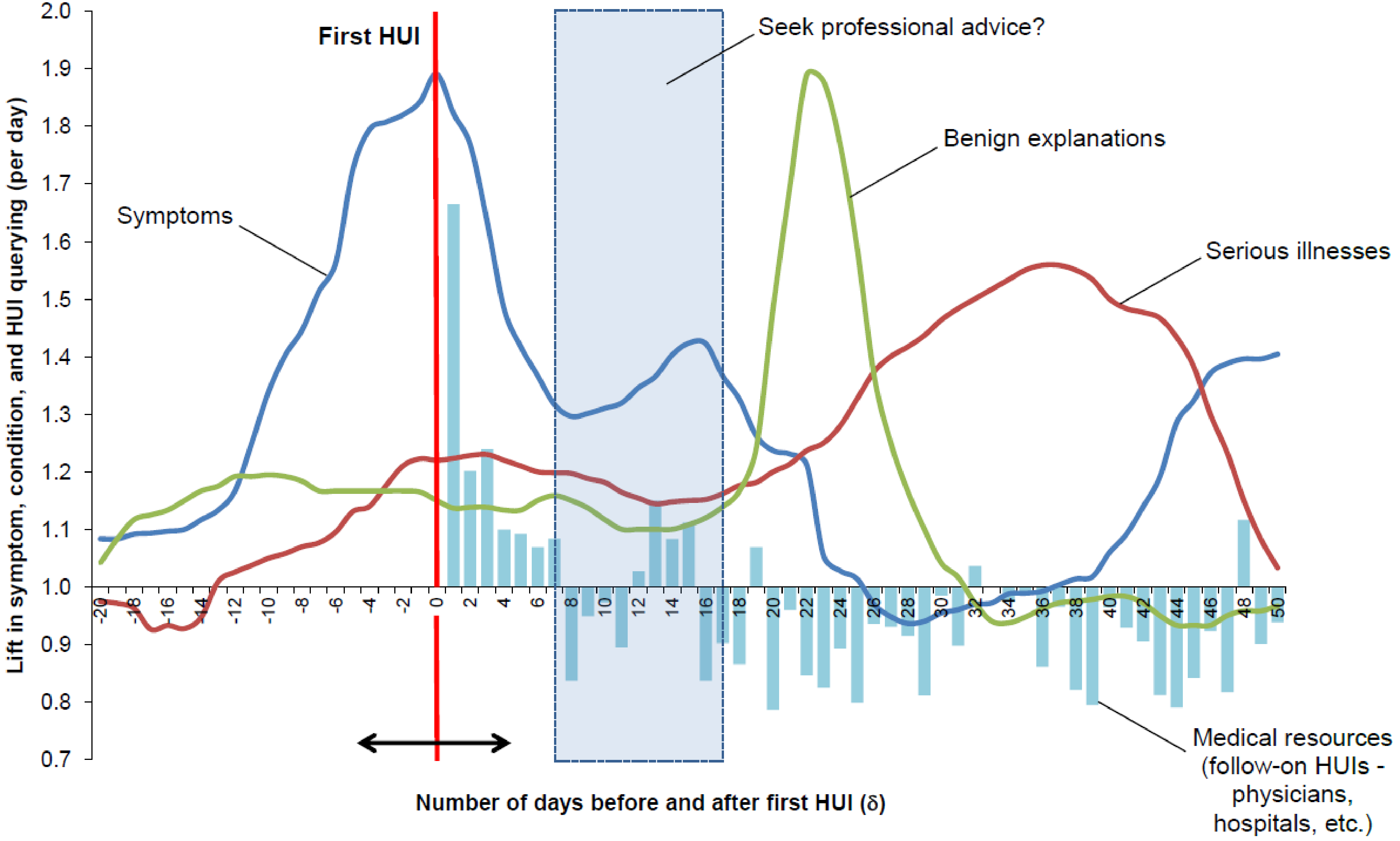
Feature	% or Avg (SD)	% change from pre-onset
% URLs medical	4.9%	+88.5 ↑
% queries medical	4.2%	+31.3 ↑
% online time on medical pages	8.0%	+247.8 ↑
# unique symptoms	0.50 (1.01)	−20.6 ↓
Symptom persistence (days)	2.46 (3.42)	−27.9 ↓
# unique conditions	1.04 (1.29)	+40.5 ↑
Condition persistence (days)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

- Medical search increases, symptom searching decreases, and condition searching increases
- **Interesting future work: Use combination of symptoms searched over time to <u>predict the onset condition</u> (early warning signs!)**

Tracking HUIs and Related Activity over Time

Align all users based on first HUI query (hospital, physician, specialist, etc.):

Lift in searching over expected search activity on each day



(White and Horvitz, JAMIA 2014)

Applications

- Quantifying skewed content distributions online
- Identifying (and down-weighting?) pages that are likely to cause escalations – challenge: the escalation may be justified
- Predicting onset of conditions over time → earlier interventions
- Better supporting Web to World transitions
    - Directing people to the right healthcare professionals, summarizing long-term search histories for sharing with the HCP

Part II : Biases in Behavior and Content

Bias in IR and elsewhere

In IR, e.g.,

- Domain bias – People prefer particular Web domains
- Rank bias – People favor high-ranked results
- Caption bias – People prefer captions with certain terms

In psychology, e.g.,

- Anchoring-and-adjustment, confirmation, availability, etc.

- **All impact user behavior**
  - Opportunity to intersect psychology and IR

Biases and Search Behavior

User behavior

Search engine behavior

- Bias can be observed in IR situations where searchers **<u>seek</u> or <u>are presented with</u> information that significantly deviates from a known or accepted truth.**

- Focus on set of Yes-No questions in Health Domain

(White, SIGIR 2013)

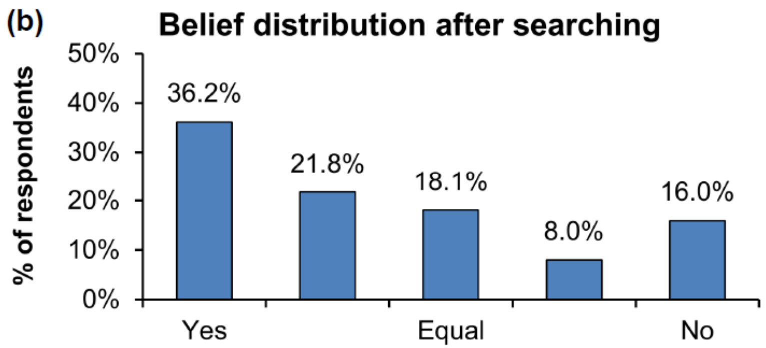
Initial Exploratory Questionnaire

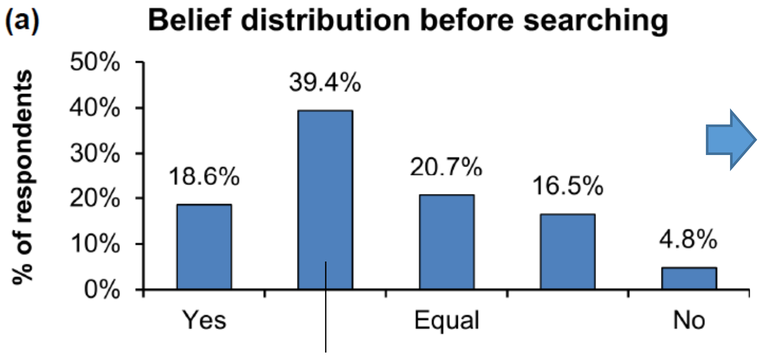
- Gain early insight into possible biases in search
- **Focus on Yes-No questions (answered with "Yes" or "No")**
  - Simplicity: Answers along single dimension (Yes → No)

- Microsoft employees; recall recent Yes-No query (in last 2 weeks)
- Asked about belief beforehand and afterwards
  - Multi-point scale: **Yes** / Lean Yes / **Equal** / Lean No / **No**

- 200 respondents. Recalled questions such as:
  - "*Does chocolate contain caffeine?*"
    "*Are shingles contagious?*"

Survey Results





Confirmation?

Belief before search	*Belief after search*: **Yes**		**Equal**		**No**
Yes	77.1%	8.6%	8.6%	2.9%	2.9%
	35.1%	40.5%	9.5%	2.7%	12.2%
Equal	23.1%	12.8%	43.6%	15.4%	5.1%
	16.1%	9.7%	19.4%	19.4%	35.5%
No	11.1%	–	11.1%	–	77.8%

Post-search belief given Pre-search belief

- Two main findings:
  1. Respondents kept strongly-held beliefs (Yes-Yes and No-No)
  2. If Before = Equal, then 2x as likely to believe Yes after search

Motivated us to:
Further explore possible impact of biases on behavior and outcomes

Log-Based Study of Yes-No Queries

- Queries, clicks, and results from Bing logs (2 weeks)
- Mined yes-no questions: start with “can”, “is”, “does”, etc.
- Focused on health since it’s important and we could get truth

- Randomly selected set of 1000 yes-no health questions
  - Each issued by at least 10 users, same top 10, same captions

- Examples include:
  - “*Is congestive heart failure a heart attack?*” (answer = No)
“*Do food allergies make you tired?*” (answer = Yes)

Answer Labeling

Yes-No Answer labels for captions/results

Physician answers for the Yes-No questions

- **Captions and result content**
- Crowdsourced (Clickworker.com)
- 3-5 judges/caption (consensus)
- Task was to assign label of:
  - **Yes** only
  - **No** only
  - **Both** (Yes and No)
  - **Neither** (not Yes and not No)
- Agreement on 96% of captions
- Performed similar labeling for each top 10 search results
  - Crowdsourced judges, agreement on 92% of pages

Example Caption Labels

Suggests AFFIRMATIVE answer (Yes only):
Question: [can i take l carnitine while pregnant]

Yes only

Is it safe to **take L-Carnitine while pregnant** - The Q&A wiki
http://wiki.answers.com/Q/Is_it_safe_to_take_L-Carnitine_while_pregnant
Is **l-carnitine** safe to **take while pregnant**? yes. Is it safe to **take** zithromax **while pregnant**? yes it is safe to **take while pregnant**. A doctor would not prescribe it ...

Suggests NEGATIVE answer (No only):
Question: [does robaxin show up on drug tests]

No only

Does robaxin show up on drug tests? | Answerbag
http://www.answerbag.com/q_view/1239474
Does robaxin show up on drug tests? no ... More Questions. Additional questions in this category. Can you have a DUI & work at a school in Pennsylvania?

Suggests BOTH affirmative and negative:
Question: [is tooth a bone]

Both

Is **tooth** consider as **a bone** - The Q&A wiki
http://wiki.answers.com/Q/Is_tooth_consider_as_a_bone
What does the **bone** in the **tooth** do? It helps u chew. **Is a tooth a bone**? Yes. Is your **tooth a bone**? No, **teeth** are not **bones**. Is the "skin" lining your stomach skin?

Suggests NEITHER affirmative nor negative:
Question: [does crestor cause bloating]

Neither

Does Crestor Cause Bloating? – HealthCentral
http://www.healthcentral.com/cholesterol/h/does-crestor-cause-bloating.html
Everything you need to know about **does crestor cause bloating**, including common uses, side effects, interactions and risks.

Physician Answers

- Two physicians reviewed the 1000 questions and gave answers
  - Inc. **50/50** = need more info, **Don't know** = really unsure
- Agreement between physicians on Yes-No was 84% (κ=0.668)

		Physician 2				
		Yes	***No***	***50/50***	***Don't know***	***Total***
Physician 1	***Yes***	**38.8%**	8.2%	3.7%	0.5%	51.2%
	No	5.7%	**31.5%**	1.2%	0.2%	38.5%
	50/50	1.8%	2.0%	1.3%	0.0%	5.0%
	Don't know	1.3%	3.1%	0.2%	0.7%	5.3%
	Total	47.5%	44.8%	6.3%	1.5%	100.0%

- Focused on the **680 questions** where both agreed Yes or No
- Distribution: **55% Yes and 45% No** (used as **TRUTH** in our study)

Result Ranking

- Volume of Yes-No content in the results

Percentage of captions or results with answer

Source	*Yes only*	*No only*	*Both*	*Neither*
Caption	28.7%	8.4%	2.7%	60.2%
Result	35.0%	12.7%	6.3%	41.0%

→ More Yes content in top-10 than No content

- Relative ranking of **top** Yes-No content when both in top 10

Percentage of SERPs where top *yes* caption or result appears above (nearer the top of the ranking than) the top *no*

Source	*Yes above No*	*No above Yes*
Caption	65.1%	34.9%
Result	62.4%	37.6%

→ Yes content ranked above No more often (when both shown)

User Behavior (Clickthrough rate)

- Studied **clickthrough rates** on captions containing answers
- Controlled for rank by just considering top result (r=1)

SERP click likelihoods for different captions given variations in answer presence in SERPs/captions, and rank

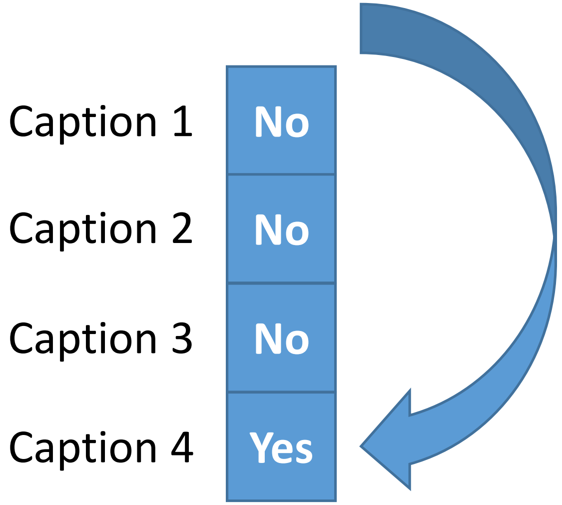
Condition(s)	*All*
$SERP_Y$	80.0%
$SERP_N$	75.9%
$SERP_{BOTH}, Caption_Y$	45.6%
$SERP_{BOTH}, Caption_N$	14.2%
$Caption_Y$	41.1%
$Caption_N$	16.3%
$Caption_{Y,r=1}$	47.4%
$Caption_{N,r=1}$	12.6%

Just considering top search result

3-4x as likely to click on captions with Yes content, even though TRUTH = 55% Yes / 45% No

User Behavior (Result skipping)

- Studied result **skipping** behavior
- Frequency with which people skipped caption w/answer to click other caption



Distribution of clicks and skips by answer

Click	*Skip* *Yes only*	*Skip* *No only*
Yes only	33.3%	$\boxed{41.5\%}$
No only	$\boxed{8.5\%}$	16.7%

- Users more likely (4x) to skip No to click Yes than vice versa

Answer Accuracy

- Examined accuracy of the top search result, as well as first click and last click in session

Answer defn.	*All*	*Physician Answer*	
		Yes	*No*
Top result	45.0%	57.1%	22.9%
First click	50.0%	59.1%	27.9%
Last click	52.3%	66.2%	29.4%

- Findings show:
  1. Top result accurate only 45% of time, less when truth is No
  2. Users improve accuracy, but only slightly (limited by top 10)

- Potential cause for low accuracy → bias in retrieved content

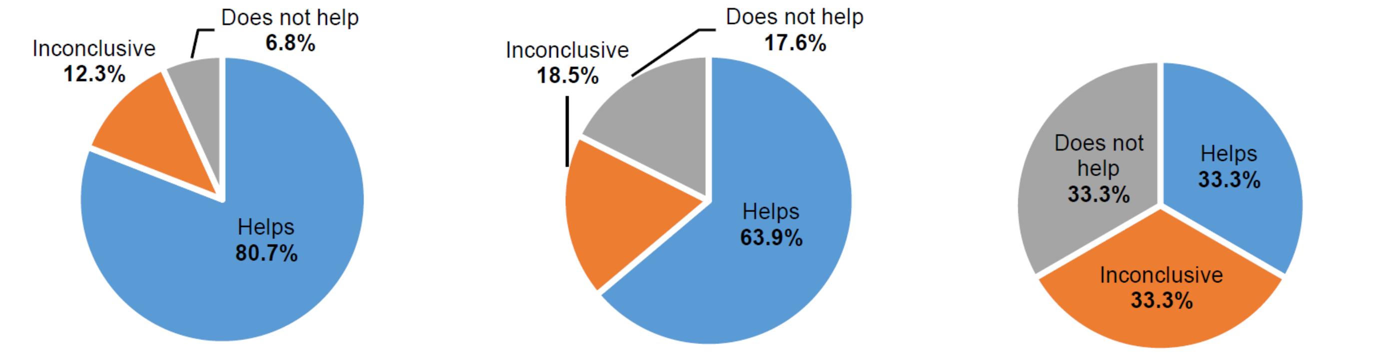
Content Biases

- Content bias in search results describes a deviation from a known or accepted truth that negatively affects result accuracy

- Used Cochrane reviews as ground truth (cochrane.org)
  - “Systematic reviews of the effects of health care” (interventions)

THE COCHRANE COLLABORATION

- Selected 3 outcomes: **Helps**, **Does not help**, **Inconclusive** (1/3 each)
- Matched queries, Hand-labeled content in top 10 and search index (top 1000)



(a) Top 10 search results (b) Search engine index (c) Expected distribution

Figure 1. Distribution of answers about interventions in (a) top results, (b) matching index content, and (c) the expected (true) distribution given our sampling criteria (33% per answer).

Example queries:
Does green tea **help** with weight loss?
Can cranberries **cure** UTIs?
Can echinacea **treat** the common cold?

Contribute to positive skew in search engine results

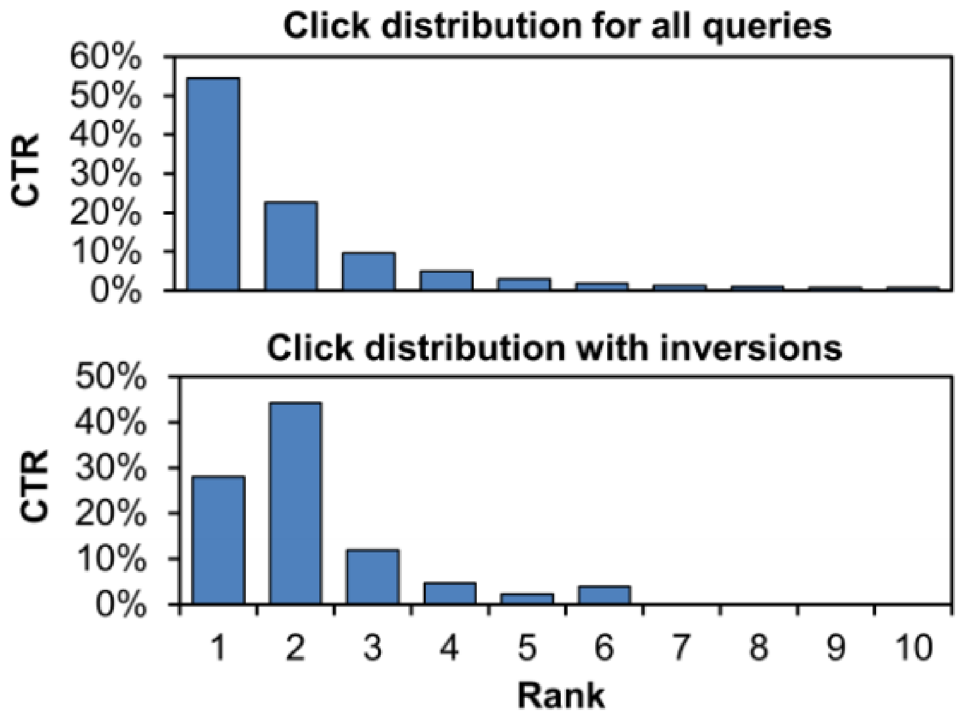
(White and Hassan, TWEB 2014)

Impact on Search Behavior

- Potentially-alarming content in captions drives clickthrough behavior, leading to changes in CTR distributions including click inversions

Click inversions (Clarke, Agichtein, Dumais, White, 2007)

Fig. 1. Top three search result captions for [chest pain]. Potentially-alarming caption content is highlighted.



Lauckner and Hsieh (CHI 2013) serious conditions in captions

→ Negative emotional outcomes for users

([White and Horvitz, TWEB 2013](#))

Building Click Prediction Models

- Features associated with clickthrough inversion caption pairs
- Learn models to predict clickthrough

Table V. Results corresponding to the features listed in Table IV with χ^2 and p-values (df = 1). Features related to inversions and supported at 95% confidence level are bold. In rows with any cell count < 5 we use a Fisher's exact test.

Category	Feature Tag	INV+	INV–	%+	CON+	CON–	%+	Diff	χ^2	p-value
Course	**Acute**	38	13	74.51	23	45	33.82	+40.69	19.309	< .0001
	Chronic	48	54	47.06	61	43	58.65	–11.59	2.7787	0.0955
Degree	**Severe**	105	65	61.76	71	99	41.76	+20.00	13.6170	0.0002
	Mild	13	52	20.00	14	7	66.67	–46.67	16.0483	< .0001
Tendency	**Malignant**	72	33	68.57	45					
	Benign	29	29	50.00	53					
Prognosis	**Deadly**	22	6	78.57	12					
	Nonfatal	4	5	44.44	7					
Transition	**Escalations**	111	54	67.27	42					
	NonEscalations	90	70	56.25	118					
Condition	**AnySeriousCondition**	274	189	59.18	236					
	AnyBenignCondition	329	302	52.14	310					
	Cancer	31	19	62.00	16					
	Pregnancy	28	22	56.00	27					
Healthcare utilization	MedicalFacility	101	105	49.03	131					
	MedicalSpecialist	6	5	54.55	13					
	MedicalProfessional	115	145	44.23	153					
Source	**MayoClinic**	75	66	53.19	90					
	WebMD	81	30	72.97	47					
	MedlinePlus	32	60	34.78	69					
	PubMed	3	10	23.08	12					
Snippet	MissingSnippet	14	20	41.18	3					
	SnippetShort	6	2	75.00	13					
Term match	TermMatchTitle	7	3	70.00	12					
	TermMatchTS	131	127	50.78	192					
	TermMatchTSU	82	94	46.59	112					
	TitleStartQuery	446	348	56.17	450	414	52.08	+4.09	2.7840	0.0952
	QueryPhraseMatch	213	154	58.04	233	233	50.00	+8.04	5.3329	0.0209
URL	**URLQuery**	16	11	59.26	13	26	33.33	+25.93	4.3535	0.0369
	URLSlashes	833	644	56.4	718	861	45.47	+10.93	36.4513	< .0001
	URLLenDiff	1471	753	66.14	1166	1218	48.91	+17.23	139.5928	< .0001
Readability	Readable	22	30	42.31	22	24	47.83	–5.52	0.3004	0.5836

Presence of following is likely to cause inversions:

- ***Acute***
- ***Severe***
- ***Malignant***
- ***Deadly***
- ***Escalations***
- ***Cancer***
- ***…***

Click perplexity, **lower = better**

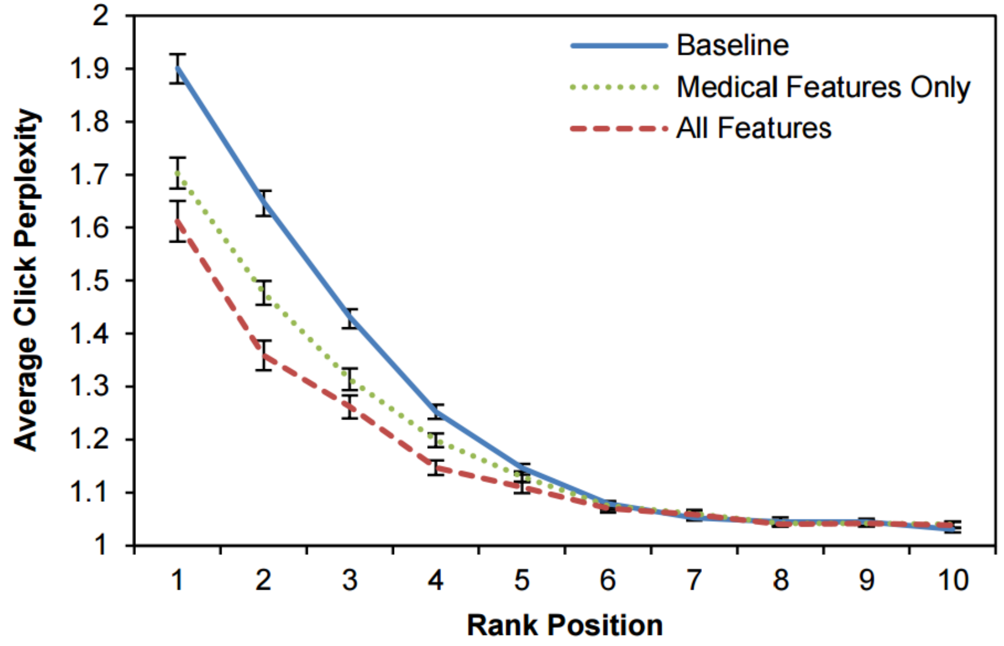


Fig. 4. Perplexity curves for DBN-model variants.
Lower perplexity represents better prediction. Error bars denote standard error.

Part III: Mining Health Search Data

3 example applications:

1. Nutrition monitoring in populations
2. Pregnancy
3. Detecting adverse drug reactions and interactions

Example 1:
Nutrition Monitoring in Populations

(Robert West, White, Horvitz, WWW 2013)
"From Cookies to Cooks: Insights on Dietary Patterns via Analysis of Web Usage Logs"

Nutrition is a Major Health Factor

- For example, annual cost of morbidity and mortality of obesity in United States and Canada: $300 billion
- Who eats what, when, and where?
- Answer usually obtained via phone surveys, medical records, diary studies, etc.
- Explored the use of **logs** and **online recipe accesses** for population-scale nutrition monitoring

(West, White, Horvitz, WWW2013)

Log Analysis – Recipe Users

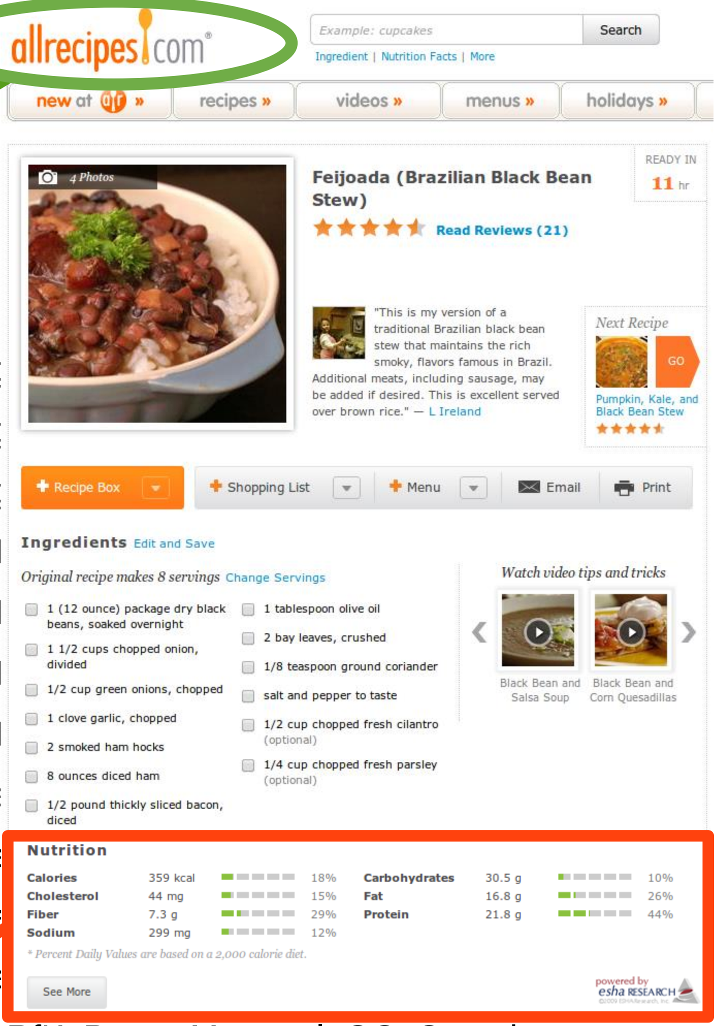
Consenting IE users, 18 months

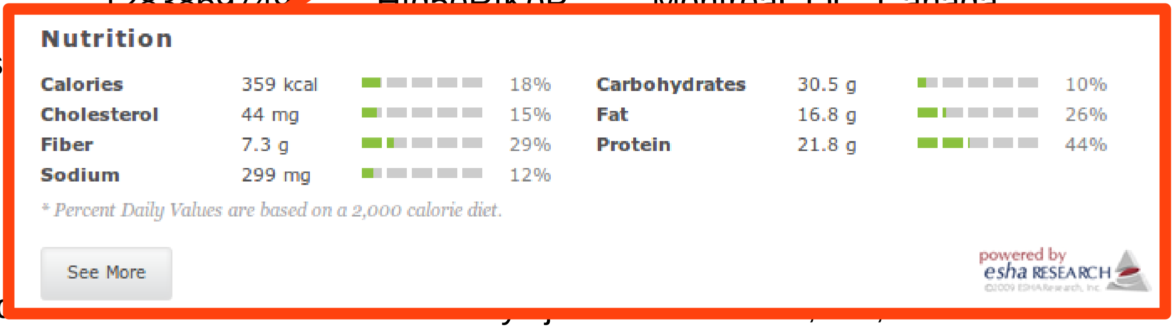
URL	Referrer	Timestamp	Anonym. UID	Geolocation
• yahoo.com?q=the+onion	yahoo.com	1283769636	h85TgdWhfg	Hackensack, NJ, USA
• theonion.com	yahoo.com?q=the+onion	1283769640	h85TgdWhfg	Hackensack, NJ, USA
• theonion.com/Area-Man-Sad	theonion.com	1283769644	h85TgdWhfg	Hackensack, NJ, USA
• bing.com		1283883335	A156N6yOjV	Blumenau, SC, Brazil
• bing.com?q=feijoada+recipe	bing.com	1283883340	A156N6yOjV	Blumenau, SC, Brazil
• allrecipes.com/tasty-feijoada	bing.com?q=feijoada+recipe	1283883346	A156N6yOjV	Blumenau, SC, Brazil
• food.com/best-feijoada-recipe	bing.com?q=feijoada+recipe	1283883397	A156N6yOjV	Blumenau, SC, Brazil
• arxiv.org		1283869645	Hfd5eRfKoP	Montreal, QC, Canada
• arxiv.org/recently-added	archiv.org	1283869649	Hfd5eRfKoP	Montreal, QC, Canada
• arxiv.org/article/cs.832590	arxiv.org/recently-added	1283869656	Hfd5eRfKoP	Montreal, QC, Canada
• google.com		1283869746	Hfd5eRfKoP	Montreal, QC, Canada
• google.com?q=cute+students	google.com	1283869749	Hfd5eRfKoP	Montreal, QC, Canada
• i.stanford.edu/~west1	google.com?q=cute+students	1283869751	Hfd5eRfKoP	Montreal, QC, Canada
• bing.com		1283877450	A156N6yOjV	Blumenau, SC, Brazil
• bing.com?q=banana+bread	bing.com	1283877458	A156N6yOjV	Blumenau, SC, Brazil
• epicurious.com/Banana-Bread	bing.com?q=banana+bread	1283877464	A156N6yOjV	Blumenau, SC, Brazil
• epicurious.com/Banana-Split	epicurious.com/Banana-Bread	1283877501	A156N6yOjV	Blumenau, SC, Brazil

Log Analysis – Recipe Users

Consenting IE users, 18 months

URL	Referrer	Timestamp	Anon
yahoo.com?q=the+onion	yahoo.com	1283769636	h85Tg
theonion.com	yahoo.com?q=the+onion	1283769640	h85Tg
theonion.com/Area-Man-Sad	theonion.com	1283769644	h85Tg
bing.com		1283883335	A156I
bing.com?q=feijoada+recipe	bing.com	1283883340	A156I
allrecipes.com/tasty-feijoada	bing.com?q=feijoada+recipe	1283883346	A156I
food.com/best-feijoada-recipe	bing.com?q=feijoada+recipe	1283883397	A156I
arxiv.org		1283869645	Hfd5e
arxiv.org/recently-added	archiv.org	1283869649	Hfd5e
arxiv.org/article/cs.832590	arxiv.org/recently-added	1283869656	Hfd5e
google.com		1283869746	Hfd5e
google.com?q=cute+students	google.com	1283869749	Hfd5eRfKeR Montreal, QC, Canada
i.stanford.edu/~west1	google.com?q=cute+students		
bing.com			
bing.com?q=banana+bread	bing.com		
epicurious.com/Banana-Bread	bing.com?q=banana+bread		
epicurious.com/Banana-Split	epicurious.com/Banana-Bread		

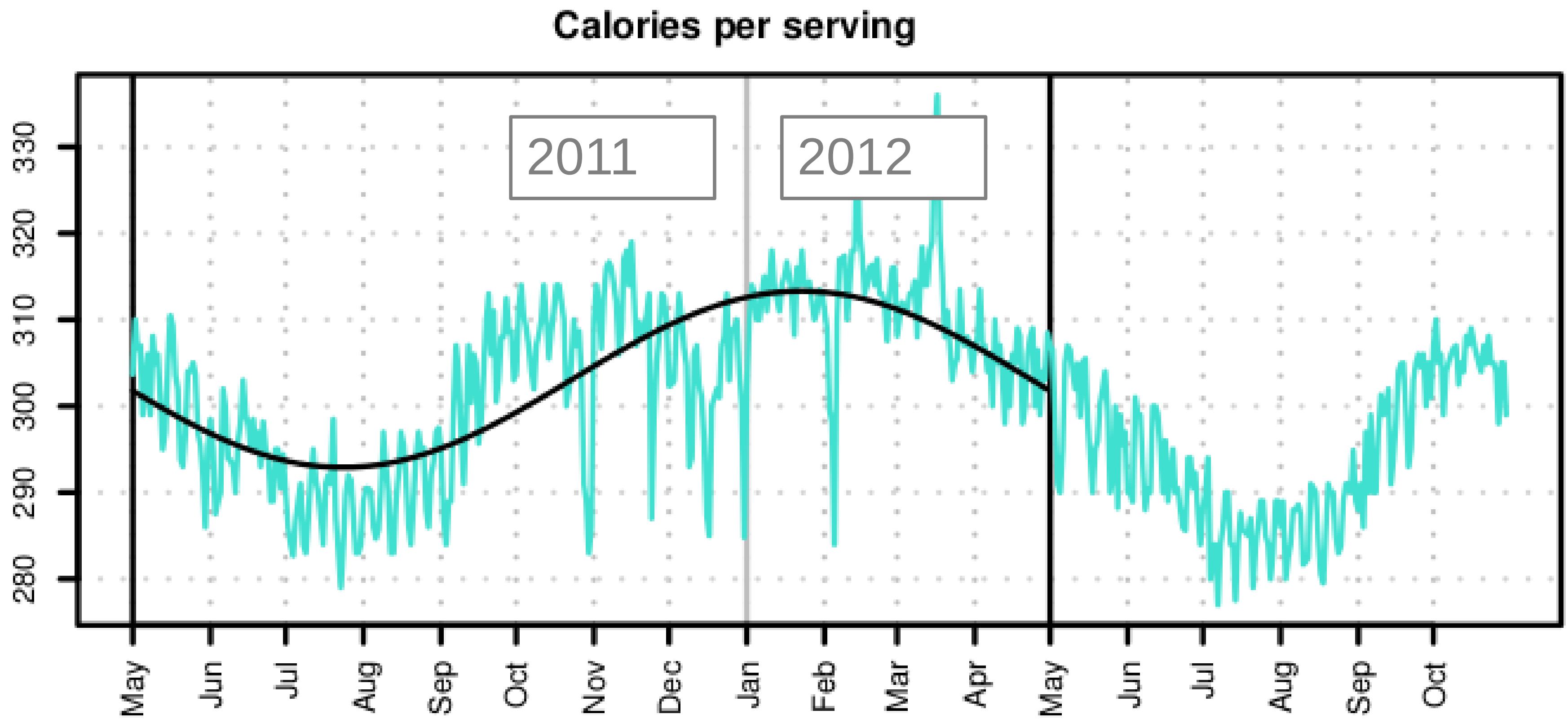




Nutritional time series

6 nutrients: total kcal, protein, fat, sodium, cholesterol

For each nutrient: average by day of year



Online Recipes for Approximating Food Popularity

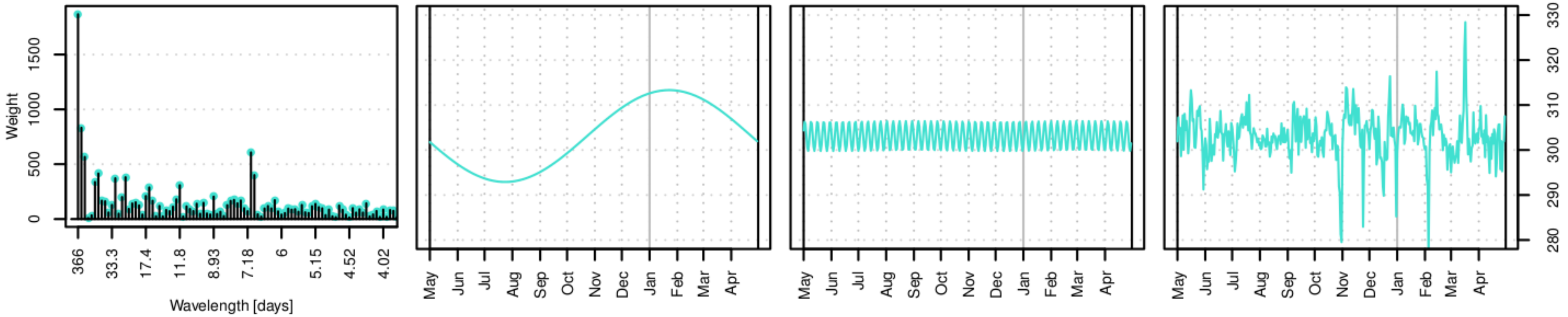
- **Fundamental assumption:** Searched online recipes ≈ eaten food
- Reality check: user survey among Microsoft employees
- “Recall last time you used an online recipe.”
  - “Did it represent well what you normally eat? **75% yes**
  - “Did you search for the specific dish you ended up cooking?” **81% yes**

Other factors:

- Population bias: online recipe users may not be representative of population
- False positives: “look but don't cook”
- False negatives: “cook but don't look”

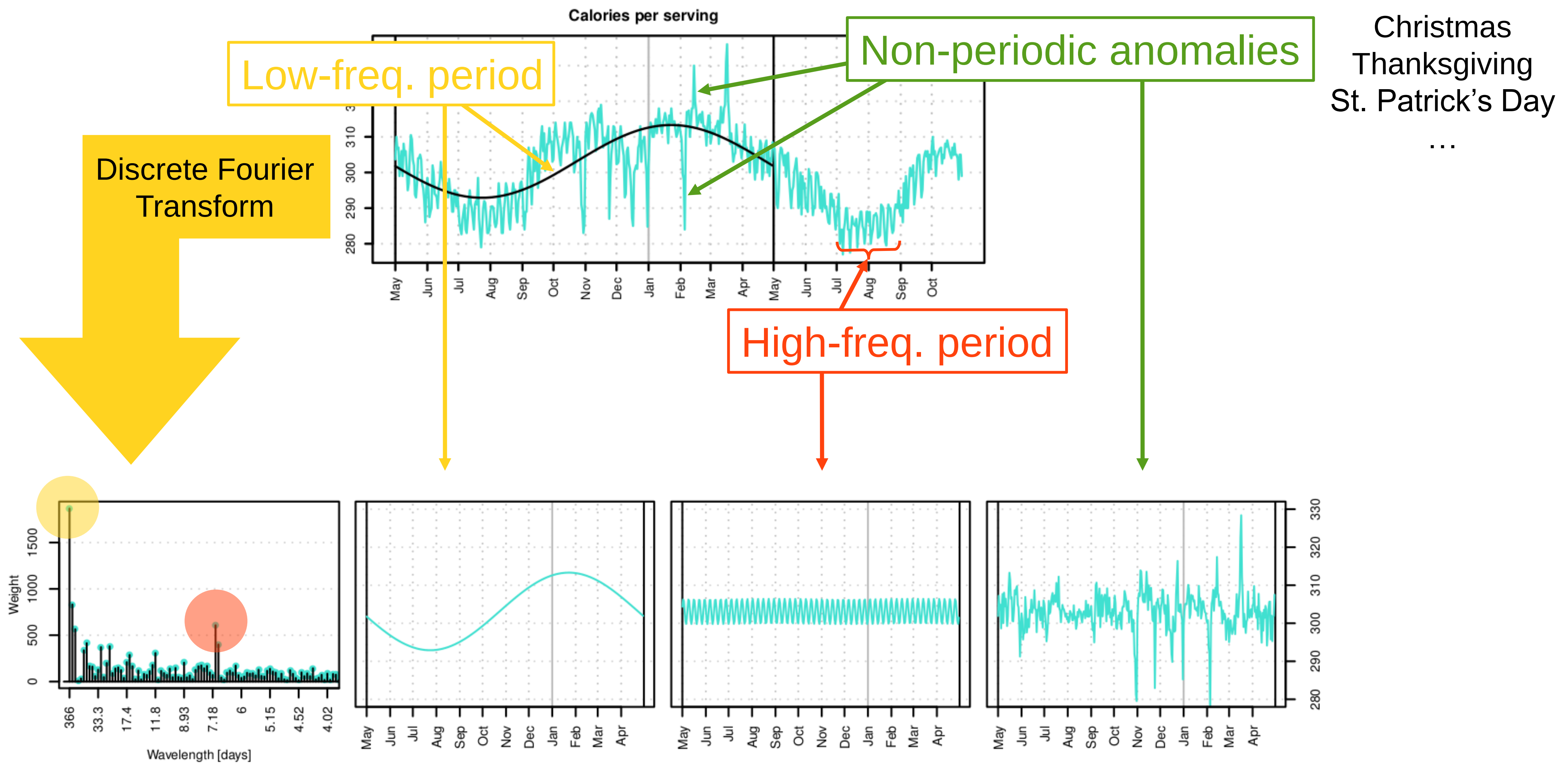
Anatomy of Nutritional Time Series

Discrete Fourier Transform



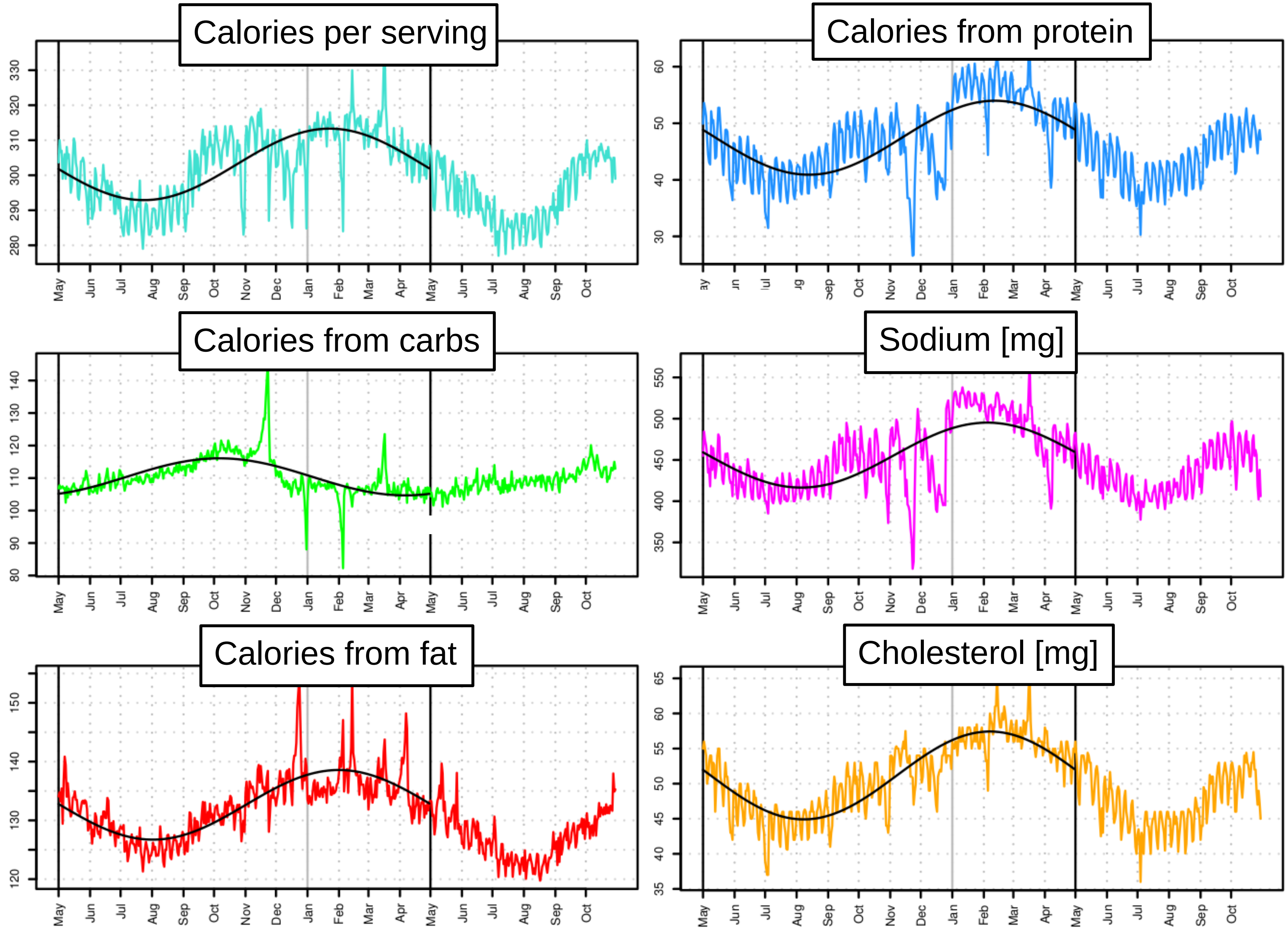
(a) Spectral density (b) Annual frequency (c) Weekly frequency (d) Residual

Anatomy of Nutritional Time Series

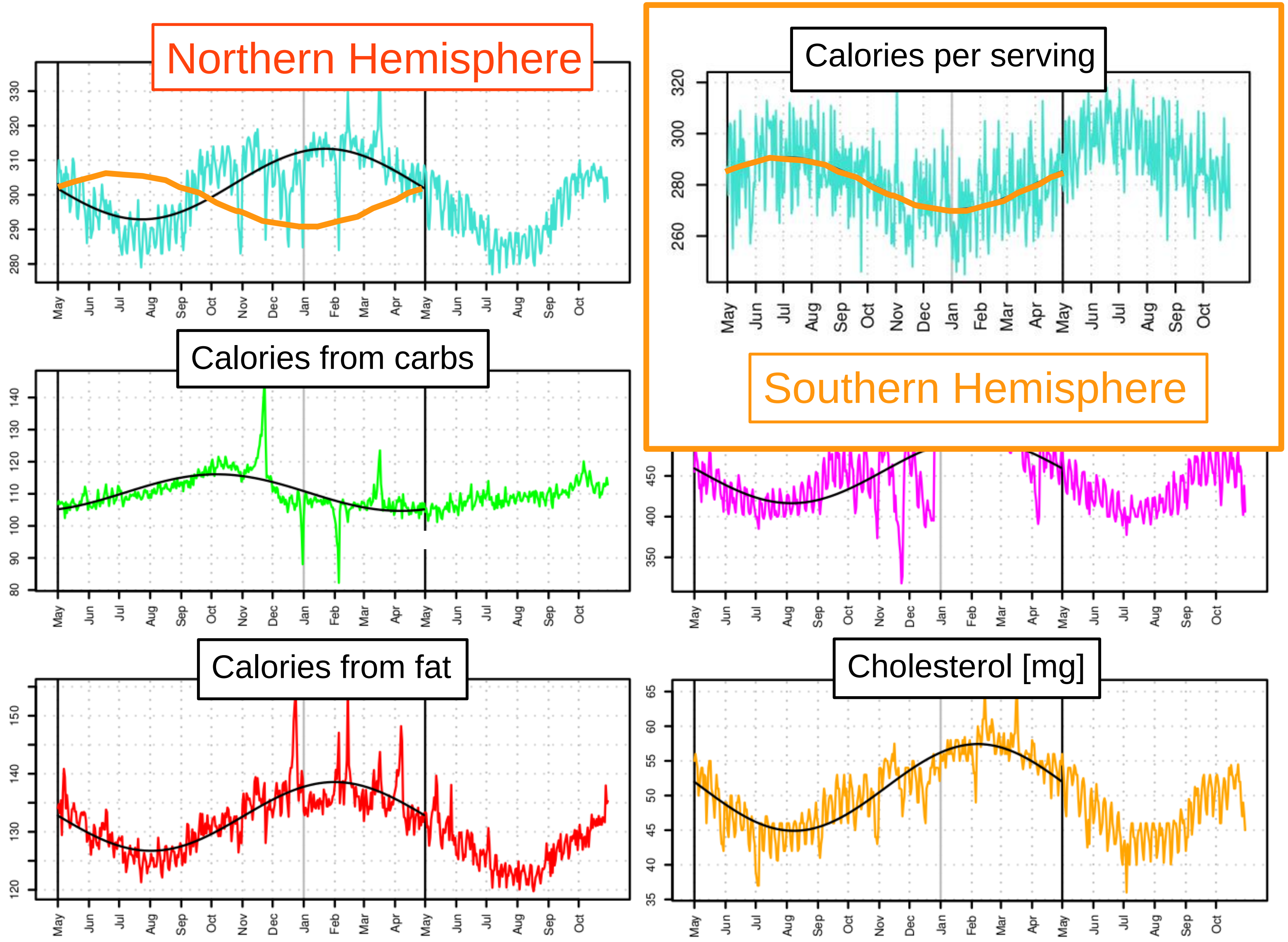


(a) Spectral density (b) Annual frequency (c) Weekly frequency (d) Residual

Nutritional time series: Annual variation



Nutritional time series: Annual variation



Effects are seasonal!

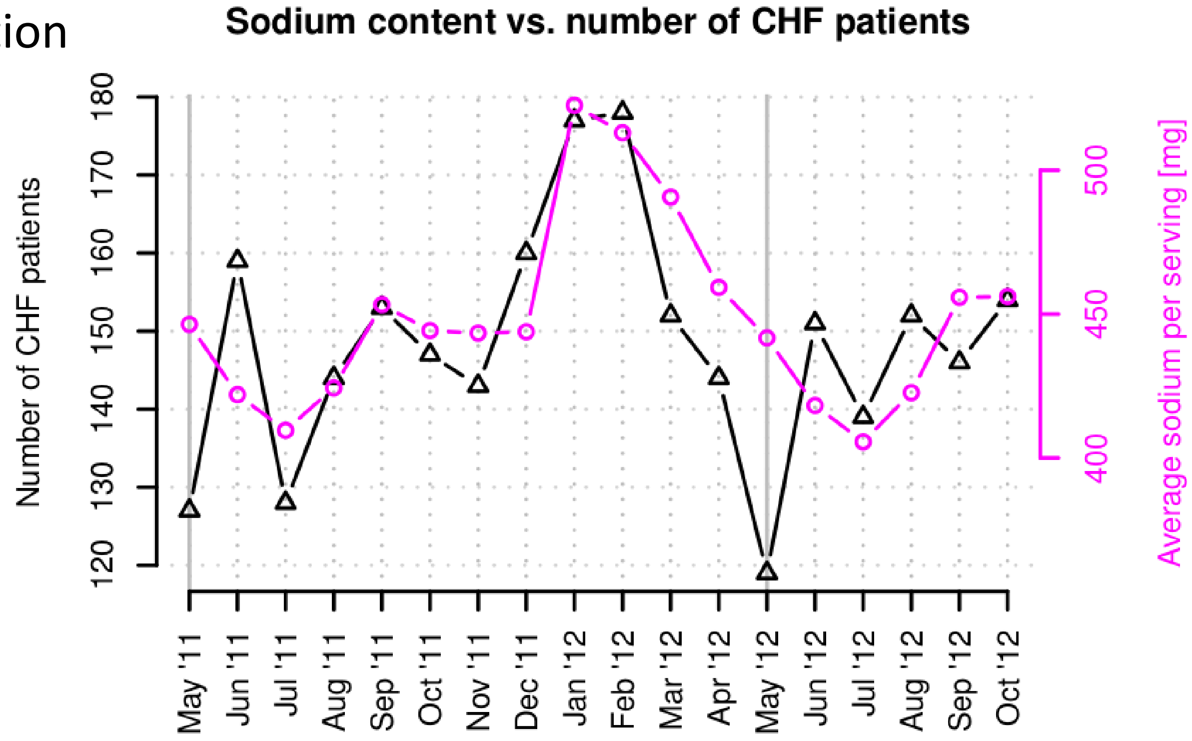
Correlations with Population Health

- Congestive heart failure (CHF): chronic condition that is dangerous and costly
- Increased sodium intake can cause acute exacerbation of symptoms
- Anecdotal reports by health practitioners:
  - Salty holiday meals with family → higher rates of CHF exacerbation
- Idea:
  - Approximate sodium intake via **recipe clicks**
  - Correlate with **hospital admission records**
- Data: All CHF admissions to emergency department for time period of our logs (Washington Hospital Center, Washington, D.C.)

Correlations with Population Health

- Congestive heart failure (CHF): chronic condition that is dangerous and costly
- Increased sodium intake can cause acute exacerbation of symptoms
- Anecdotal reports by health practitioners:
  - Salty holiday meals with family → higher rates of CHF exacerbation
- Idea:
  - Approximate sodium intake via **recipe clicks**
  - Correlate with **hospital admission records**
- Data: All CHF admissions to emergency department for time period of our logs (Washington Hospital Center, Washington, D.C.)

Pearson correlation: $r = 0.62$, $p = 0.0028$



Applications

Estimating nutritional intake from logs enables:

- Insights about public health via online activities
- Cheap and fast tracking of population-wide dietary preferences
- Guide targeted public-health campaigns
- Understand and intervene on chronic conditions
- Support of users interested in changing eating habits

Example 2:
Exploring Time-Dependent Concerns about Pregnancy and Childbirth from Search Logs

(Adam Fourney, White, Horvitz, SIGCHI 2015)

Pregnancy

- Last year, 3,957,577 babies were born to parents in the United States

- Many of those parents searched online for information about their pregnancy or their newborns

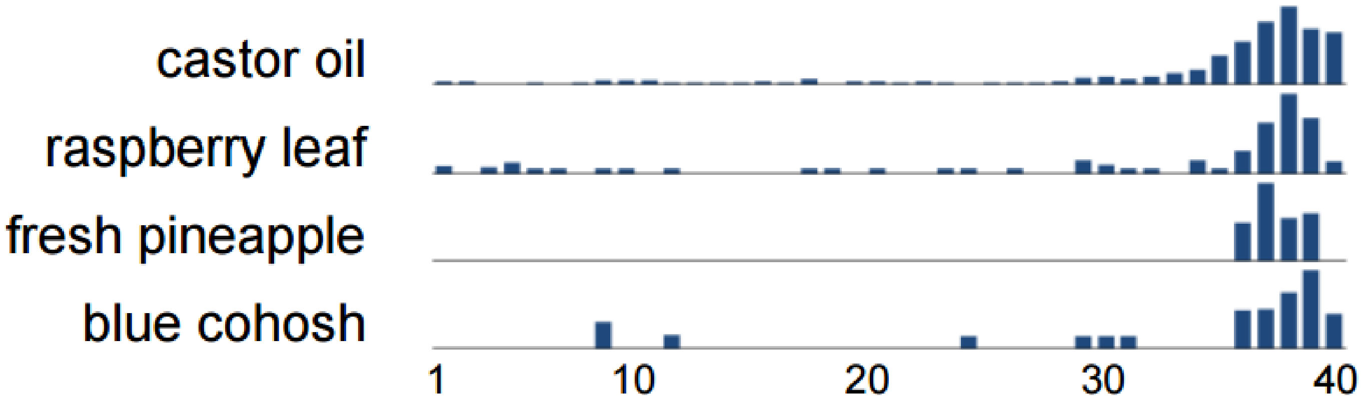
- Their queries tell, in exquisite detail, the very human and personal story of pregnancy and childbirth

- Advertisers know this
  - A person's attention becomes **220 times** more valuable to advertisers if it is known they are pregnant*

*E Steel. (2013), *Financial worth of data comes in at under a penny a piece.* Financial Times.

Pregnancy

Can this data also be of value as a tool for public health research?
e.g., studying querying for self-induction of labor over 40 weeks:



Interest in self-induction of labor peaks at week 38. Induction not recommended before week 39.

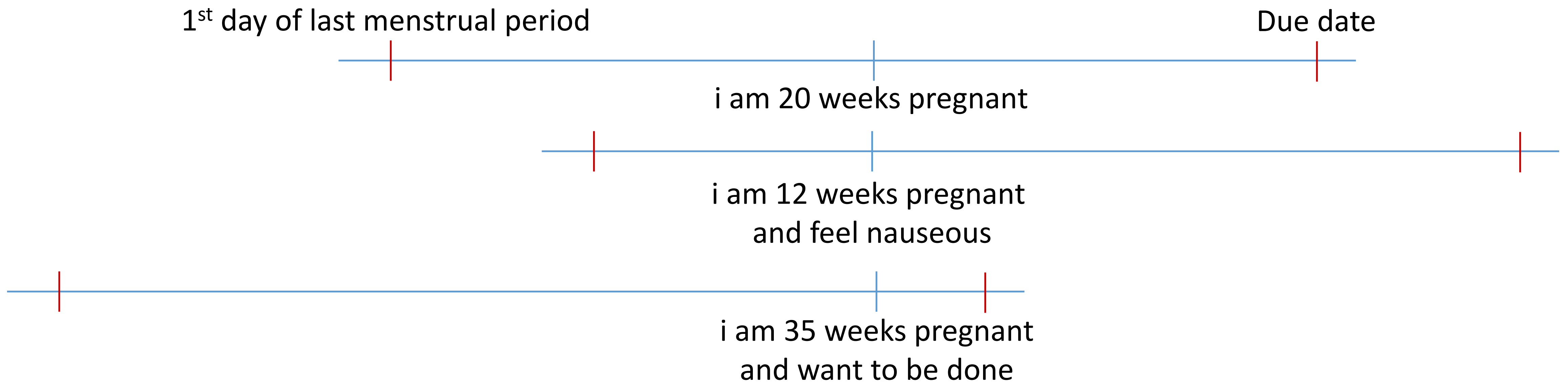
- Can tackle questions such as:
  - How do the experiences of pregnancy & childbirth manifest in the logs?
  - Can we predict who is pregnant, how far along they are, and when they give birth?

(Fourney, White, Horvitz, SIGCHI 2015)

Leveraging Self-Report Queries

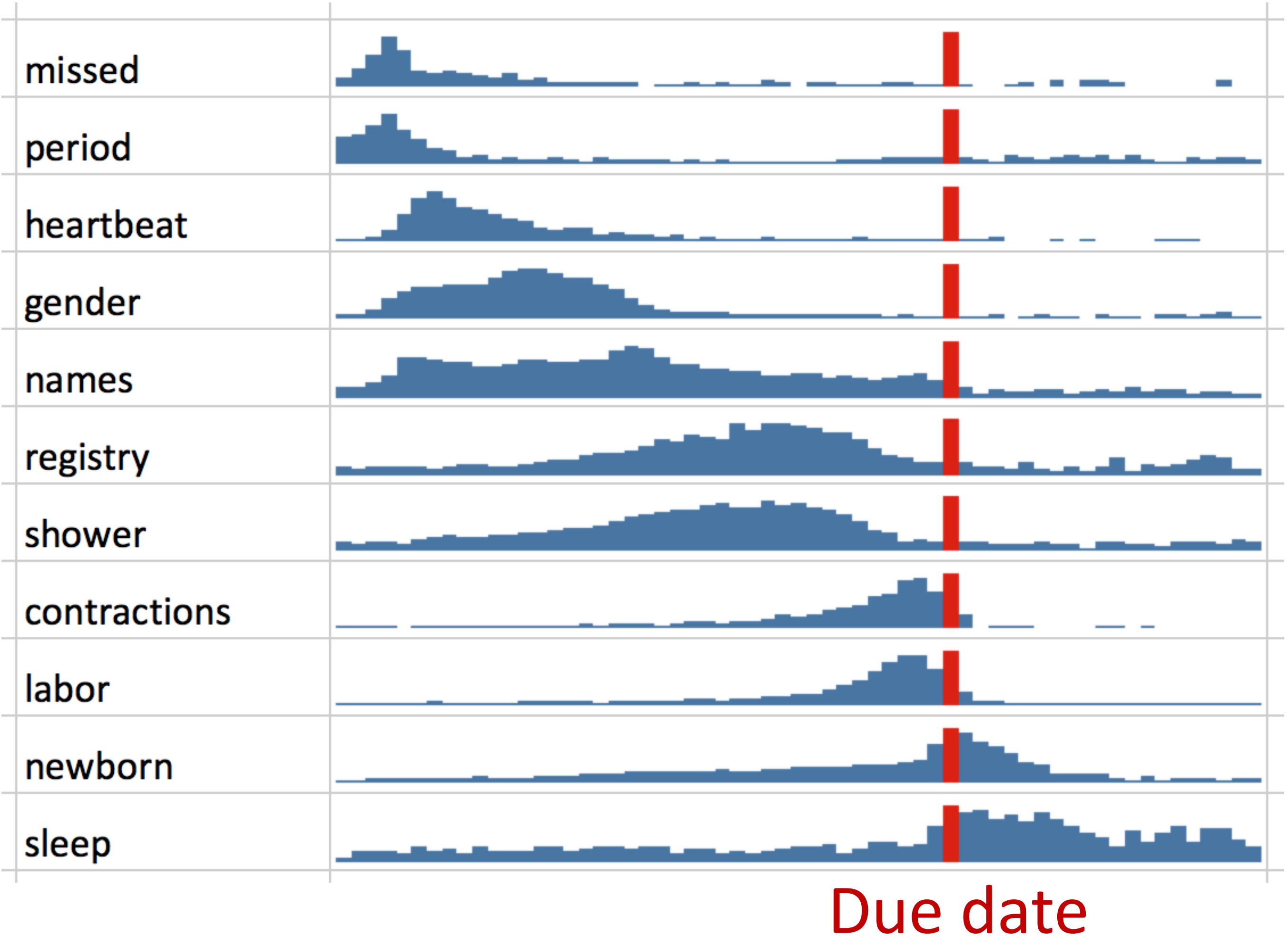
“I am *N* weeks pregnant”

- ~13,000 users searched this phrase on Bing.com, between June 2012 and December 2013
- Assume are as reliable as survey responses, especially since unprompted
- Places users on a well-known timeline



Characterizing and Predicting

Compute temporal distributions of key query terms



Predict for non self-report users

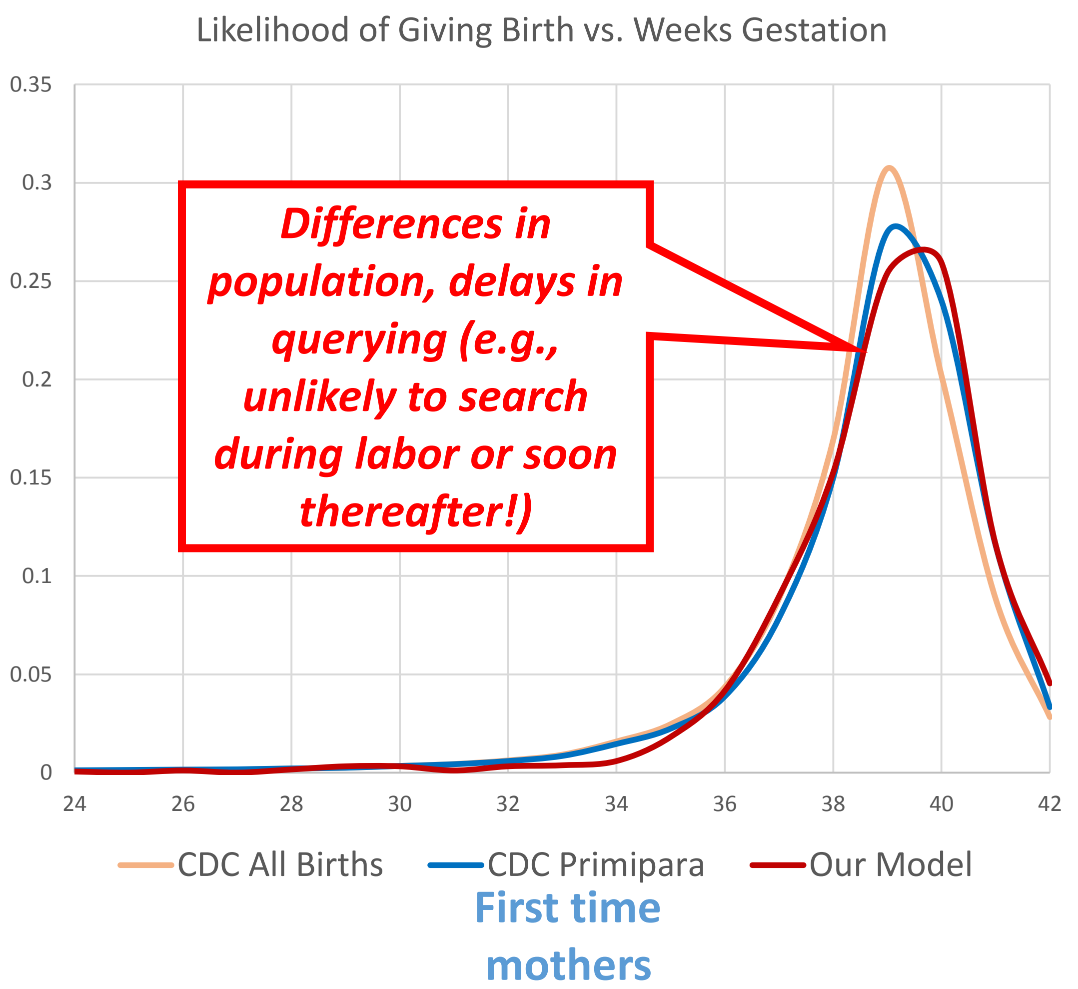
1. Fit features to a linear model via linear regression
2. If line slope 0.85-1.15, user → (likely to be) pregnant

Improves coverage beyond self-reporting searchers (8x increase in coverage)

Prediction error = ±0.685 weeks on average (median: 0 weeks)

Validation With External Data

Comparison with CDC data on birth @ weeks gestation



Many tests performed during pregnancy:

Compare spikes in query interest against **when tests are performed**

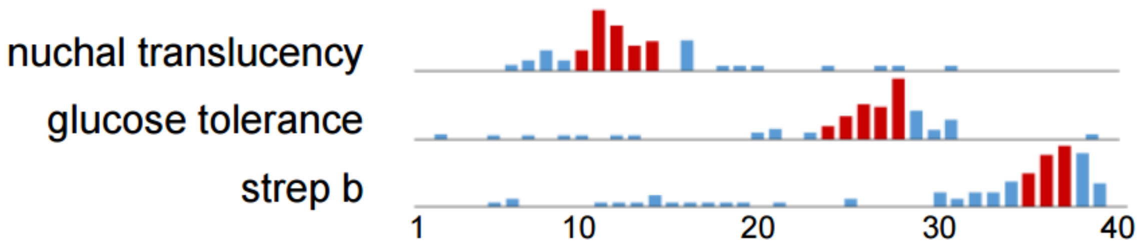


Figure 5: Histograms detail about how interest in three standard prenatal screening procedures vary over 40 gestational weeks. Bars show the proportion of searchers who have searched at least once for the bigram of interest in the associated week. Red bars report weeks in which each test is typically performed, as reported in [2], [18], and [24] respectively.

Applications

- Supporting mothers through personalized search
  - Tailoring search experience to stage or first time moms
  - Providing advice or guidance during related activities (e.g., flying when pregnant)

- Providing support via signals mined from logs
  - Quantifiable data to support assertions about pregnancy experiences
  - [back pain] → "Many expectant mothers query for this in Tri. 1, drops off in Tri. 2"

- Public health research
  - Studying sensitive issues, e.g., early induction of labor or drug abuse while pregnancy via search activity

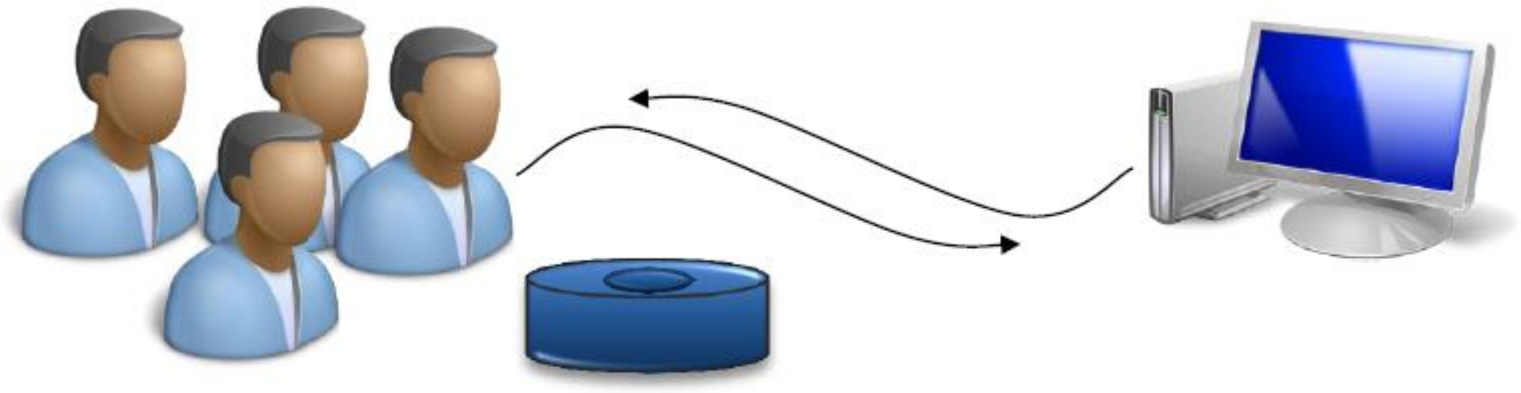
Example 3:
Identifying Drug Interactions and Adverse Drug Reactions from Search Logs

(White et al., JAMIA 2013)
(White et al., Nature CPT 2014)

Signals on Medication Adverse Effects

→ Web search as sensor for side effects?

1 in 250 of people query on top-100 drugs



- Adverse drug effects – 4th leading cause of preventable death in U.S.

Signals on Medication Adverse Effects

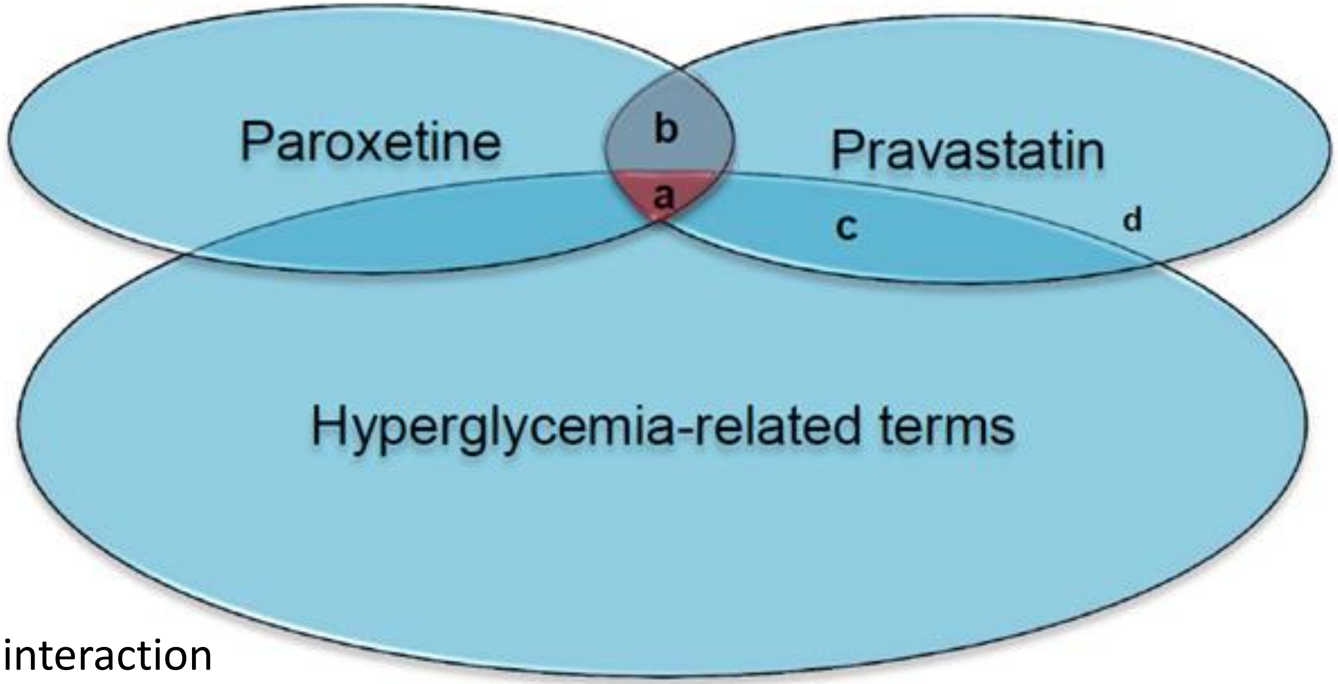
- Pharmacovigilence: spontaneous reports FDA *Adverse Event Reporting System* (AERS) – reports from patients, clinicians, drug companies

- 2011 finding in AERS analysis (Tatonnetti, et al.):
- *Paxil + Pravachol* → ✓ *Hyperglycemia*
- *Pravachol* → X *Hyperglycemia*
- *Paxil* → X *Hyperglycemia*

(Tatonnetti et al., Nature CPT 2011)

Web-Scale Pharmacovigilance

- Disproportionality analysis **using logs pre 2011**
- Reporting ratios (RR) -- observed vs. expected: $RR = \frac{\frac{a}{b}}{\frac{c}{d}}$



RR >> 1 → drug interaction

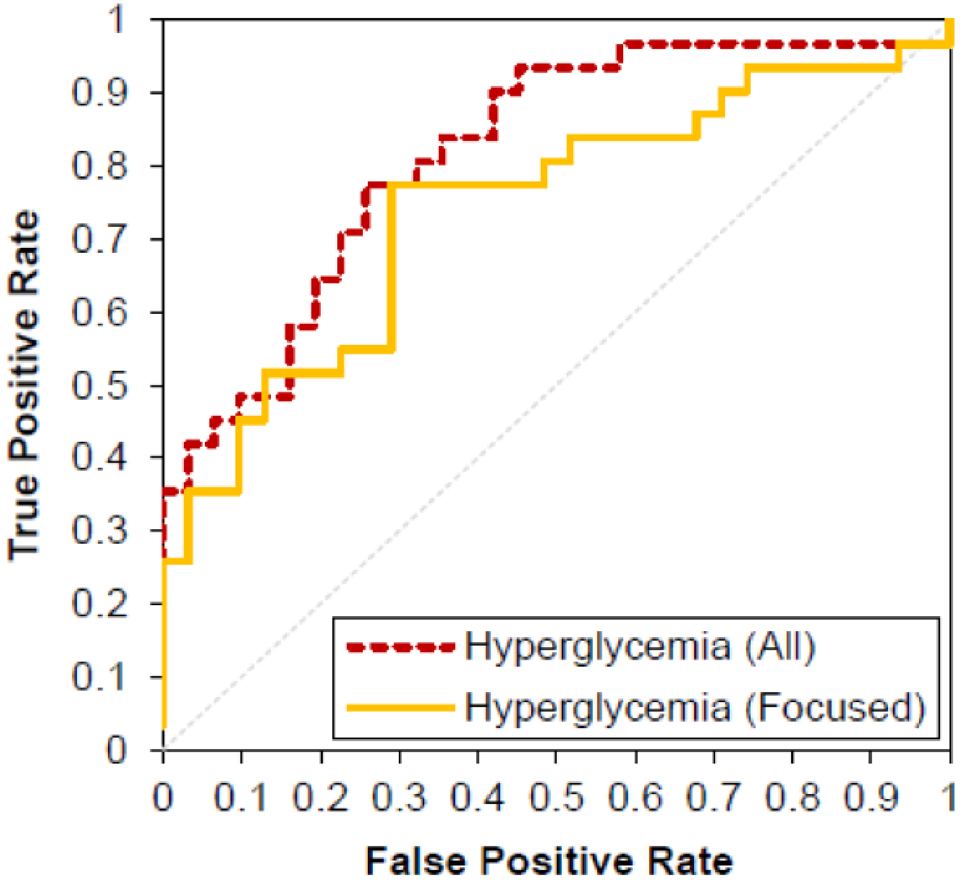
	a	*b*	*c*	*d*	*RR*	*95% CI (Lower, Upper)*	*p-value (one-tailed)*
Expected (pravastatin)	342	2716	2581	56302	2.747	2.438, 3.094	< 0.0001
Expected (paroxetine)	342	2716	3645	71243	2.461	2.189, 2.767	< 0.0001

Hyperglycemia terms:
polydipsia
thirst
thirstiness
thirsty
polyphagia
appetite increase
increased appetite
hunger
hungry
polyuria
frequent urinating
frequent urination
increased urination
hyperglycemia
hyperglycaemia
high glucose
glucose high
high blood glucose
blood glucose high
high blood sugar
blood sugar high
increase blood sugar
blood sugar increase

(White et al., JAMIA 2013)

Characterizing Sensor Error

- Test on known interactions
- 31 true positives for hyperglycemia (TP)
- 31 true negatives for hyperglycemia (TN)



Focused = Subset of terms with clearer connection to hyperglycemia

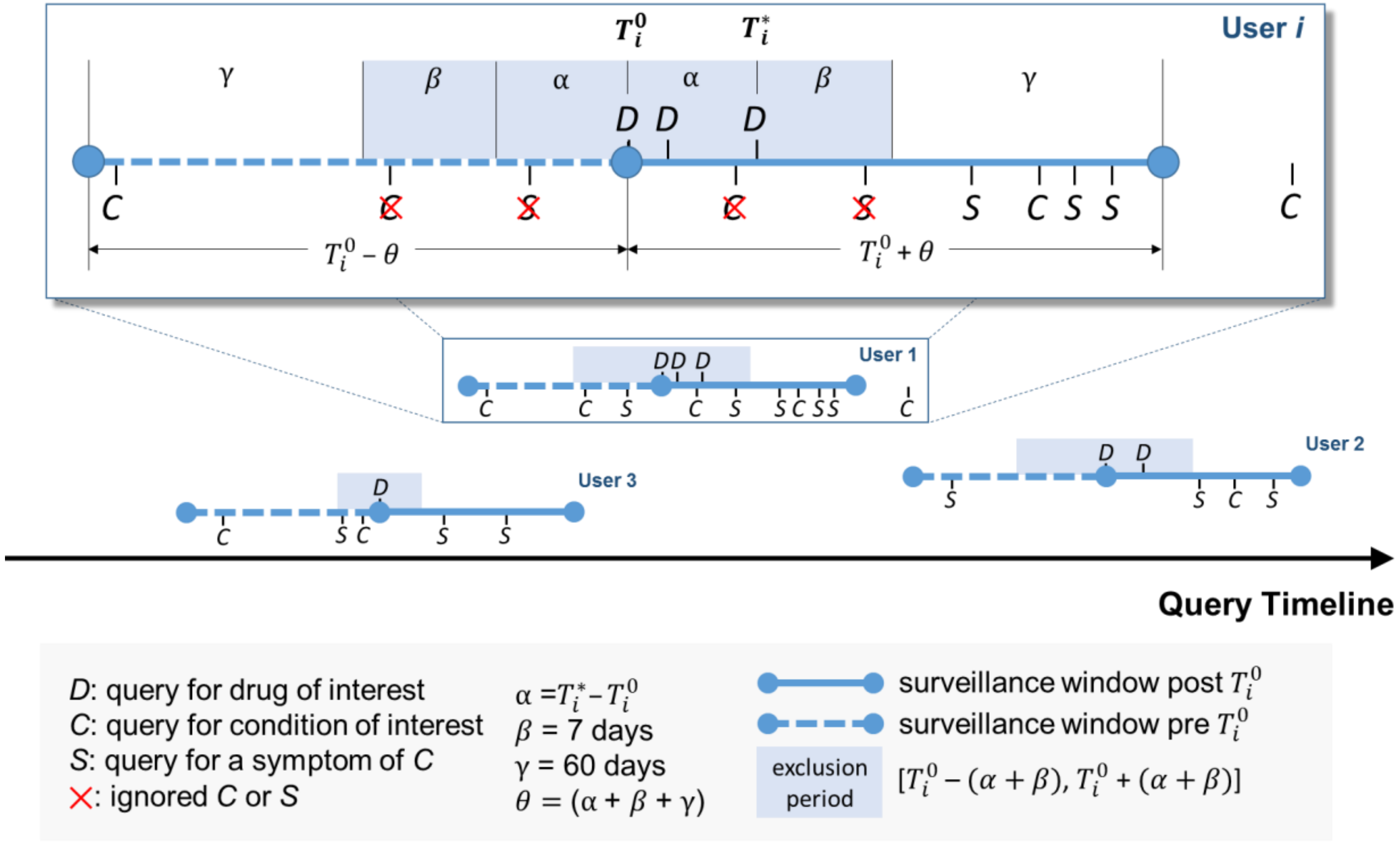
Label	*Drug 1*	*Drug 2*
TP	dobutamine	hydrocortisone
TP	dobutamine	triamcinolone
TP	dobutamine	prednisolone
TP	betamethasone	dobutamine
TP	glipizide	phenytoin
TP	dobutamine	methylprednisolone
TP	prednisolone	salmeterol
TP	salmeterol	triamcinolone
TP	betamethasone	terbutaline
TP	dexamethasone	dobutamine
TP	budesonide	salmeterol
TN	hydrochlorothiazide	tazobactam
TN	clindamycin	montelukast
TN	lamotrigine	nystatin
TN	methylprednisolone	rosuvastatin
TP	budesonide	formoterol
TN	loratadine	nystatin
TN	hydroxychloroquine	prochlorperazine
TN	labetalol	sertraline
TN	ciprofloxacin	vecuronium

Users as their Own Control

- Use search logs to detect adverse drug reactions not drug interactions
- Using ground truth from drug safety community (OMOP): 400 drugs + outcomes
  - Four outcomes: renal failure, GI bleed, liver injury, MI
- Within-user analysis: before and after **first instance of drug**

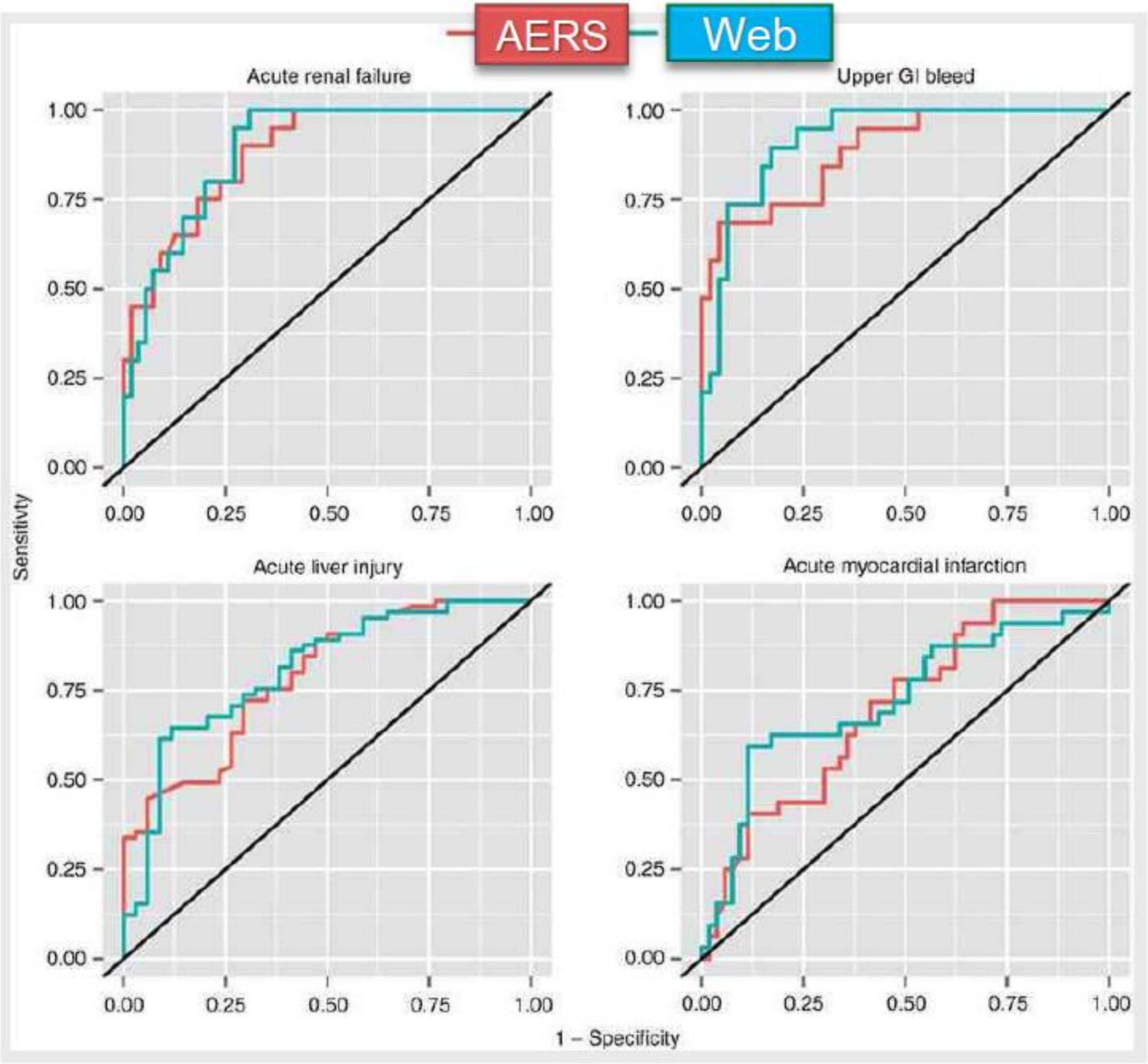
Exclusion periods to reduce effect of web on search behavior

→ More "experiential" signal



(White et al., Nature CPT 2014)

Rare, Serious Adverse Effects



FDA uses AERS & Multi-item Gamma Poisson Shrinker algorithm (DuMouchel and Pregibon, KDD)

White, Harpaz, DuMouchel, Shah, Horvitz. *Nature Clinical Pharmacology and Therapeutics*, 2014

Complementarity of Signals (AUROC)

	AERS	Search	Together
Acute Renal Failure	0.88	0.88	0.93
Upper GI Bleed	0.89	0.92	0.92
Acute Liver Injury	0.79	0.81	0.86
Acute Myocardial Infarction	0.70	0.73	0.75
Average	0.81	0.83	0.86

AUROC improvements over separate are statistically significant ($p < 0.05$)

Applications

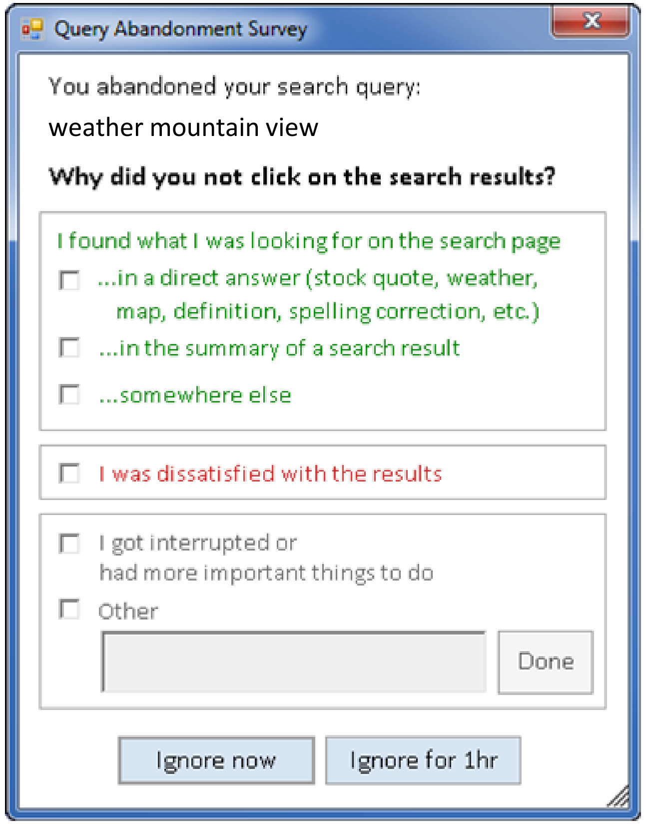
- Prediction of **unknown** drug interactions and adverse drug reactions
- Inform follow-on studies and clinical trials
- BLAERS (internal MSR tool)
- Prospective analysis
  - Needs time-indexed ground truth

- Impact through:
  - Early alerting for patients
  - Partnerships with government agencies
  - Partnerships with drug companies

Part IV: Opportunities and Challenges

(Some of the) Limitations of Log Analysis

- Logs offer ***SCALE*** but should be used in combination with more traditional instruments (intake logs, surveys, clinical studies)
  - Logs provide information about the "what", not the "why"
  - Opportunities for log-survey linking methodologies, **in-situ** monitoring of behavioral rationales via focused surveys →
- Experiential vs. exploratory
  - Difficult to distinguish those affected from those interested
- Multiple people using the same machine (intertwined behavioral signals for 50%+ of userids)
  - Recent research on **activity attribution** may help (White et al., WWW2014)

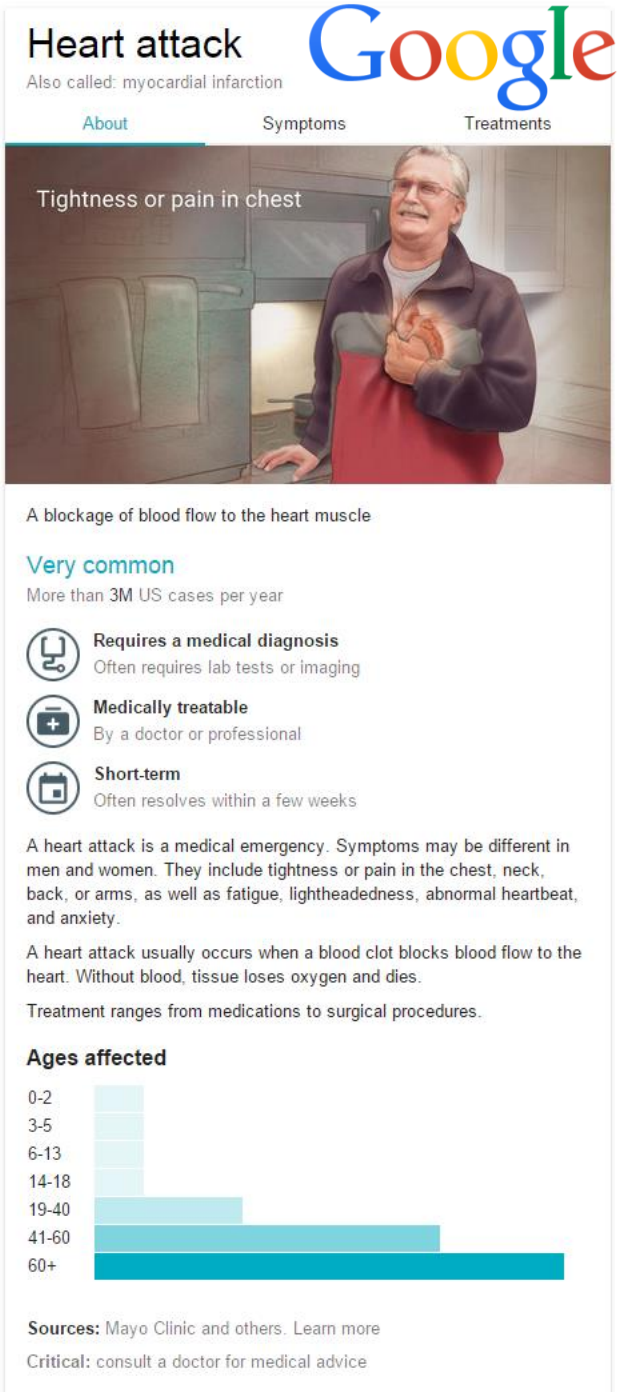


(Diriye et al., CIKM 2012)

Opportunities and Challenges

- Health information seeking → Important, prevalent
- Clear benefit to people (in surfacing reliable content), cou
- Mixed methods important to fully understand observed b

- Searchers need help in finding reliable content, learning a managing decisions about self-treatment & pursuing prof



Opportunities and Challenges

- Health information seeking → Important, prevalent
- Clear benefit to people (in surfacing reliable content), could save lives!
- Mixed methods important to fully understand observed behavior

- Searchers need help in finding reliable content, learning about conditions, managing decisions about self-treatment & pursuing professional care, etc.

- Significant ethics and privacy implications – health is personal
- Need clearer paths to impact – connections with companies/agencies
- Emphasized big data—"small data" is important too

Opportunities and Challenges

- Sensor systems for public health monitoring
  - Search is a limited lens on online behavior – also tweets, social media posts, etc.

- Need to understand biases in data – validate data against known truth

- Mining can’t occur in isolation – needs partnerships for impact

- Small data mining → personal health management
  - Triangulate signals from many sources, devices, EHR, etc. (with informed user consent!), logs as memory aid

Thanks for listening!

Thanks to the BCS-IRSG and Microsoft Research for the KSJ Award. I’m deeply honored.