

Search Engines, their Evaluation and the Impact on Health Decisions

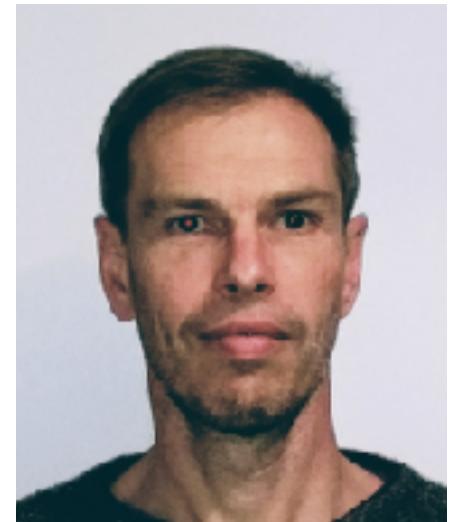
Dr Guido Zuccon

g.zuccon@uq.edu.au – ielab.io/guido

ITEE, University of Queensland
St Lucia, QLD, Australia

Work presented here is in collaboration with: Anton van der Vegt, Jimmy,
Dr Bevan Koopman (AEHRC/CSIRO), Dr Gianluca Demartini

Our research addresses problems in information retrieval (search engine technology), information seeking, data science, and health informatics



Anton van der Vegt

PhD Student

Bridging the Human-Task Cognitive Gap in Medical Search



Jimmy

PhD Student

Search Engines that Help People Make Better Health Decisions



Harrisen Scells

PhD Student

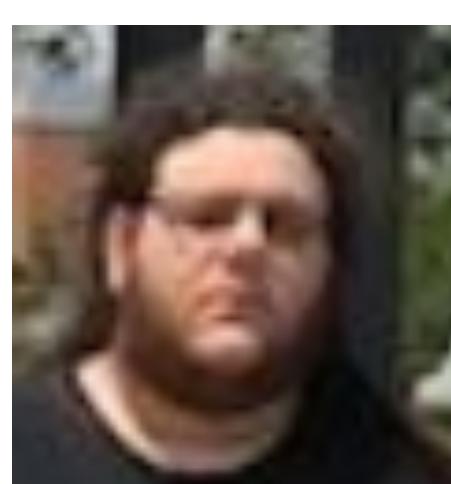
Systematic Review Creation with Information Retrieval



Shengyao Zhuang

PhD Student

Online Learning to Rank



Sebastian Cross

PhD Student

Health Conversational Search



Sitthichoke Subpaiboonkit

PhD Student

Causality Discovery in Drug-Drug Interactions



Suresh Pokaharel

PhD Student

Similarity Computing on Complex Health Data



Daniel Locke

PhD Student

Methods for Improving the Effective Retrieval of Case Law



Rischan Mafrur

PhD Student

Insights Recommendation for Exploring IoT Data

Health Data Science at UQ

- Dr Mahsa *Baktashmotlagh* – image processing and classification
- Dr Gianluca *Demartini* – human intelligence
- A/Prof Anders *Eriksson* – computer vision and machine learning
- Dr Wen *Hua* – knowledge graph mining
- Prof Shazia *Sadiq* – data quality
- Dr Sen *Wang* – health data mining
- Dr Guido *Zuccon* – health search
- (Prof Xue *Li* – currently at Neusoft)



ICU & ED



THE AUSTRALIAN
EALTH
RESEARCH CENTRE

Gold Coast University Hospital



Australian Government
Australian Research Council

电子科技大学医学院
附属肿瘤医院
The affiliated cancer hospital,school of medicine,UESTC
1956



中南大學湘雅醫院
XIANGYA HOSPITAL CENTRAL SOUTH UNIVERSITY

江苏省中医院
Jiangsu Province Hospital of Chinese Medicine



UQ-Neusoft Joint Lab in Health Data Science

Research Activities:

- Heterogeneous, Multilingual **Health Knowledge Graph** creation, scale-up, and inference for information extraction, clinical case review, evidence based medicine
- **Multimodal clinical image analysis** for abnormalities identifications, boundary detection



Dr Guido Zuccon
Health Search Investigator &
Lab Director



Dr Wen Hua
Knowledge Graph
Investigator



Ekaterina Kramtsova
PhD Student
Image Analysis



Dr Mahsa Baktashmotlagh
Image Analysis
Investigator



Dr Sen Wang
Machine Learning
Investigator



Linh Le
PhD Student
Knowledge Graph



Dr Gianluca Demartini
Human Intelligence
Investigator



Dr Tony Chen
PostDoc
Knowledge Graph &
Image Analysis



Bing Liu
PhD Student
Knowledge Graph

In this talk...

- I will present two (user-based) studies:
 1. Clinicians searching for literature/guidelines for evidence based practice (CDS)
 2. Consumers searching for health advice online (CHS)
- I will reflect on the **implications** of our empirical results and observations with respect to:
 - How we **evaluate** search systems in Information Retrieval
 - What problems and areas **future research** effort in advancing current search technology may need to address

CDS: Our health and longevity depends on good clinical decisions

Clinical Questions: An information need that arises as a result of a patient encounter directly pertaining to the diagnosis and/or treatments for a patient condition

- What is the diagnosis?
- What medical tests to order?
- What treatments to suggest?



- 1 question every 2 patient visits
- More than half (53%) of clinical questions don't get answered

[Del Fiol et al., 2014]

Does it matter if clinical questions are answered?

- Considerable supporting evidence that **answering clinical questions** leads to significant changes to provided care: if relevant information is provided to clinicians, then **changes to clinical decisions are observed** with respect to:
 - Patient advice (72%)
 - Treatment decision (59%)
 - Choice of diagnostic test (45%)
 - Choice of drug (treatment) (45%)
 - Diagnosis decision (29%)
 - Thus avoiding: additional tests and procedures (49%), additional tests and procedures (49%), additional surgery (21%), patient mortality (19%)
 - **Correctly answering** clinical questions **impacts** patient **outcomes** and the **efficiency** and **effectiveness** of the healthcare system
- Marshall JG. The impact of the hospital library on clinical decision making: the Rochester study. Bull Med Libr Assoc. 1992;80(2):169.
- Marshall JG, et al. The value of library and information services in patient care: results of a multisite study. J Med Libr Assoc JMLA. 2013;101(1):38.

How can clinicians answer their questions?

- Current medical search engines, such as PubMed, can help clinicians to answer their clinical questions.
- However, they are often not used?
 - Clinicians report **lack of time** as the primary **barrier to using such systems** [Cook et al., 2013]
- In fact, clinicians frequently operate under **time pressure**:
 - Average consultation time across more than half of the world's population is less than 5 minutes [Irving et al., 2017]
 - Across ten industrialized countries, over one-third of all primary care physicians are dissatisfied with the time available per patient [Osborn et al., 2015]



What we have been interested in studying

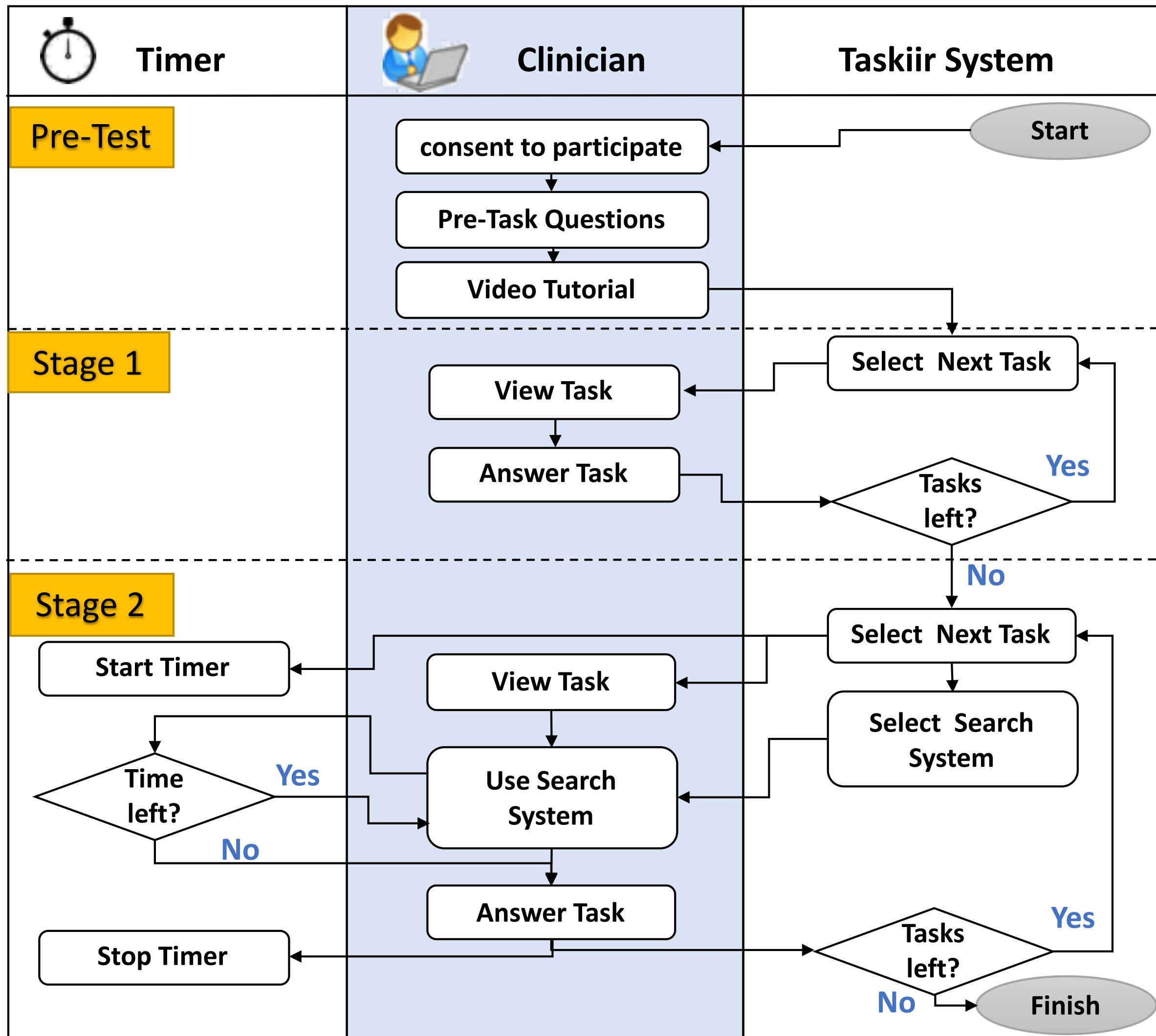
1. Does the use of a **search engine** enable clinicians to make **better clinical decisions**?
 - Does finding (relevant) information translates into a making a good decision? Can clinicians actually find relevant information?
2. How does **time pressure** impact clinical **decision** making, when using a search engine?
 - If I am short in time, do I still make good decisions when using a search engine?
3. Does a **significantly better search engine** translate to **better and faster clinical decisions**?
 - After all, We (the IR community) have spent some much time (and \$\$) in building better and better search engines

Based on: (1) van der Vegt A, Zucccon G, Koopman B, Deacon A. Impact of a Search Engine on Clinical Decisions Under Time and System Effectiveness Constraints: Research Protocol. JMIR Res Protoc 2019;8(5):e12803
... and others under review...

Methodology: User Study

- 85 final year medical students, 16 doctors and 8 nurses
- 16 questions: 1,653 samples
- Answer= Yes, No, Conflicting
- Questions are real-life scenarios:

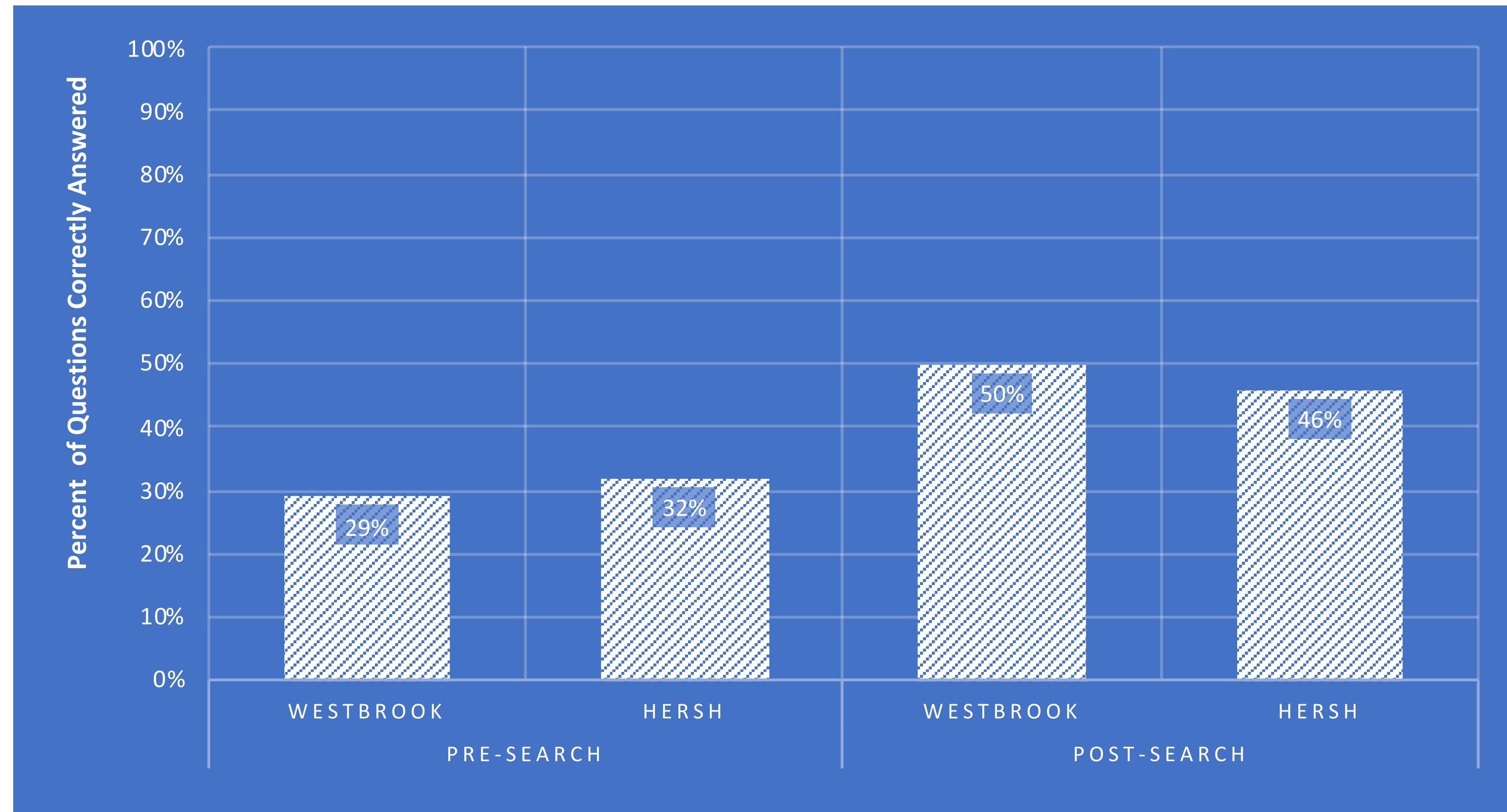
A 48 year old man presents with severe right sided loin pain and is diagnosed with a 4mm distal ureteric calculus. Has Tamsulosin been shown to increase the chances of the calculus passing?



Does the use of a search engine enable clinicians to make better clinical decisions?

Similar Prior Studies:

- [Westbrook et al, 2005]:
75 clinicians, 8 clinical questions
- [Hersh et al, 2002]:
45 medical students, 21 nursing students, total 324 questions

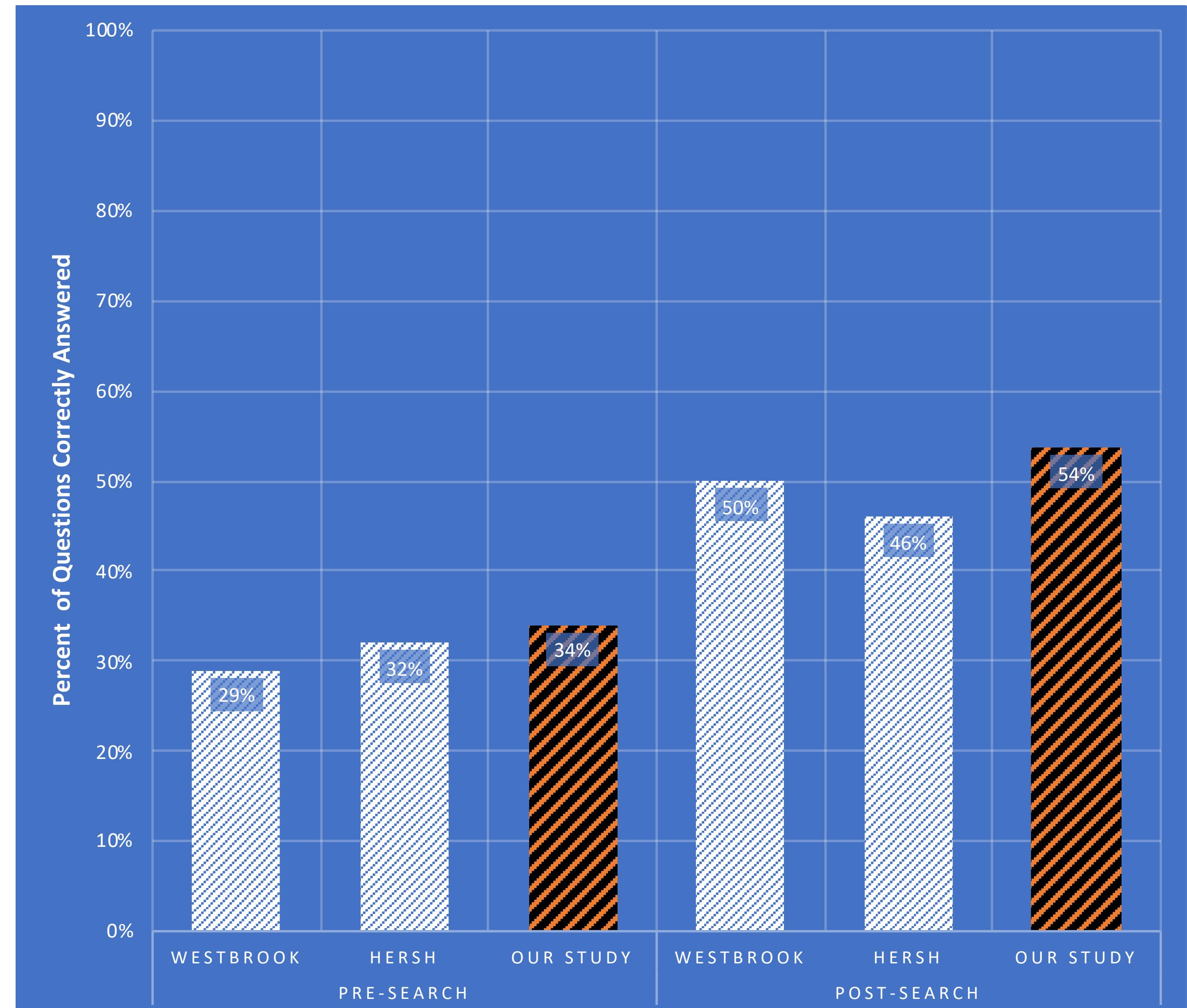


Does the use of a search engine enable clinicians to make better clinical decisions?

Our results:

- Slightly higher pre-search correct answer rate
- Slightly higher post-search correct answer rate
- Similar **significant benefit** from using a medical **search** engine

Note: All differences between pre-search and post-search answers are significant
(McNemar's Chi-squared Test for symmetry)

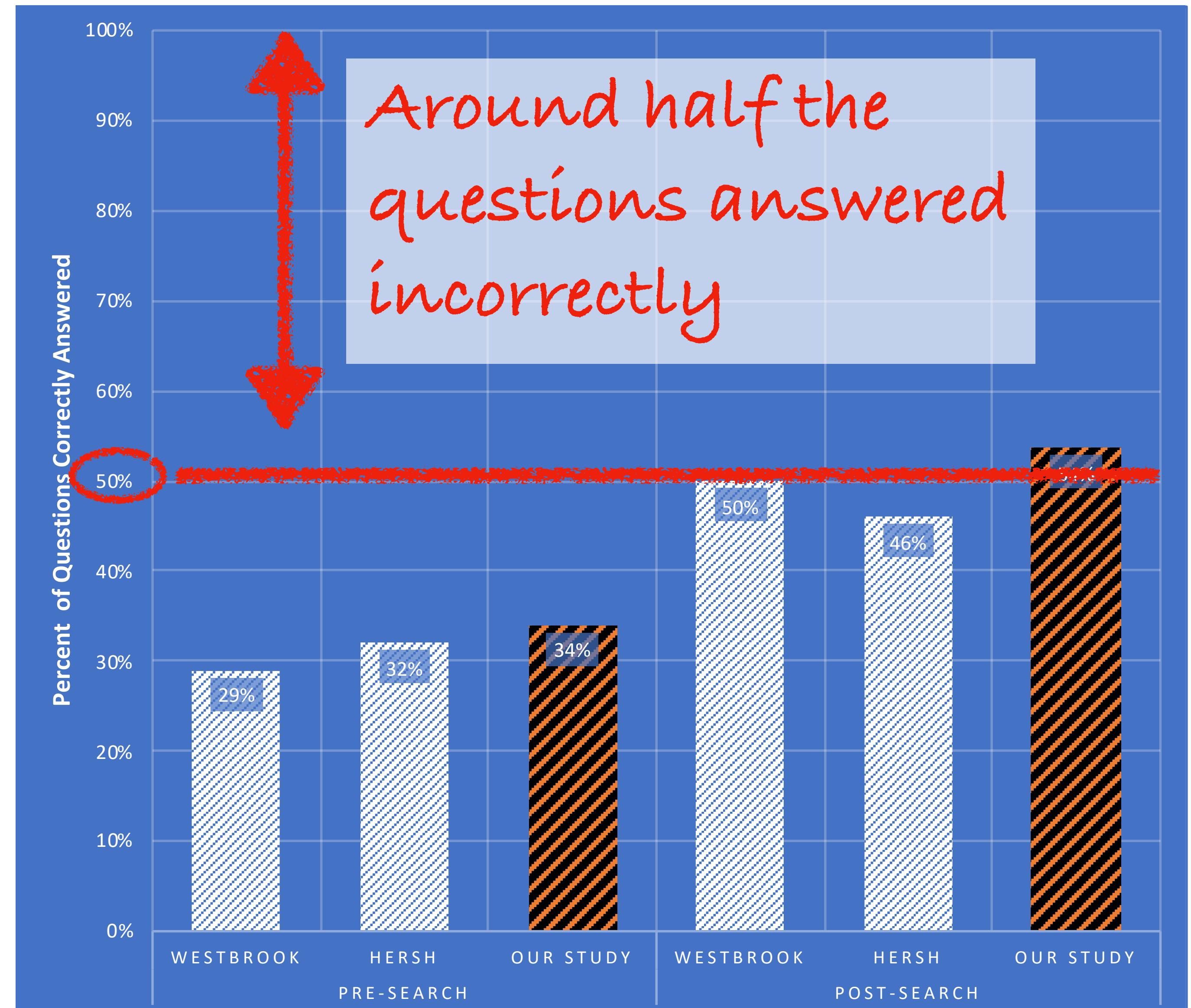


Does the use of a search engine enable clinicians to make better clinical decisions?

Our results:

- Slightly higher pre-search correct answer rate
- Slightly higher post-search correct answer rate
- Similar **significant benefit** from using a medical **search engine**

Note: All differences between pre-search and post-search answers are significant
(McNemar's Chi-squared Test for symmetry)

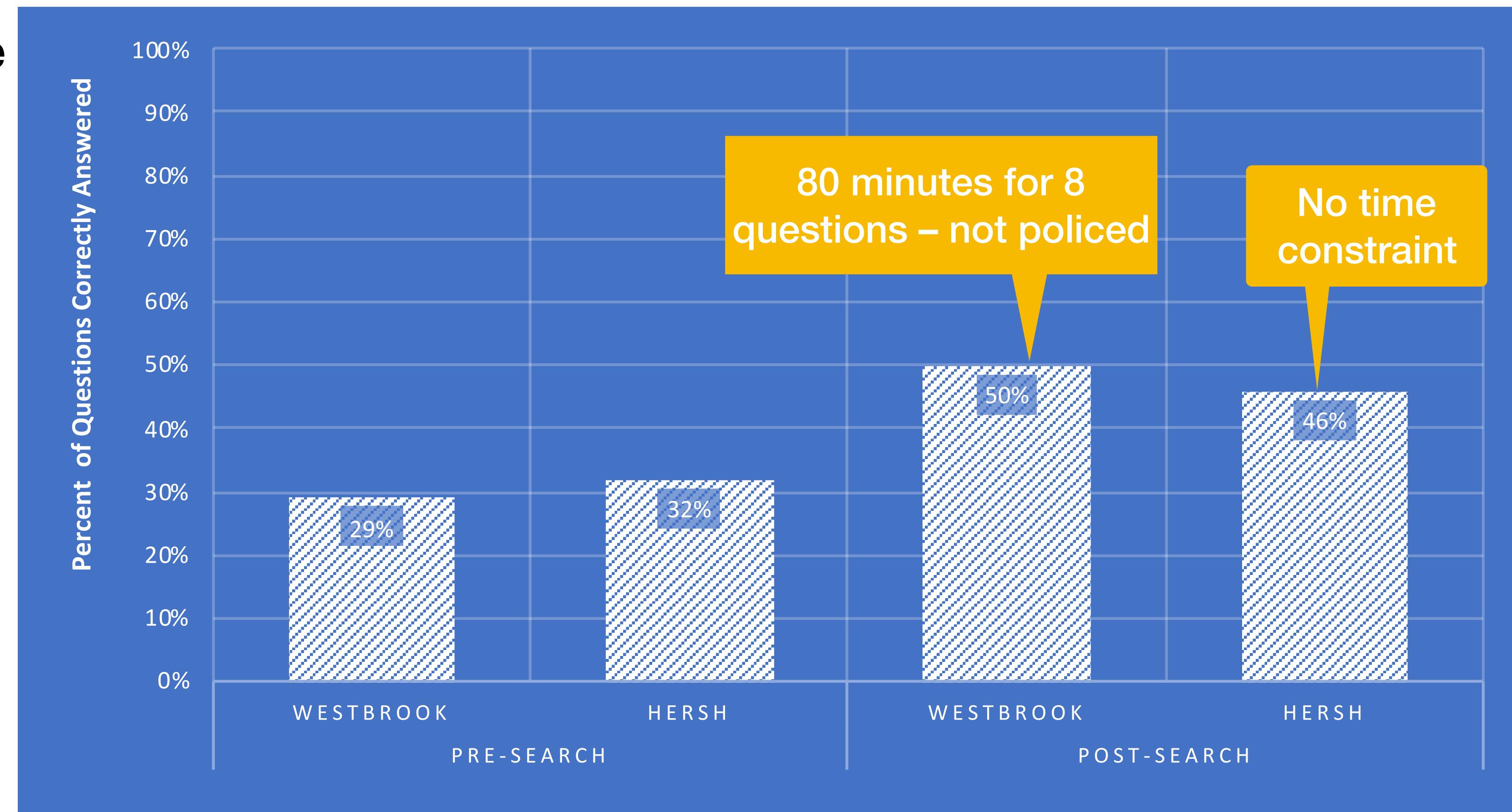


How does time pressure impact clinical decision making, when using a search engine?

Similar Prior Studies: time pressure not controlled for, or measured

[Crescenzi et al, 2016]: Impact of time constraints on user during information search. Users report:

- Greater time pressure
- Higher perceived task difficulty
- Less satisfaction with their performance



Study Design: Imposing Time Pressure



Search time constraints applied to induce time pressure



Based on [Westbrook et al, 2005]: mean search time = 6.1mins, std. dev = 3 mins. Therefore, 3 Time constraints applied to search:
3 minute, 6 minute and 9 minute

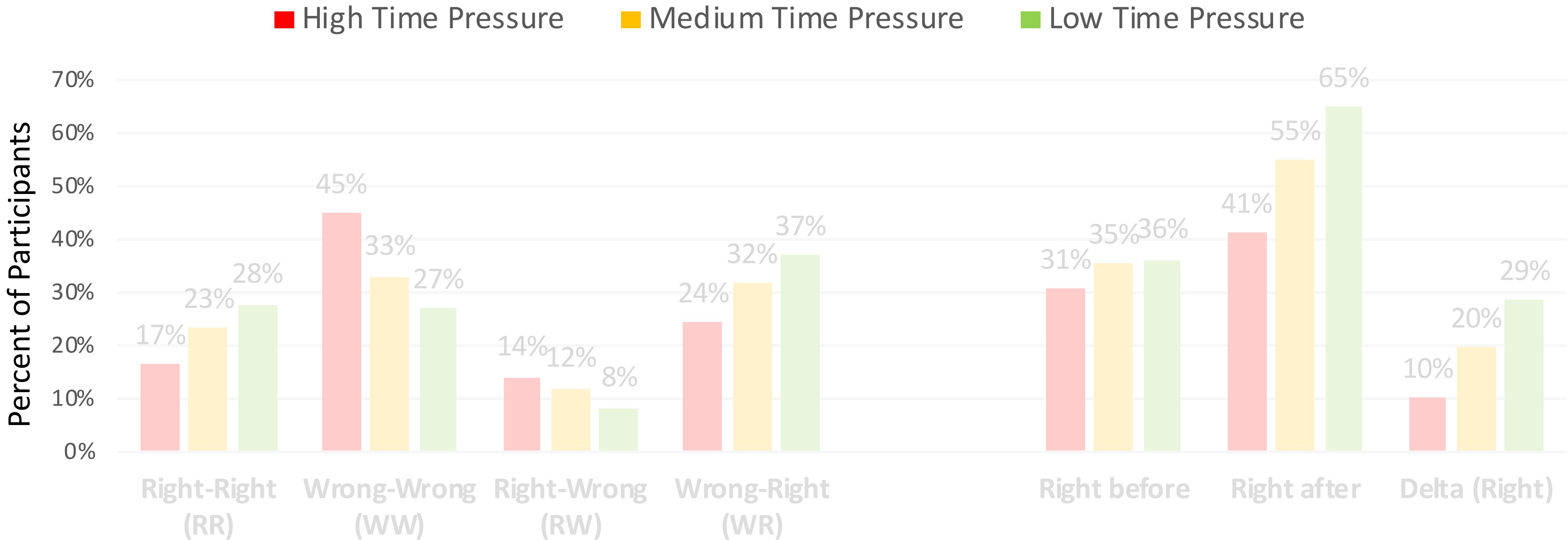


Within-subject, time constraint changes each question

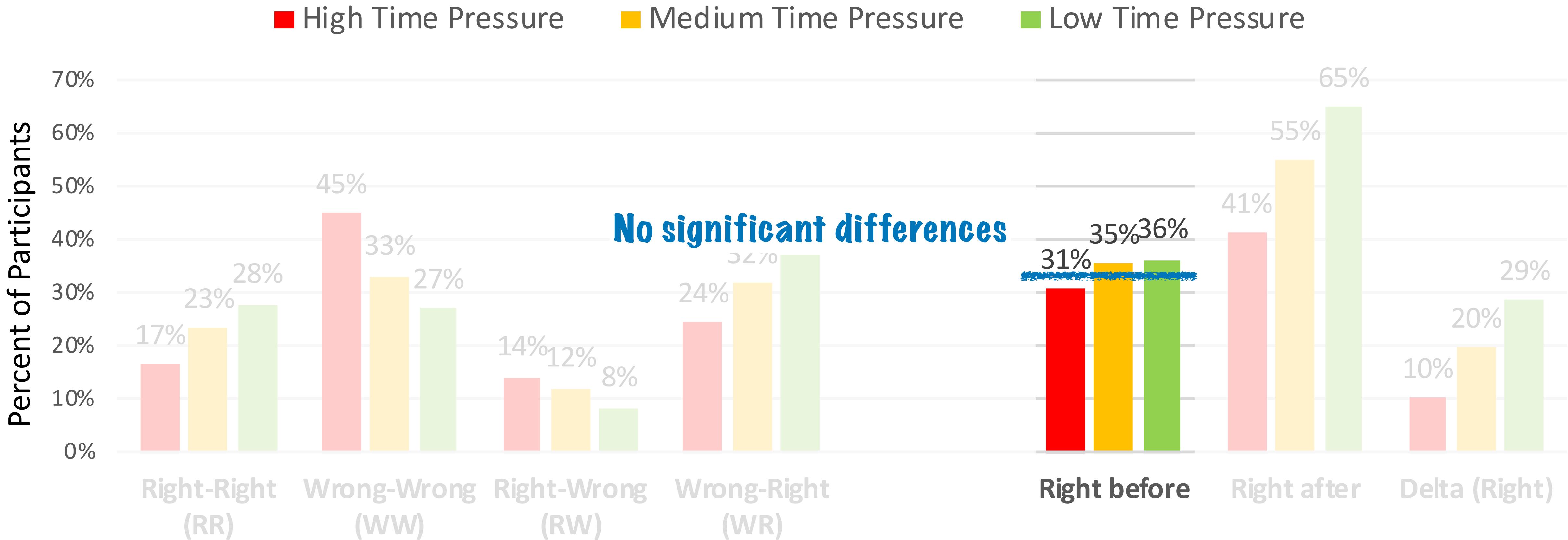


Participant notified of time at start of each question, then given a minute-by-minute countdown timer, then warning at 30 seconds to go, then forced to stop and answer question when time is up

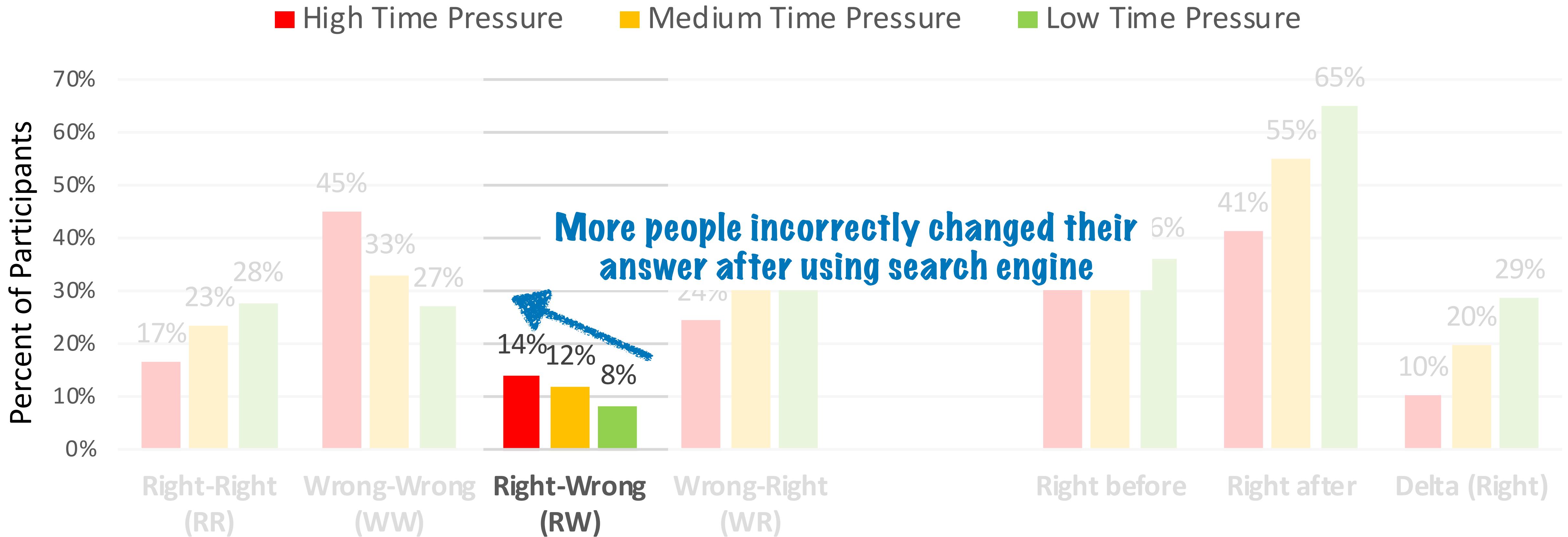
Impact of Time Pressure of Decisions



Correctness Before Search

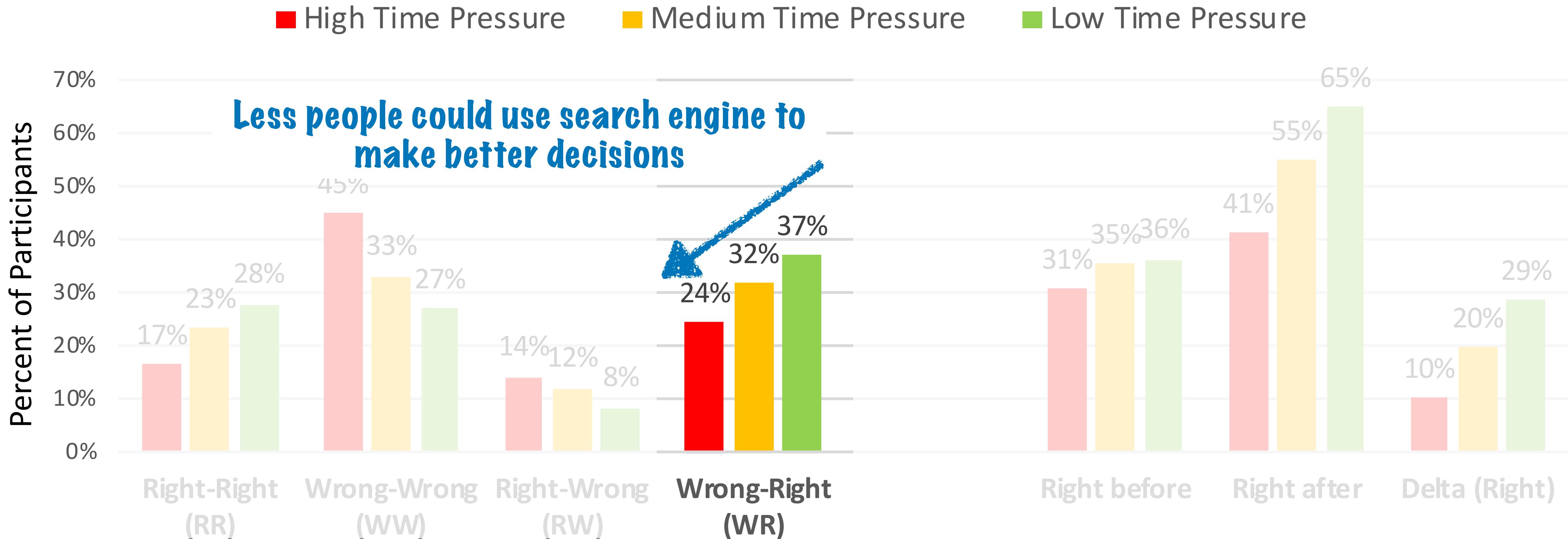


Right before search, then wrong after



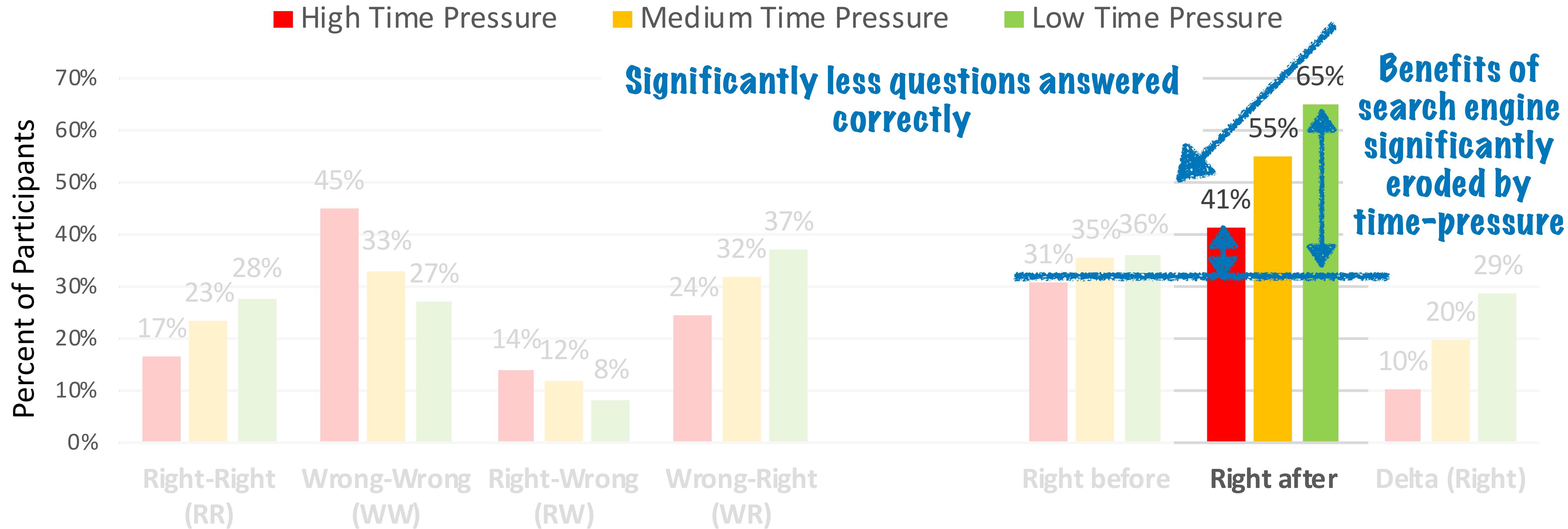
Increases with increased time pressure

Wrong before search then right



Decreases with increased time pressure

Right after using search system

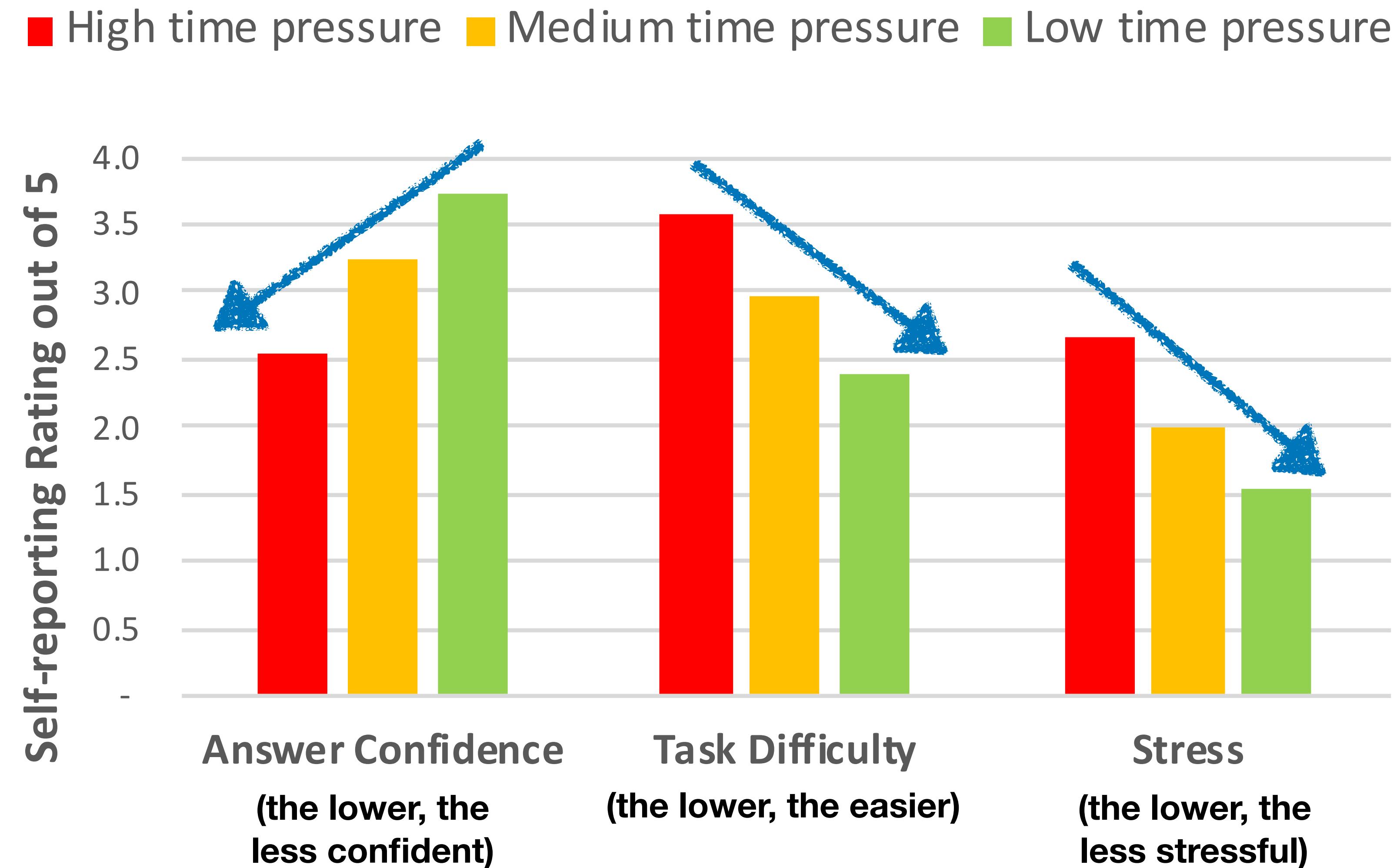


Decreases with increased time pressure

Impact of time pressure on participants

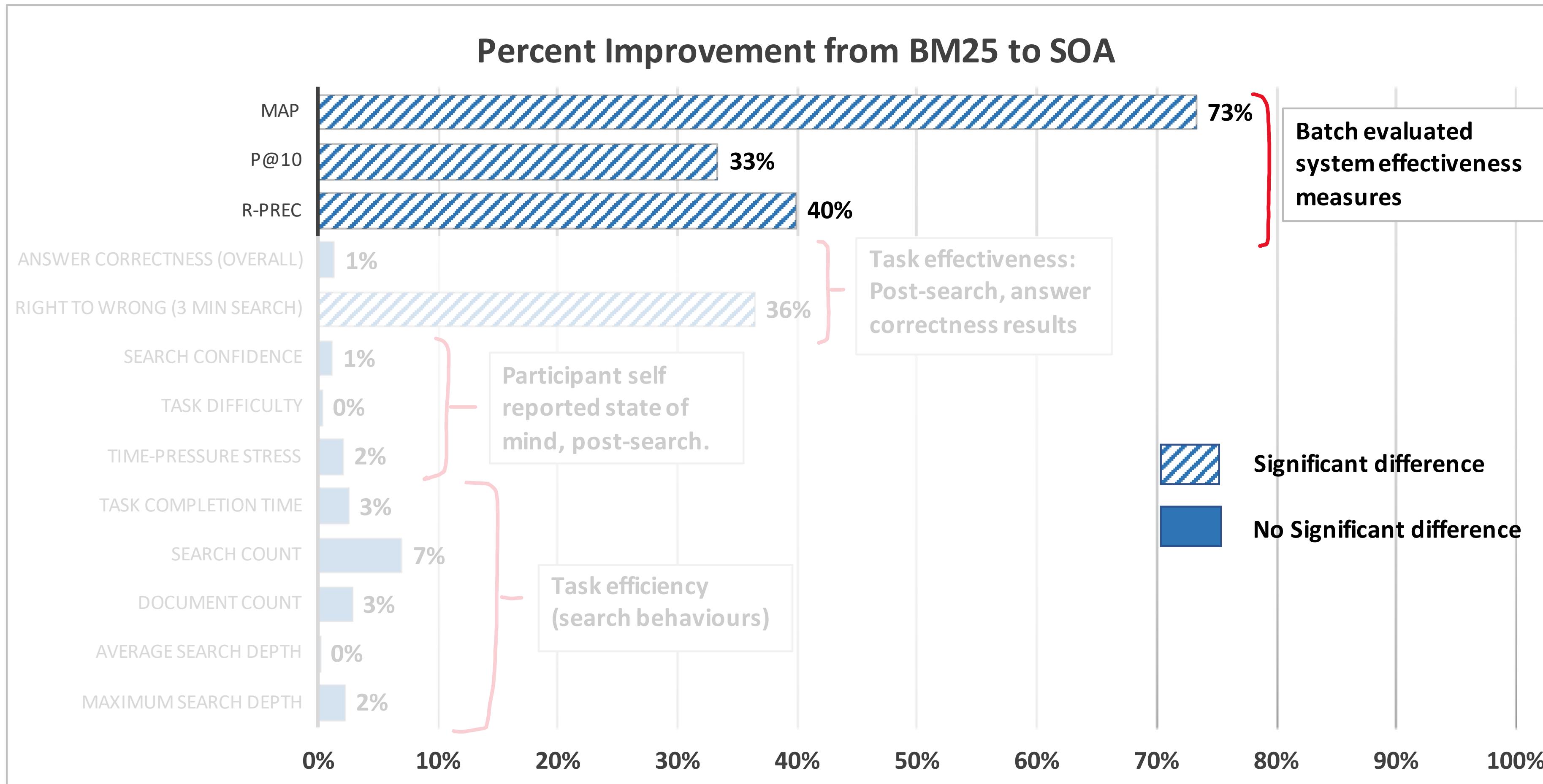
- Decreased answer confidence
- Increased task difficulty
- Increased stress

Note: All differences between time pressure levels are significant
(TukeyHSD p<.05)



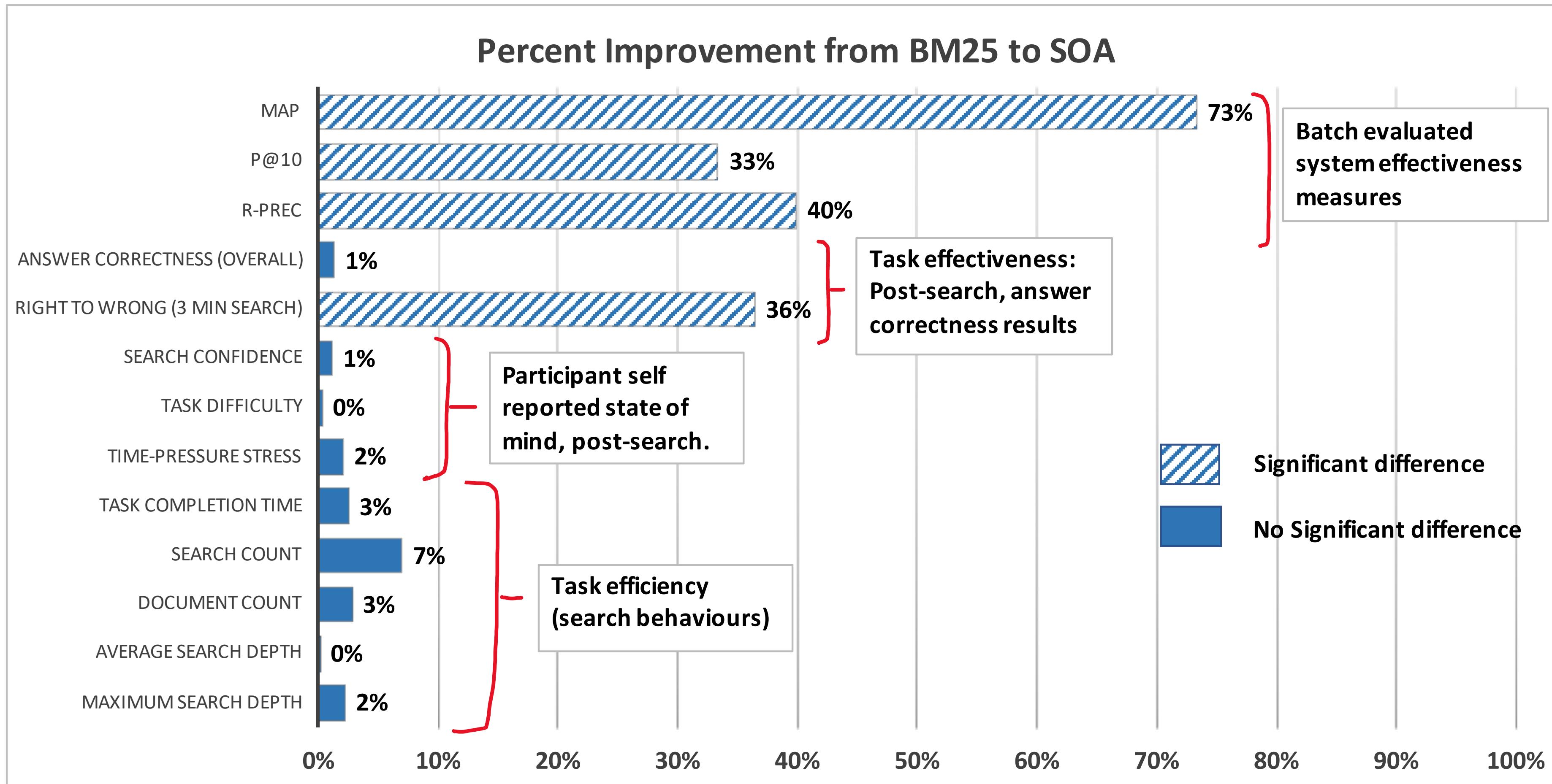
Time pressure occurs because of increased numbers of searches, documents read, search time, read time and depth of search

Does a significantly better search engine translate to better and faster clinical decisions?



- Significant difference in off-line traditional IR performance, e.g., +73% MAP

Does a significantly better search engine translate to better and faster clinical decisions?



- Significant difference in off-line traditional IR performance, e.g., +73% MAP
- **No/Minimal Impact on decisions!**
 - Under the tightest, time-constrained search conditions, a better performing search engine **prevents** more **clinicians** from **incorrectly changing** their clinical **decisions**

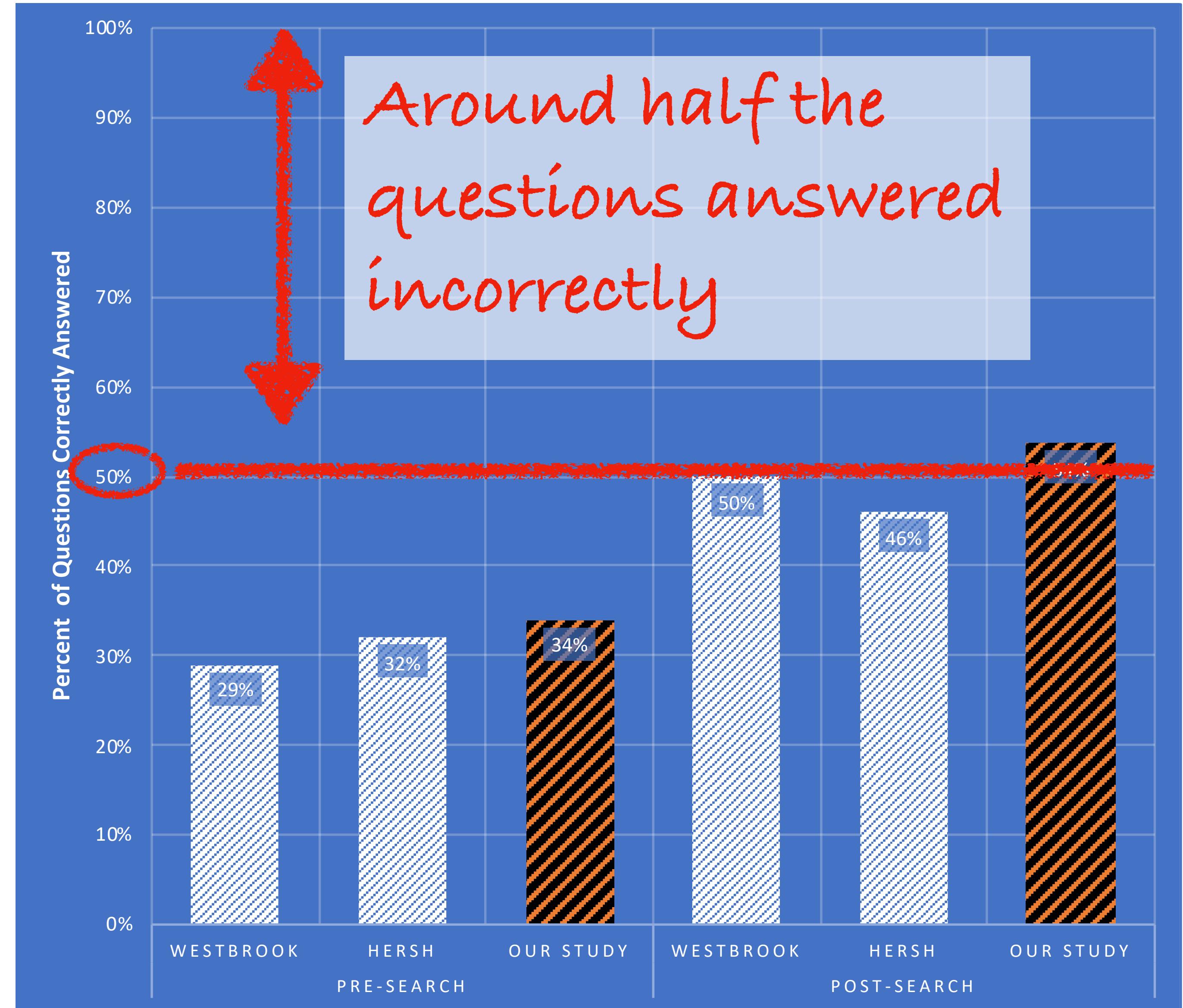
More General Implications for Search Engine Evaluation

- Traditionally, **search engine evaluation** has measured a **function** of the presence of **relevant documents** in the ranking, e.g. precision, recall, MAP, nDCG
- BUT – users do not search to find relevant documents
- They **search to fulfil tasks**, and often to **make decisions**
- Our results show that **good rankings** in terms of relevance are **NOT equal** to increase in **correct decisions**
- In cases such the one seen here, **evaluation should be more holistic**: it should consider the root cause for using the search engine, i.e. making decisions
- (see next for what may actually impact decisions)

So, why do clinicians fail to correctly answer around half the questions?

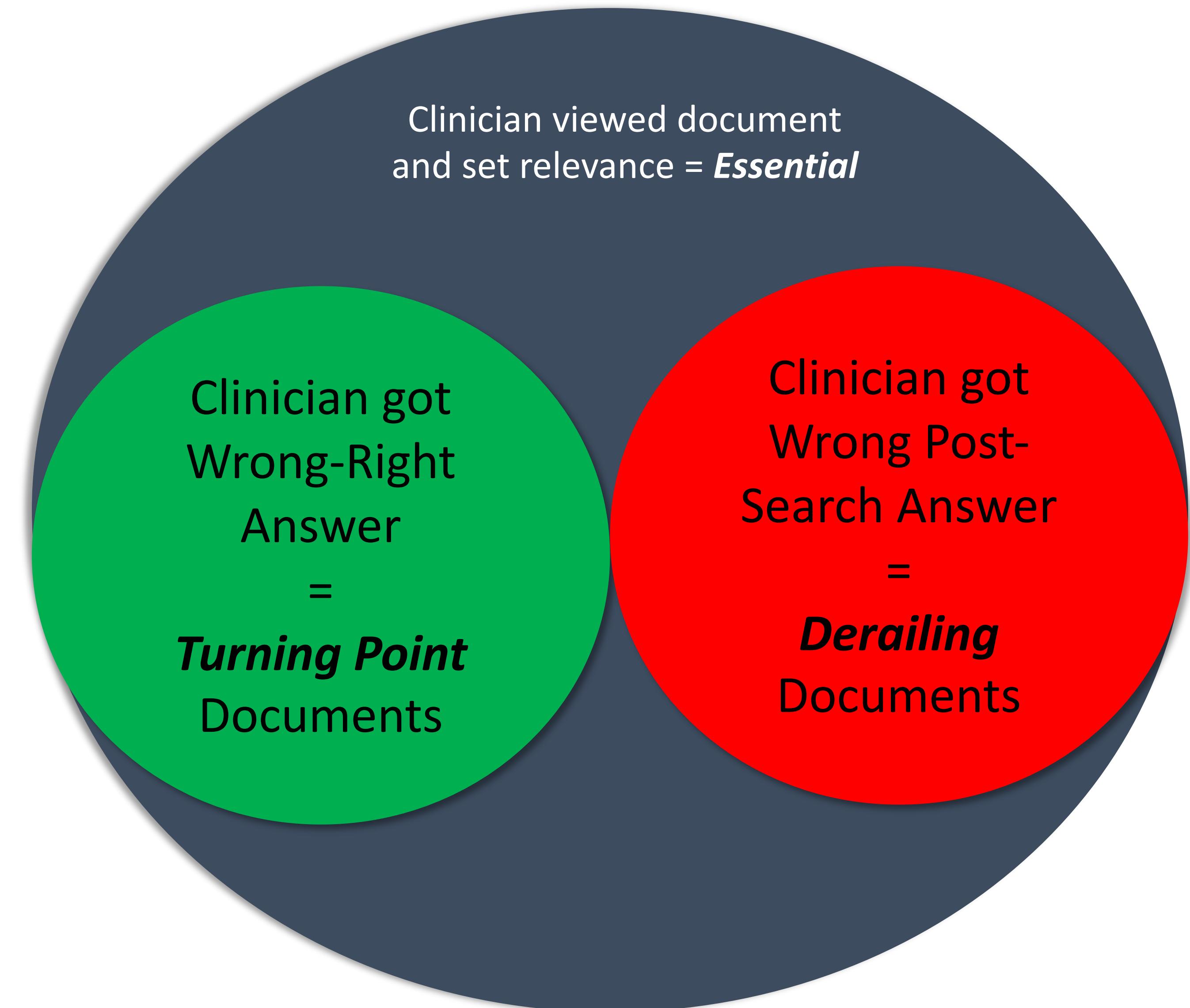
Is it because of:

- the corpus?
- the search engine?
- the participant?



Analysis: Success Factor Model

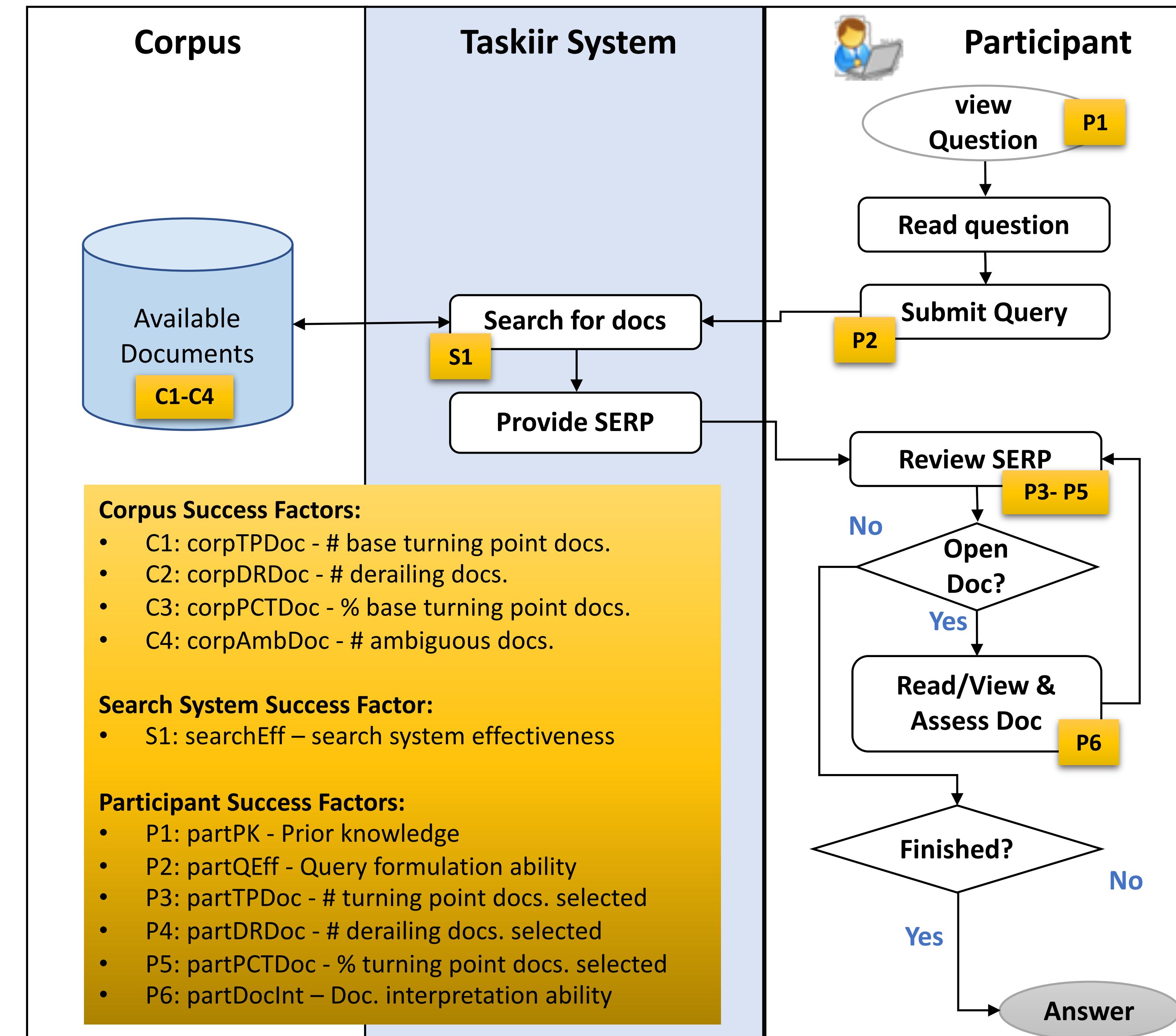
- Define 2 types of documents:
 - Turning Point
 - Derailing
- Define factors based on these docs



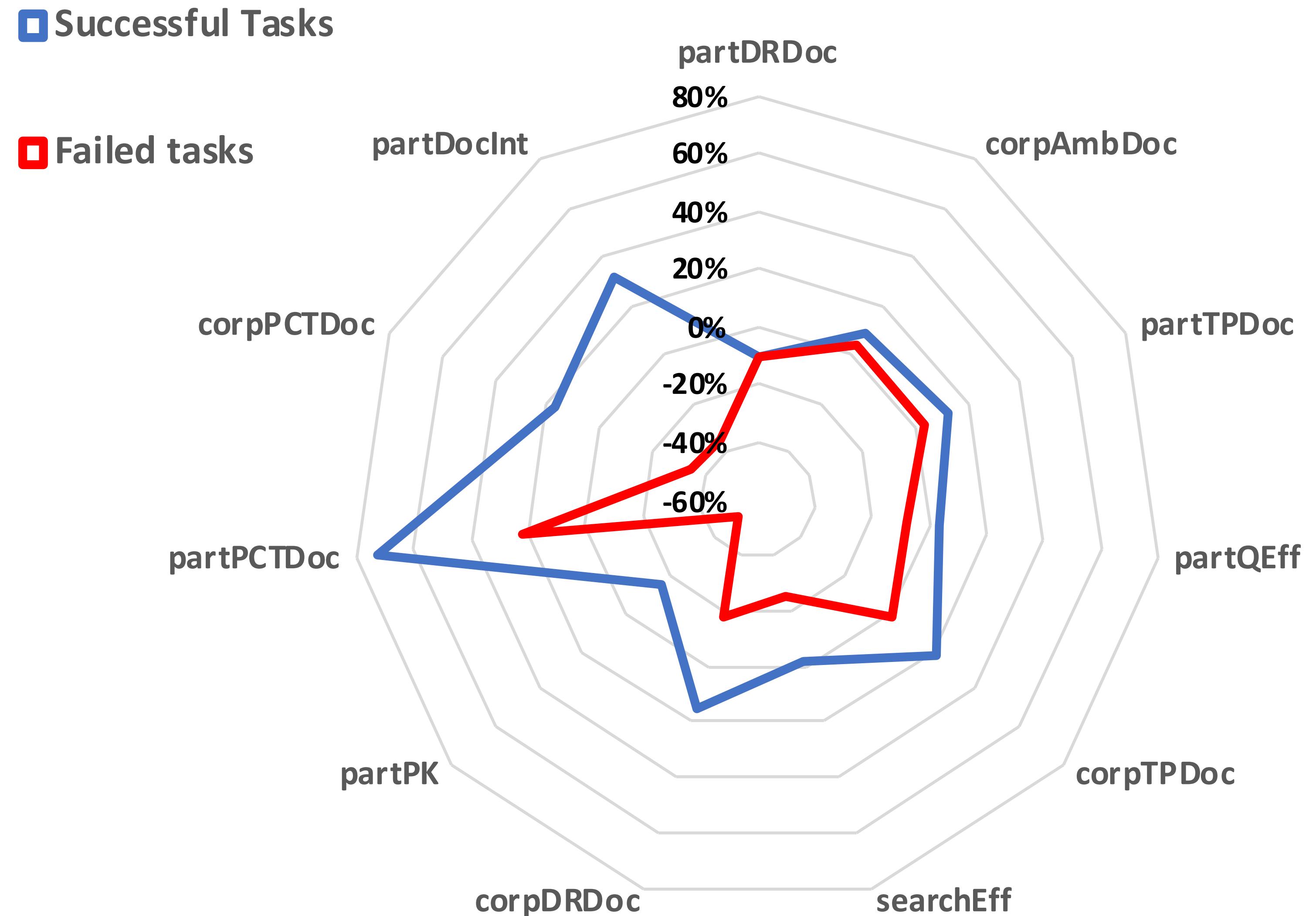
The Success Factors

Each factor has 3 grades

- Green: Above normal (+1SD)
- Amber: normal
- Red: below normal (-1SD)

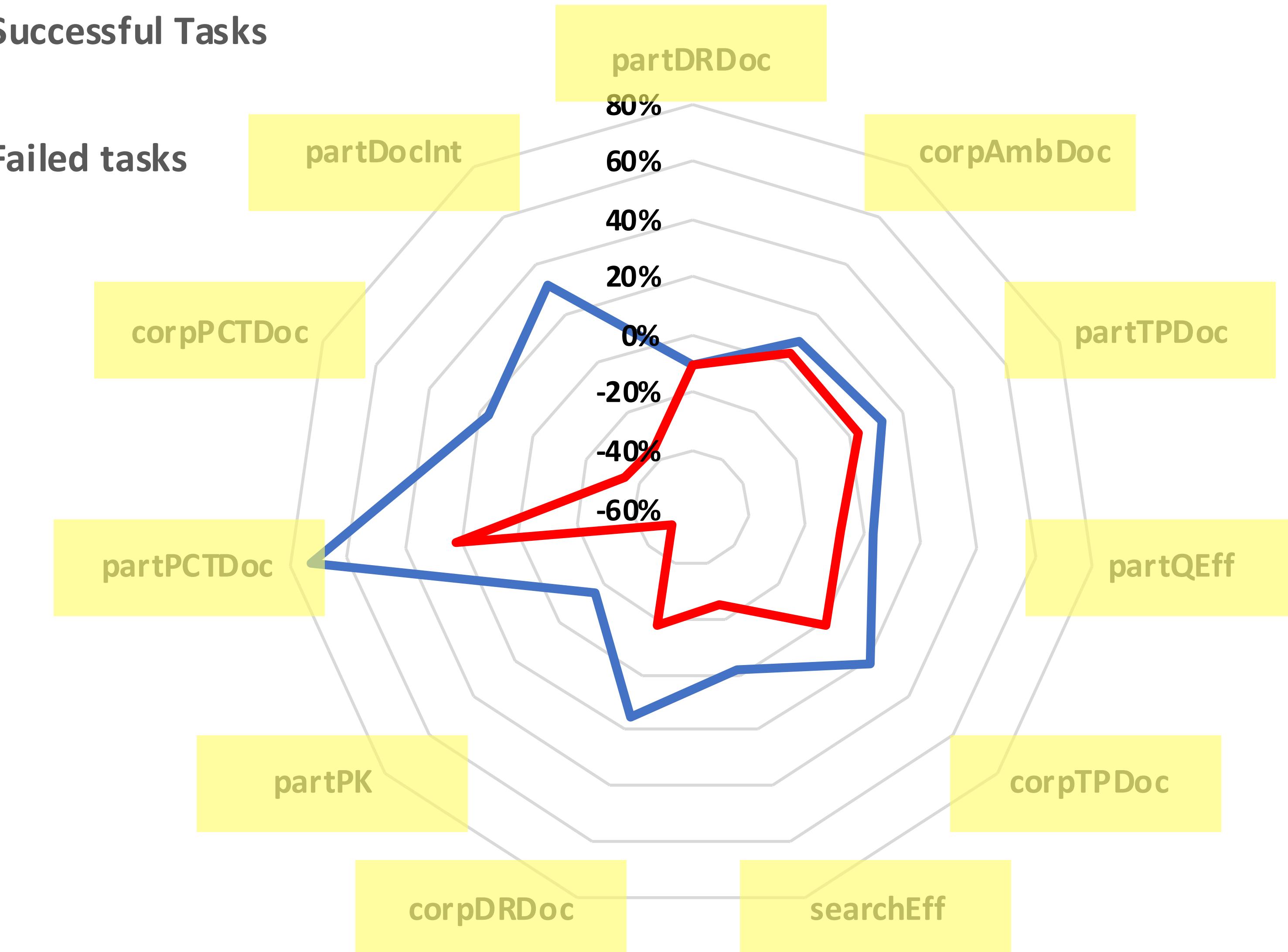


Which Factors are most important?



Which Factors are most important?

Factors are listed around the radar graph



Which Factors are most important?

All participant questions answered **correctly** post-search

All participant questions answered **incorrectly** post-search

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

partPK

corpDRDoc

partDRDoc

partDocInt

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

Failed tasks

corpPCTDoc

partPCTDoc

80%

60%

40%

20%

-20%

-40%

-60%

corpAmbDoc

partTPDoc

partQEeff

corpTPDoc

searchEff

Successful Tasks

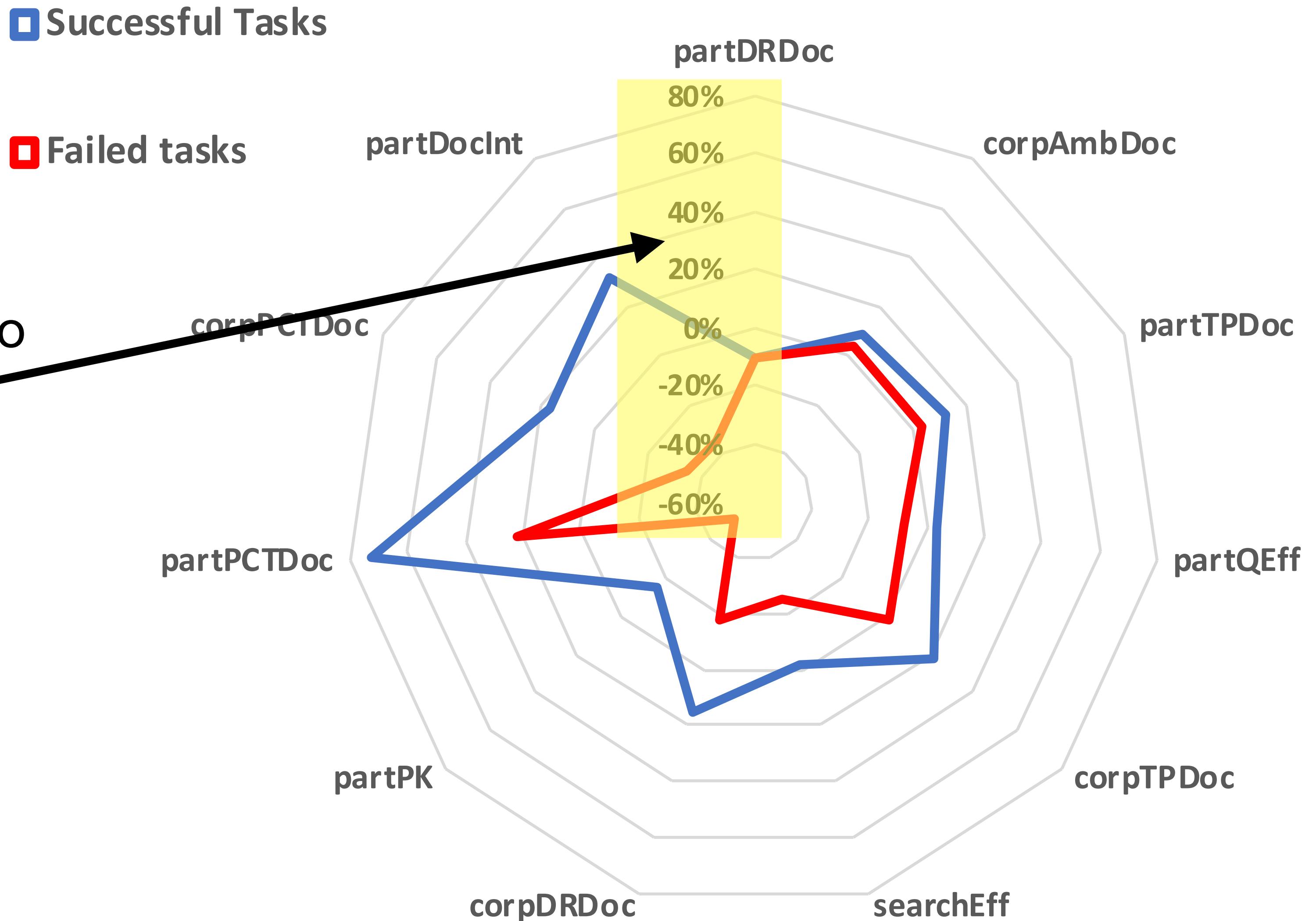
Failed tasks

corpPCTDoc

partPCTDoc</

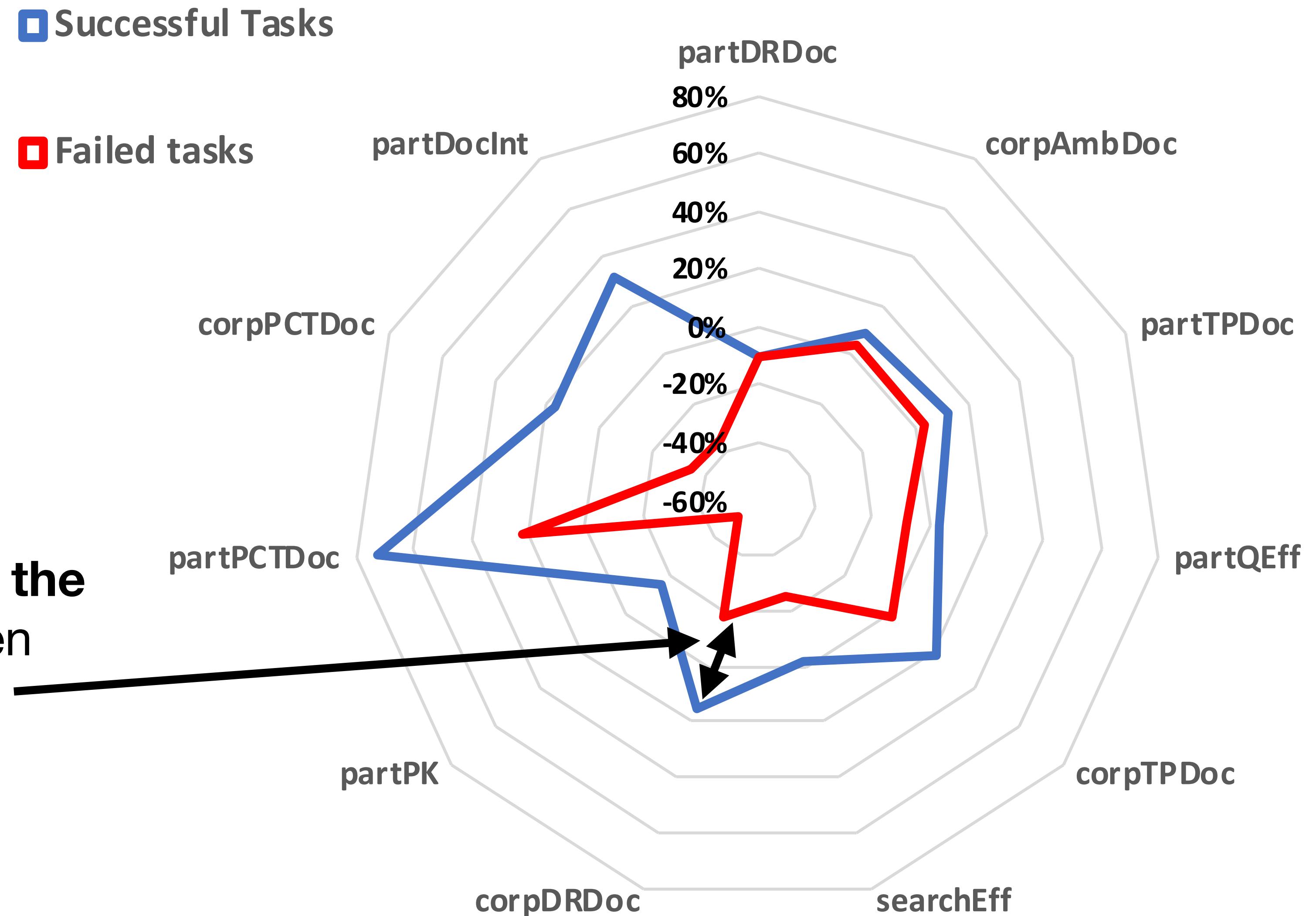
Which Factors are most important?

The scale:
Ratio of Green graded factors to
Red graded factors. The **more**
positive, the more green
graded tasks in that cohort for
that factor



Which Factors are most important?

The **bigger the gap, the greater the difference in that factor between success and failure**

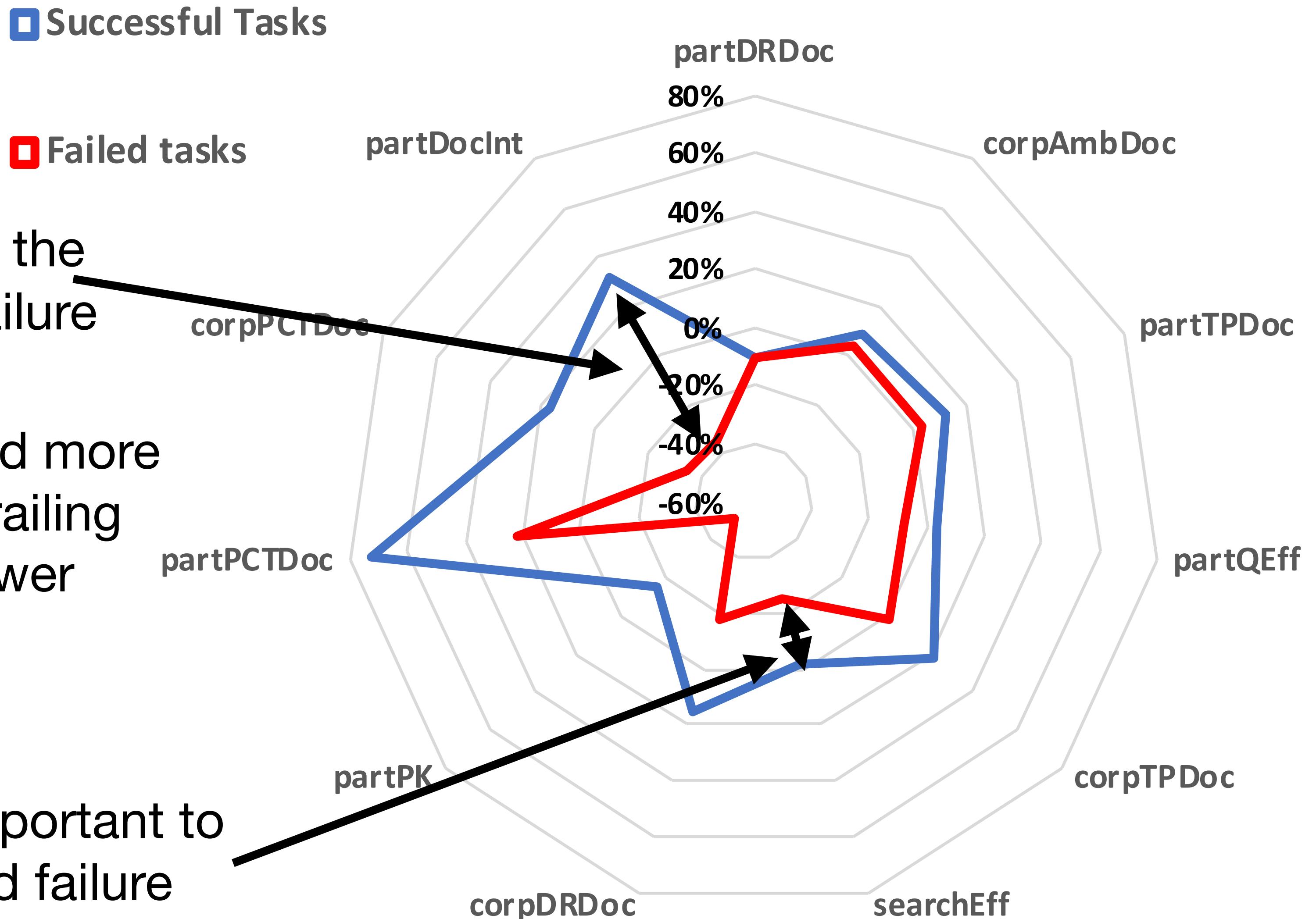


Which Factors are most important?

Poor document **interpretation** is the greatest factor accounting for failure

Document Interpretation = Read more turning point documents that derailing documents, but get a wrong answer

Search Effectiveness not so important to distinguish between success and failure



More General Implications for Search Engine Research (Design, Evaluation)

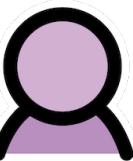
From the results of this study, we have learnt that:

- The **ranking of documents** (and thus, traditional search effectiveness) is **not the only factor** affecting users decisions
 - actually, it is a weak factor
- Documents **ambiguity**, document **misinterpretation** are far **more pressing factors** that influence users decisions
- Holistic Evaluation: should consider the whole search task
- How can search engine technology aid document interpretation, resolve document ambiguity? See next for an example, in a different, but related task

Consumer Health Search (CHS)

- CHS: **people searching for health advice online.**
 - 59% of U.S. adults has searched online for health information [Fox & Duggan, 2013].
 - 7% of Google searches are health related ~ 70,000 health queries/minute [Murphy, 2019].
- Search results strongly **bias people's health decisions** [Pogacar, 2017; White, 2009].
- People **struggle to understand** health search results and **fail to make correct health decisions** [Alpay, 2009; Kobayashi & Ishizaki, 2019].
- 59% of self-diagnosers decided **NOT to confirm their condition with a health professional** [Fox & Duggan, 2013].

Search Activities



User Activities



Information need:

1. What condition do I have?
2. What to do?

1. Formulating queries

swollen tender pink eye

Re-formulating queries

2A. Understanding Search Results Snippets

5 home remedies for pinkeye - Medical News Today
<https://www.medicalnewstoday.com/articles/324487.php> ▾
Feb 19, 2019 - Pinkeye is an eye condition that is usually temporary and the result of a bacterial infection, viral infection, or allergic reaction. ... Pinkeye, which doctors call conjunctivitis, usually resolves within 2 weeks without treatment. ... However, antibiotics will only help with pinkeye ...
Types of pinkeye · Cool compress · Damp cloth clean · Avoid touching

Pink eye (conjunctivitis) - Symptoms and causes - Mayo Clinic
<https://www.mayoclinic.org/diseases-conditions/pink-eye/symptoms.../syc-20376355> ▾
Jan 3, 2019 - Pink eye (conjunctivitis) is an inflammation or infection of the transparent membrane (conjunctiva) that lines your eyelid and covers the white part of your eyeball. When small blood vessels in the conjunctiva become inflamed, they're more visible. This is what causes the whites of your eyes to appear reddish or pink.

Pink Eye: Usually Mild and Easy to Treat | Features | CDC
<https://www.cdc.gov/features/conjunctivitis/index.html> ▾
Pink eye is common and spreads easily. It sometimes needs medical treatment. Learn the symptoms when to seek treatment, and how to help prevent it.

Pink Eye (Conjunctivitis): Symptoms, Causes, Treatment, Prevention
<https://www.webmd.com/Eye-Health/Reference> ▾

2B. Understanding Search Results Documents

ALL ABOUT VISION
Conditions Glasses Sunglasses Contact Lenses Vision Insurance Vision Surgery Eye Care
Home > Conditions > Eye Conditions, A-Z | En Español
How to avoid swollen eyelids
By Amee Rodriguez; reviewed by Gary Heiting, OD
Symptoms Causes
WebMD
Eye Health > Slideshows > Pinkeye (Conjunctivitis) Slideshow: Causes, Symptoms, & Treatments
CHECK YOUR SYMPTOMS FIND A DOCTOR FIND A DENTIST FIND LOWEST DRUG PRICES
HEALTH A-Z DRUGS & SUPPLEMENTS LIVING HEALTHY FAMILY & PREGNANCY NEWS & EXPERTS
Pinkeye (Conjunctivitis) Slideshow: Causes, Symptoms, & Treatments
What Is Pinkeye? Mayo Clinic
Patient Care & Health Info Departments & Centers Research Education For Medical Professionals
Patient Care & Health Information > Diseases & Conditions
Pink eye (conjunctivitis)
Symptoms & causes Diagnosis & treatment
Overview
Pink eye (conjunctivitis) is an inflammation or infection of the transparent membrane (conjunctiva) that lines your eyelid and covers the white part of your eyeball. When small blood vessels in the conjunctiva become inflamed, they're more visible. This is what causes the whites of your eyes to appear reddish or pink.
Pink eye is commonly caused by a bacterial or viral infection, an allergic reaction, or — in babies — an incompletely opened tear duct.
Though pink eye can be irritating, it rarely affects your vision. Treatments can help ease the discomfort of pink eye. Because pink eye can be contagious, early diagnosis and treatment can help limit its spread.
Symptoms
The most common pink eye symptoms include:

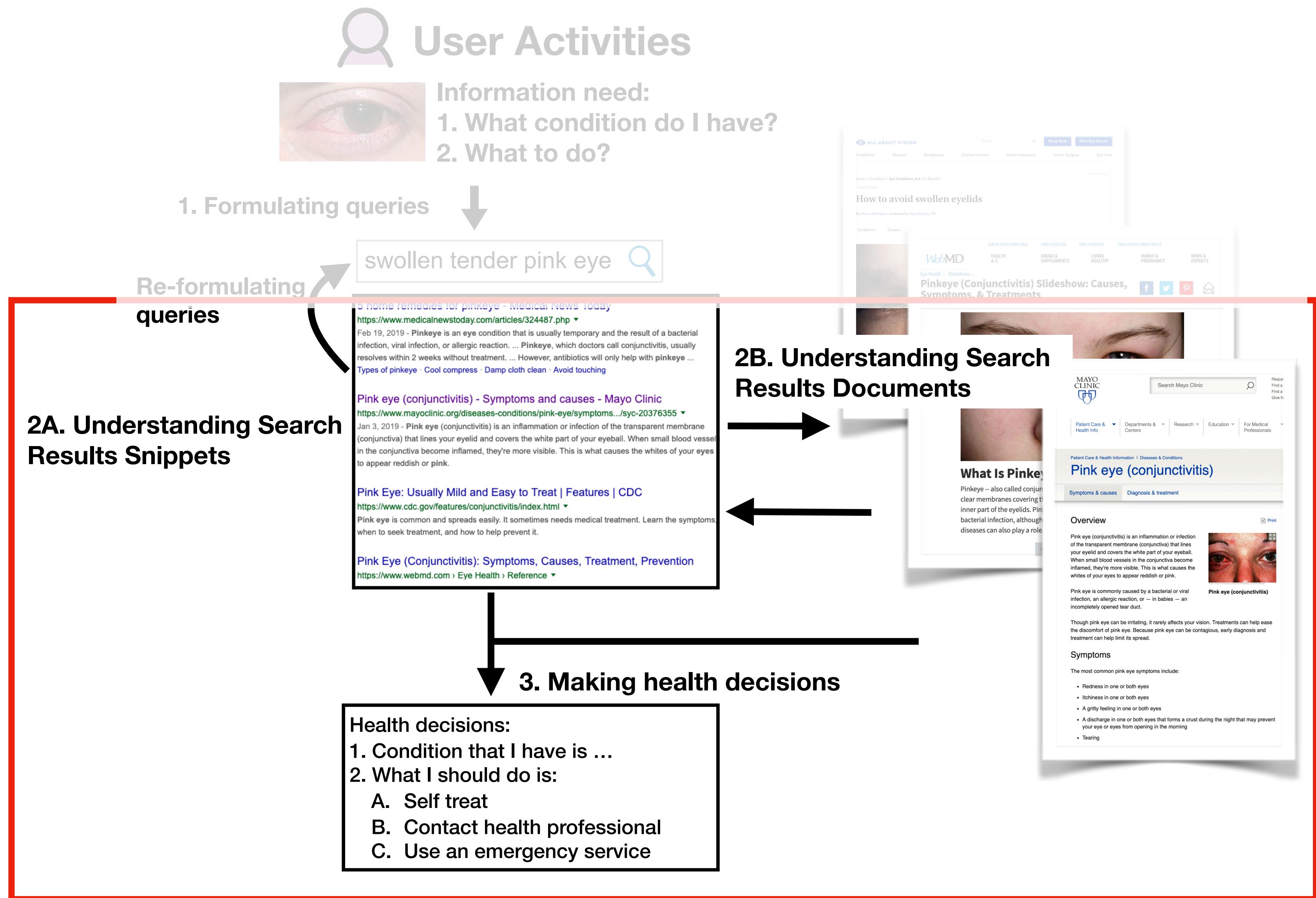
- Redness in one or both eyes
- Itching in one or both eyes
- A gritty feeling in one or both eyes
- A discharge in one or both eyes that forms a crust during the night that may prevent your eye or eyes from opening in the morning
- Tearing

3. Making health decisions

Health decisions:

1. Condition that I have is ...
2. What I should do is:
 - A. Self treat
 - B. Contact health professional
 - C. Use an emergency service

Search Activities



- Ranking and content of **snippets heavily bias people** [White et al., 2009; Dumais et al., 2010]
- People often **misunderstand health search results snippets** [White et al., 2009], e.g. tend to consider the ranking of search results as the likelihood of health condition
- Search results are from **different sources with varying quality**
- Health web pages are **difficult to read** [D'Alessandro et al., 2001; Becker, 2004].
- And contain **medical terms** that are **difficult and intimidating** for laypeople [Kindig et al., 2004; Zeng et al., 2005; Kobayashi & Ishizaki, 2019].

Assist People Appraising Search Results

- Entity cards have been used by commercial search engines to present **coherent, easy to understand and trustworthy information** (Radlinski, 2015)
- In current search engines, the appearance of a health card is triggered by queries that contain **health condition name or its aliases**
- We investigate the benefits of health cards for broader search tasks - beyond just find out more information about a specific health condition (e.g. symptom-based query)

child stomach ache vomiting diarrhea temperature

About 1,750,000 results

Why Is My Child Throwing Up With No Fever? Nausea and ...
<https://www.webmd.com/children/guide/child-throw-up-no-fever>
Usually, your child will also have nausea, watery diarrhea, and stomach pain. It's possible for food poisoning to cause fever, but it's common for it to cause throwing up with no fever, too.

Diarrhea, Fever, Nausea or vomiting and Stomach cramps ...
<https://symptomchecker.webmd.com/multiple-symptoms?symptoms=diarrhea%7Cfever%7Cnausea-or-vomiting%7C...>
Diarrhea, Fever, Nausea or vomiting and Stomach cramps. WebMD Symptom Checker helps you find the most common medical conditions indicated by the symptoms diarrhea, fever, nausea or vomiting and stomach cramps including Gastroenteritis, Food poisoning, and Irritable bowel syndrome.

What are the Causes of Stomach Ache and Fever?
<https://www.wisegeek.com/what-are-the-causes-of-stomach-ache-and-fever.htm>
When a person has a stomach ache and fever, one of the most likely causes of his symptoms is a stomach flu. Despite the misleading name, this condition is not related to influenza, but is caused by a virus that leads to such symptoms as pain, fever, diarrhea, nausea, and vomiting.

Viral Gastroenteritis (Stomach Flu): Symptoms and Treatment
<https://www.healthline.com/health/viral-gastroenteritis>
Symptoms can last from 1 to 10 days, including diarrhea, vomiting, and fever. Here are some self-care tips and home remedies for relief. Viral gastroenteritis, also known as the stomach flu, is an ...

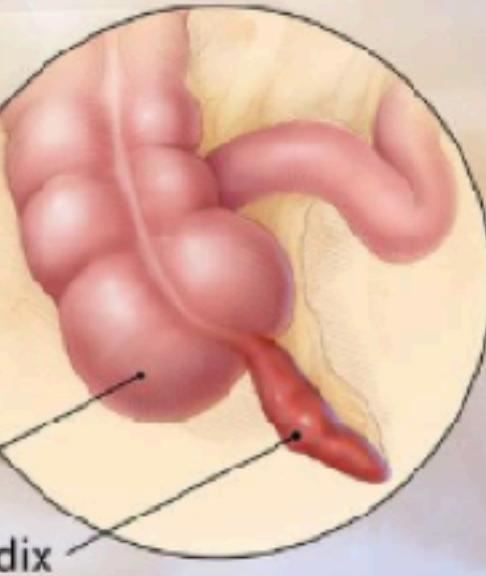
Norovirus | Symptoms | CDC
<https://www.cdc.gov/norovirus/about/symptoms.htm>
It's probably norovirus, a common virus that is not related to the flu. Norovirus is the most common cause of foodborne diarrhea and vomiting. Norovirus causes inflammation of the stomach or intestines. This is called acute gastroenteritis. A person usually develops symptoms 12 to 48 hours after ...

Child Vomiting | Ask Dr Sears
<https://www.askdrsears.com/topics/health-concerns/childhood-illnesses/vomiting>
This is a virus that causes sudden onset of vomiting, high fever and stomach pain. Diarrhea usually begins during the first or second day. The length of vomiting varies from the "12 hour flu" to the dreaded "72 hour flu." There is no blood or stool test to diagnose this.

Diarrhea, Fever and Stomach cramps: Common Related ... – WebMD
<https://symptomchecker.webmd.com/multiple-symptoms?symptoms=diarrhea%7Cfever%7Cstomach-cramps&sympto...>
Food poisoning can cause abdominal pain, diarrhea, nausea, vomiting, fever, chills, and weakness. Traveler's diarrhea. Traveler's diarrhea causes watery diarrhea and cramps, sometimes with a low-grade fever. Diverticulosis. Diverticulosis is a condition of the colon and causes pain, cramping, infection, bleeding, and other symptoms. Drug overdose

Appendicitis

About **Symptoms** **Treatments**



Causes severe abdominal pain

Large intestine **Inflamed appendix**

A condition in which the appendix becomes inflamed and filled with pus, causing pain.

- Requires a medical diagnosis
- Lab tests or imaging often required
- Treatable by a medical professional
- Short-term: resolves within days to weeks

Appendicitis begins with pain near the belly button and then moves to the right side. This is often accompanied by nausea, vomiting, poor appetite, fever and chills.

Critical: Consult a doctor for medical advice
Sources: Mayo Clinic

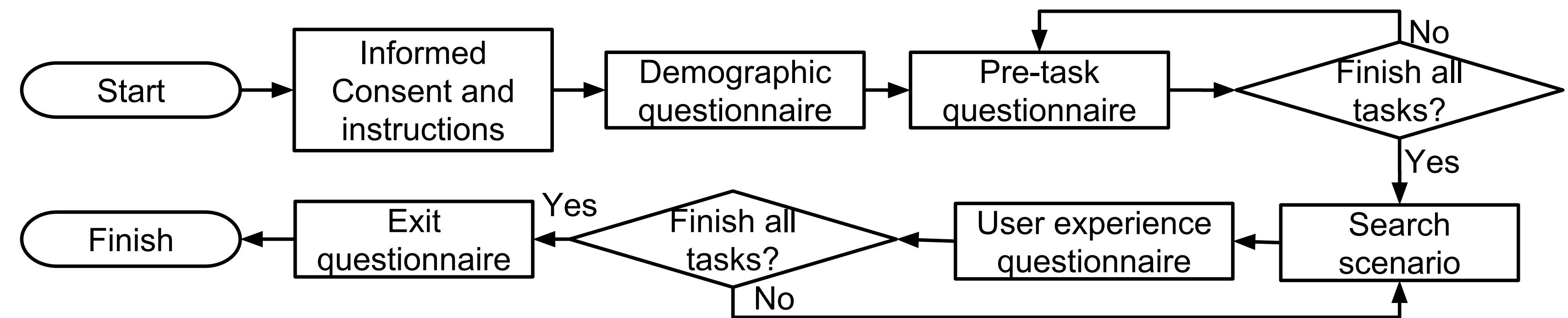
Based on: (1) Jimmy, G Zuccon, B Koopman, G Demartini, Health Cards for Consumer Health Search, SIGIR 2019
(2) Jimmy, G Zuccon, G Demartini, B Koopman, Health Cards to Assist Decision Making in Consumer Health Search, AMIA 2019
(3) Jimmy, G Zuccon, B Koopman, G Demartini, Health Card Retrieval for Consumer Health Search, CIKM 2019

Do Health Cards Help People Making Effective Health Decisions?

- In a previous study we compared a search engine with health cards VS no health cards, and we found that:
 - The use of health cards in general **reduces the effort spent, increases the good abandonment rate, and improves the user's satisfaction** while not significantly impacting other considered measurements
 - Presenting health cards also **benefits the not knowledgeable searchers** to identify the correct health information
- What I am describing next is a study where we focus on **Health Decisions** with respect to **self-diagnosis** (what condition and what to do)
- And we consider settings where health cards are used, but to varying degree (Single VS. Multiple Cards, Correct VS Incorrect Cards)

Methodology: User Study

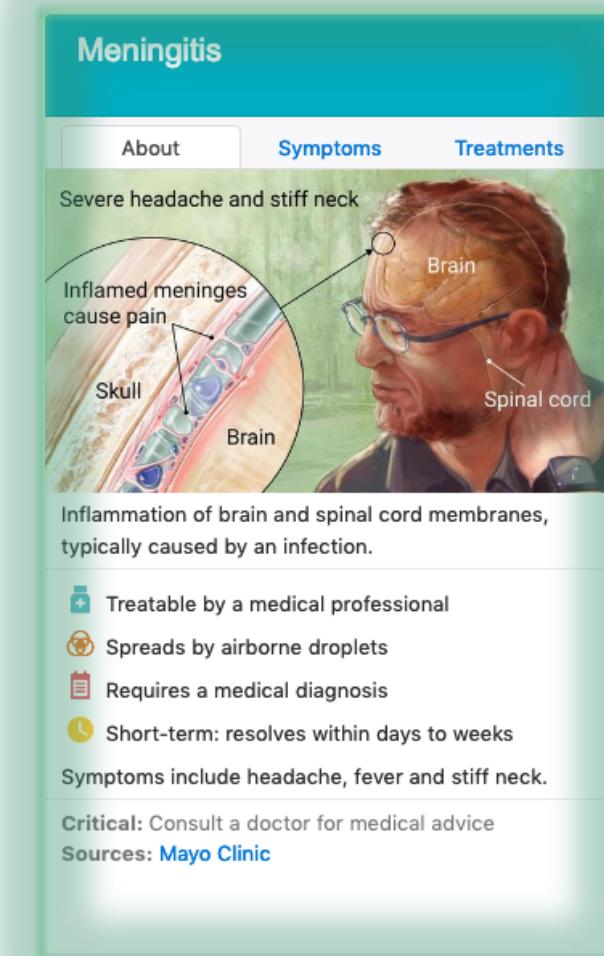
- 8 health search tasks X 64 Participants = 512 Data Points.
- Graeco-latin square design to avoid order effects.
- Types of user interface:
 - ❖ Single VS. Multi-Cards
 - ❖ Correct VS Incorrect Cards



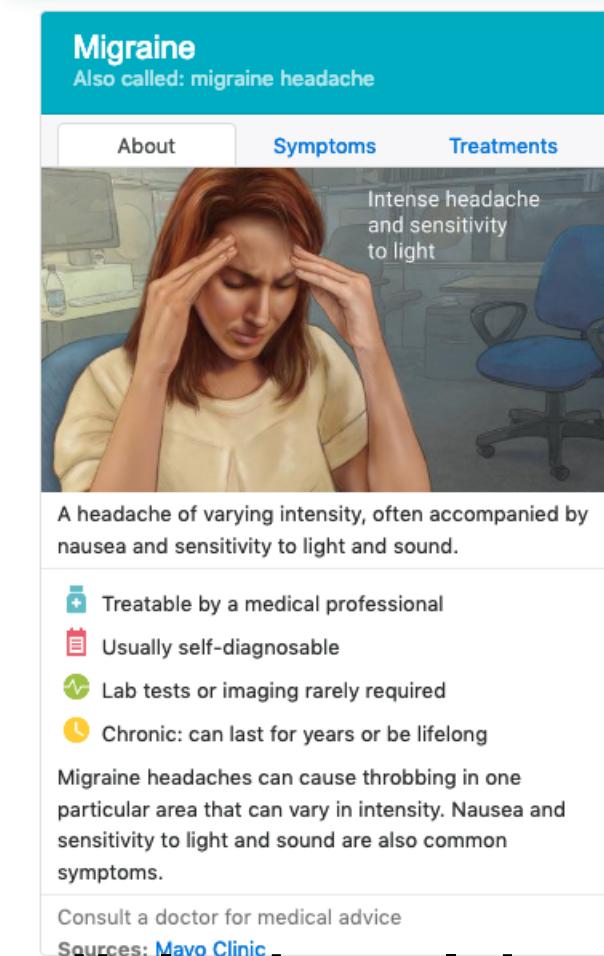
The Four Search Interfaces

Single-Card

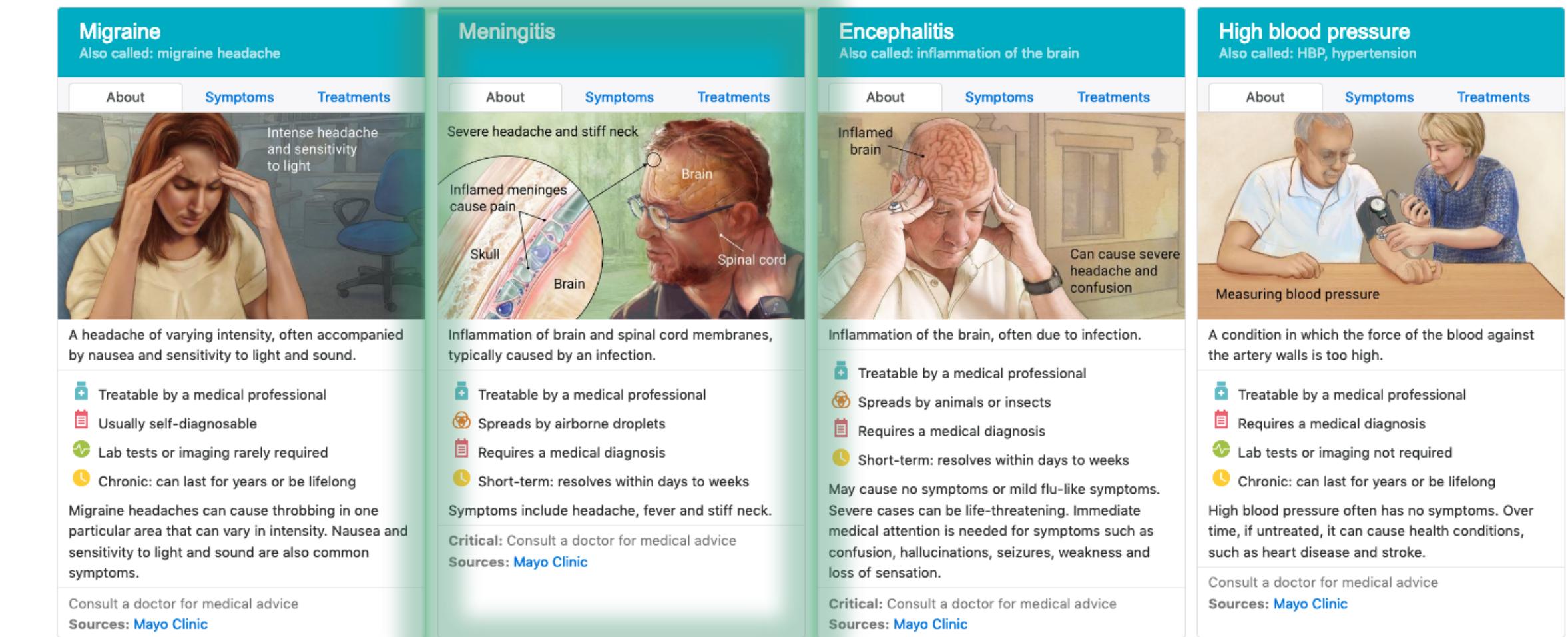
With a Correct Card



No Correct Card



Multi-Cards



Why Multi-Cards? Presenting possible health conditions side-by-side may enable people to perform differential diagnosis + increases chances to present a correct health card

Search Task

For a health scenario, make the following two decisions:

1. What is the most possible condition for the scenario?
2. What to do / level of urgency ?
 - A. Use emergency service
 - B. Contact a health professional
 - C. Self treat

Sample scenario:

Scenario: Your 12-year-old daughter had a sudden severe abdominal pain with nausea, vomiting, and diarrhea. Her body temperature is 40C.

Correct Diagnosis: Appendicitis

Correct Urgency Level:
Emergency

[Semigran et al., 2015]

Is this a difficult task?

- For users: Our previous findings show that **health card provide less support** for such explorative type of search.
- For systems: People submit **under-specified queries** [Stanton et al., 2014; Chen & Turner, 2010; Zuccon et al., 2015; Luger et al., 2014] which make determining the one correct health card difficult (if possible at all).
- For example: people search for one symptom at a time say ‘headache’ [Luger et al., 2014]. This is a symptom of many health conditions.

Measurements

1. Selected as a source of information



2. Lead to more **correct** answers



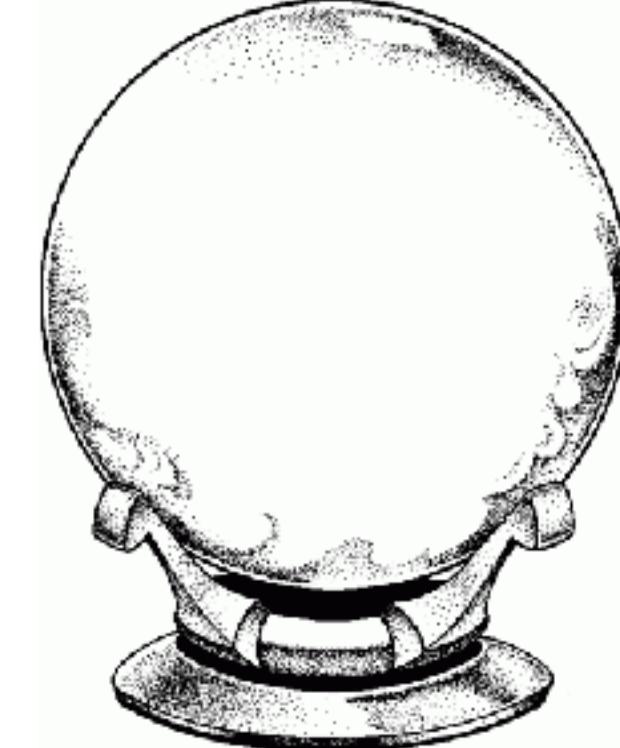
3. Reduce **time** to complete a task



4. Reduce the **effort** to complete a task



5. Increase rate of
Good abandonment*



6. Reduce the **workload** to complete a task



7. Improve **confidence** on decisions



* Good abandonment is when people manage to complete a task without clicking on the search results.

Measurements

1. Selected as a source of information



2. Lead to more **correct** answers



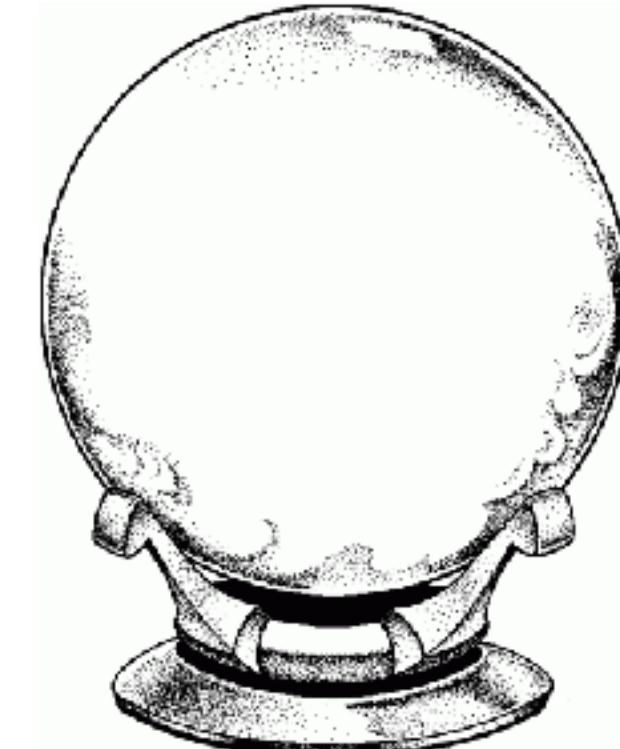
3. Reduce **time** to complete a task



4. Reduce the **effort** to complete a task



5. Increase rate of
Good abandonment*



6. Reduce the **workload** to complete a task



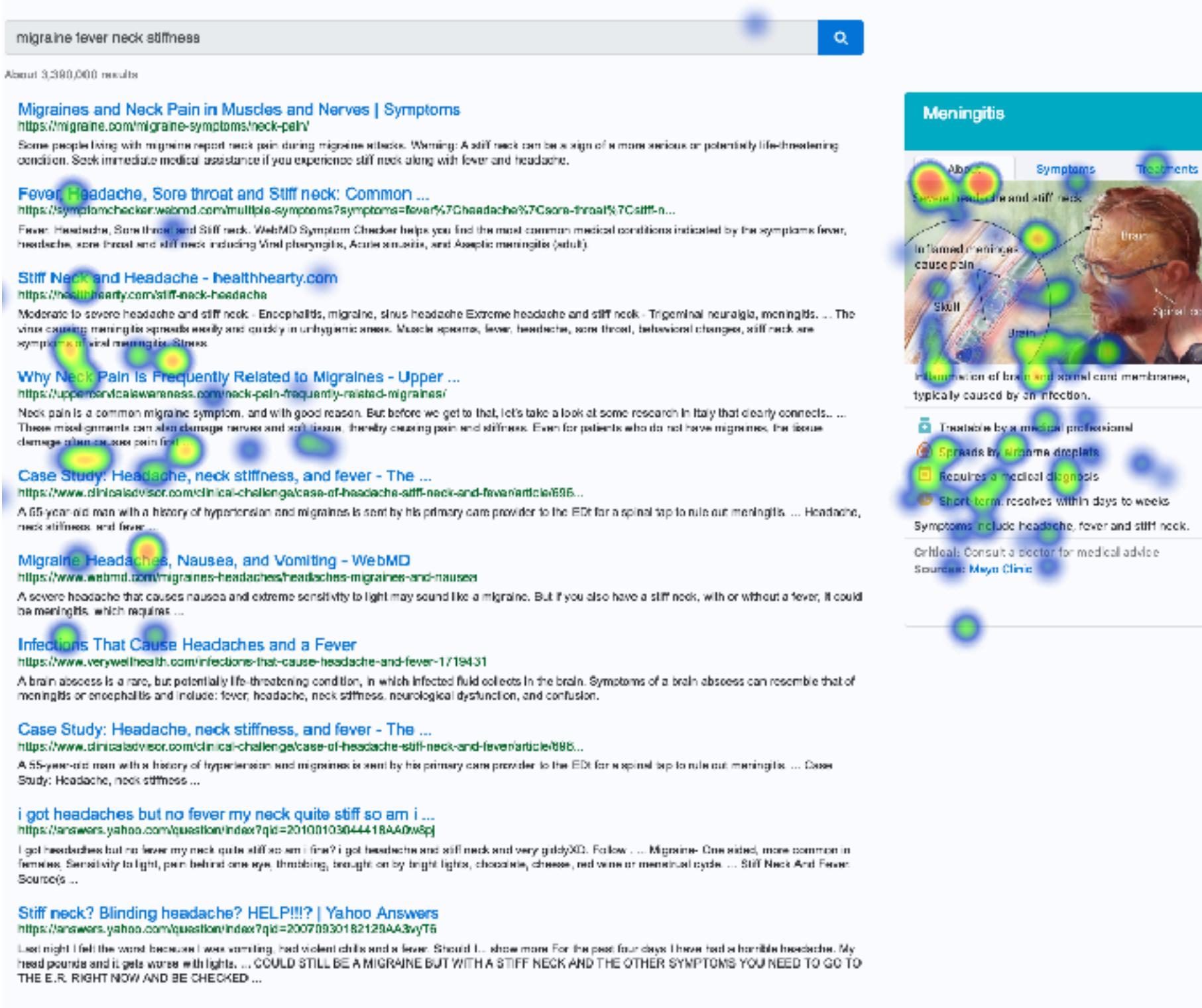
7. Improve **confidence** on decisions



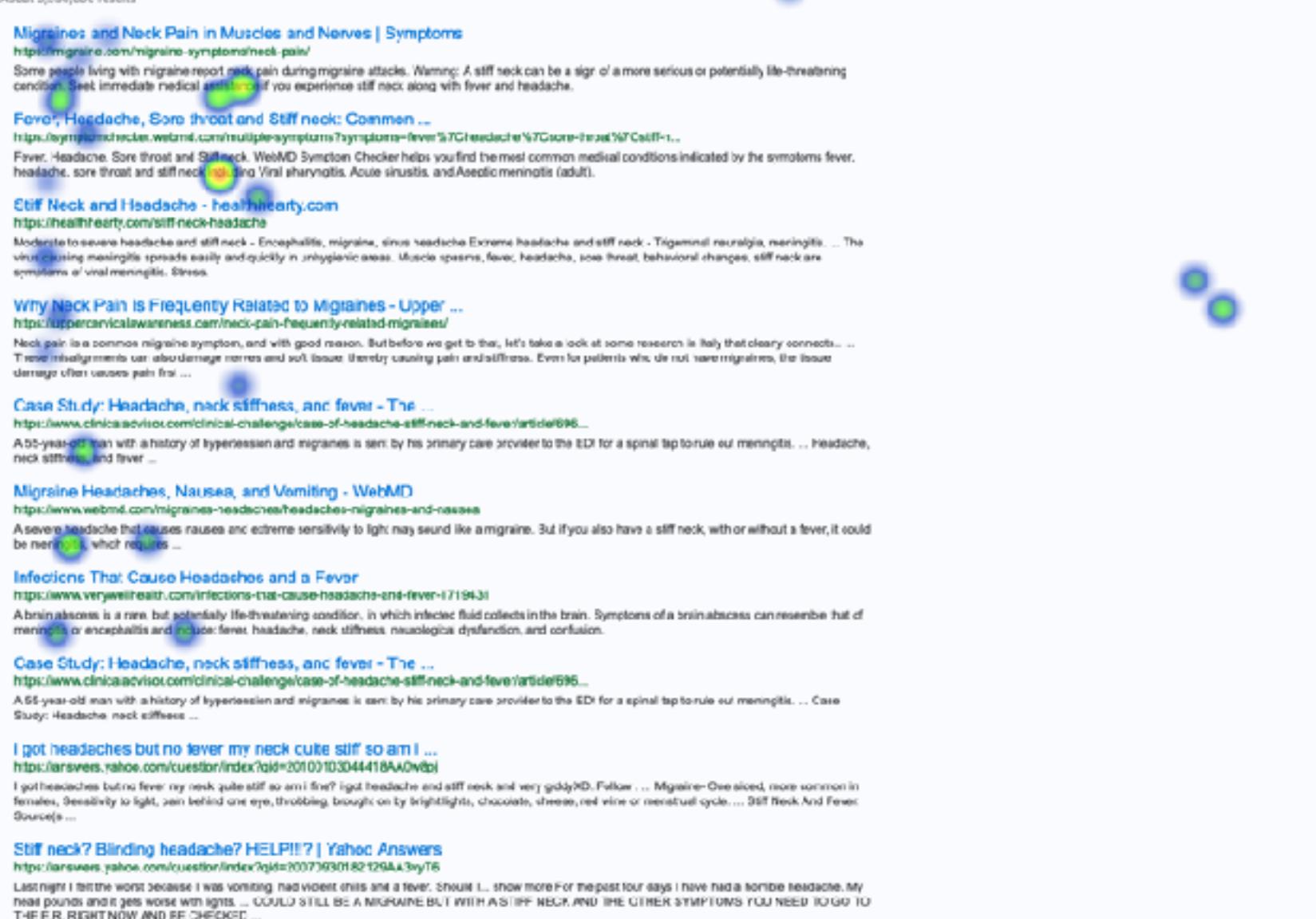
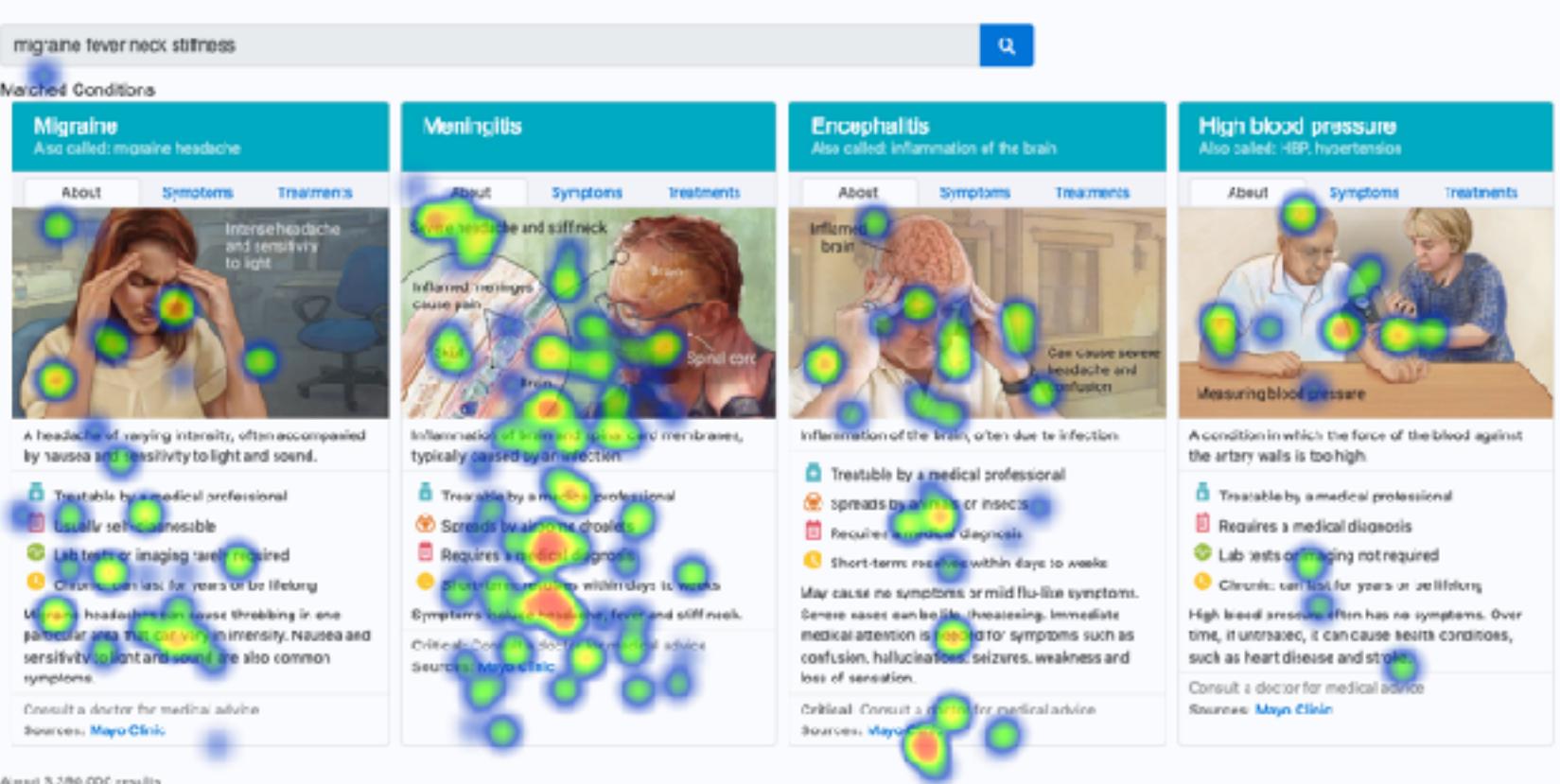
* Good abandonment is when people manage to complete a task without clicking on the search results.

Impact of Single VS. Multi-Cards on Search Behaviour

49% of time was spent to interact with the Single-Card



82% of time was spent to interact with the Multi-Cards interface.

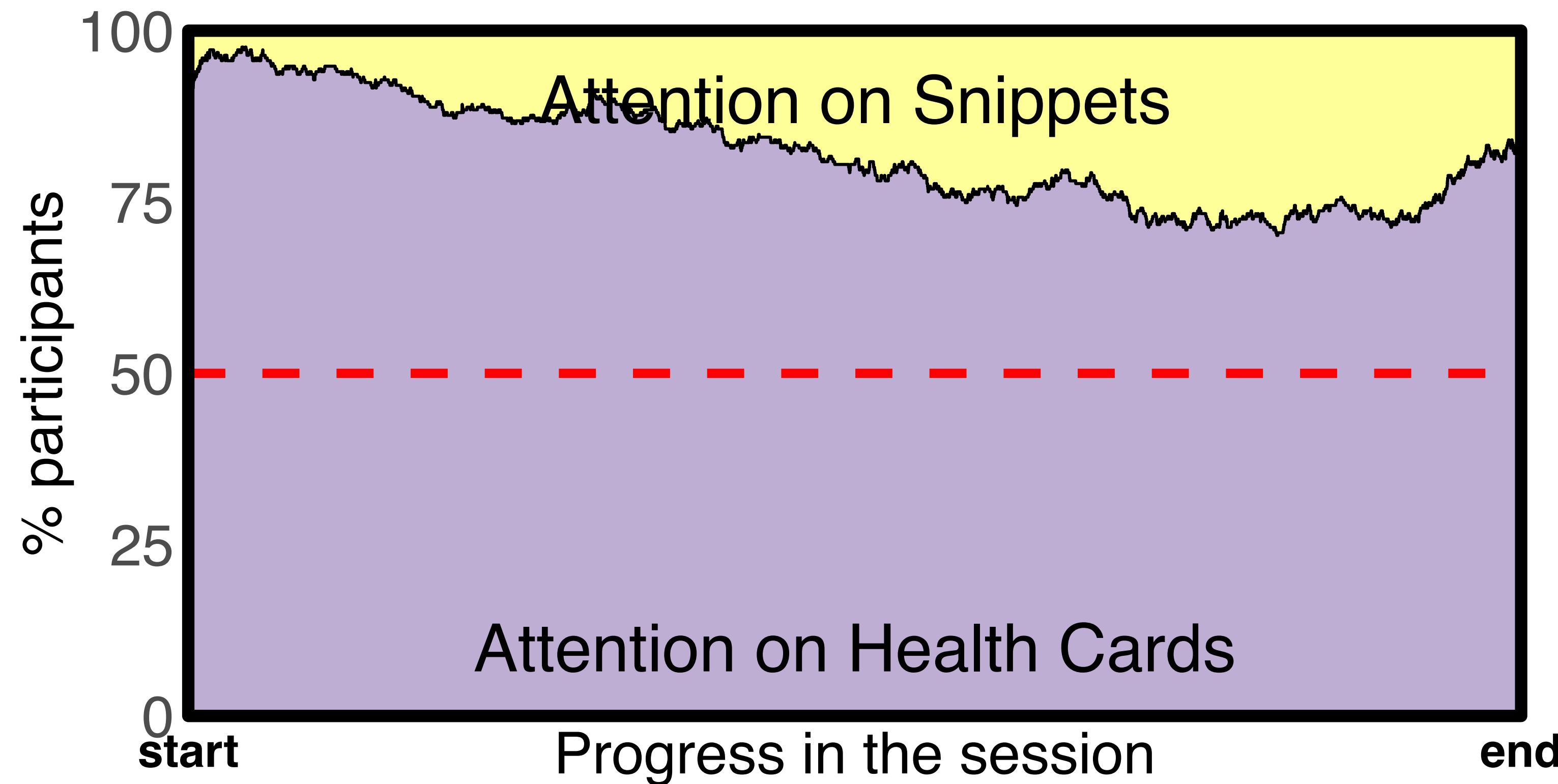


Impact of Single VS. Multi-Cards on Search Behaviour

49% of time was spent to interact with the Single Card

82% of time was spent to interact with the Multi-Cards interface

Participants tend to consider health cards earlier and at the end of a session

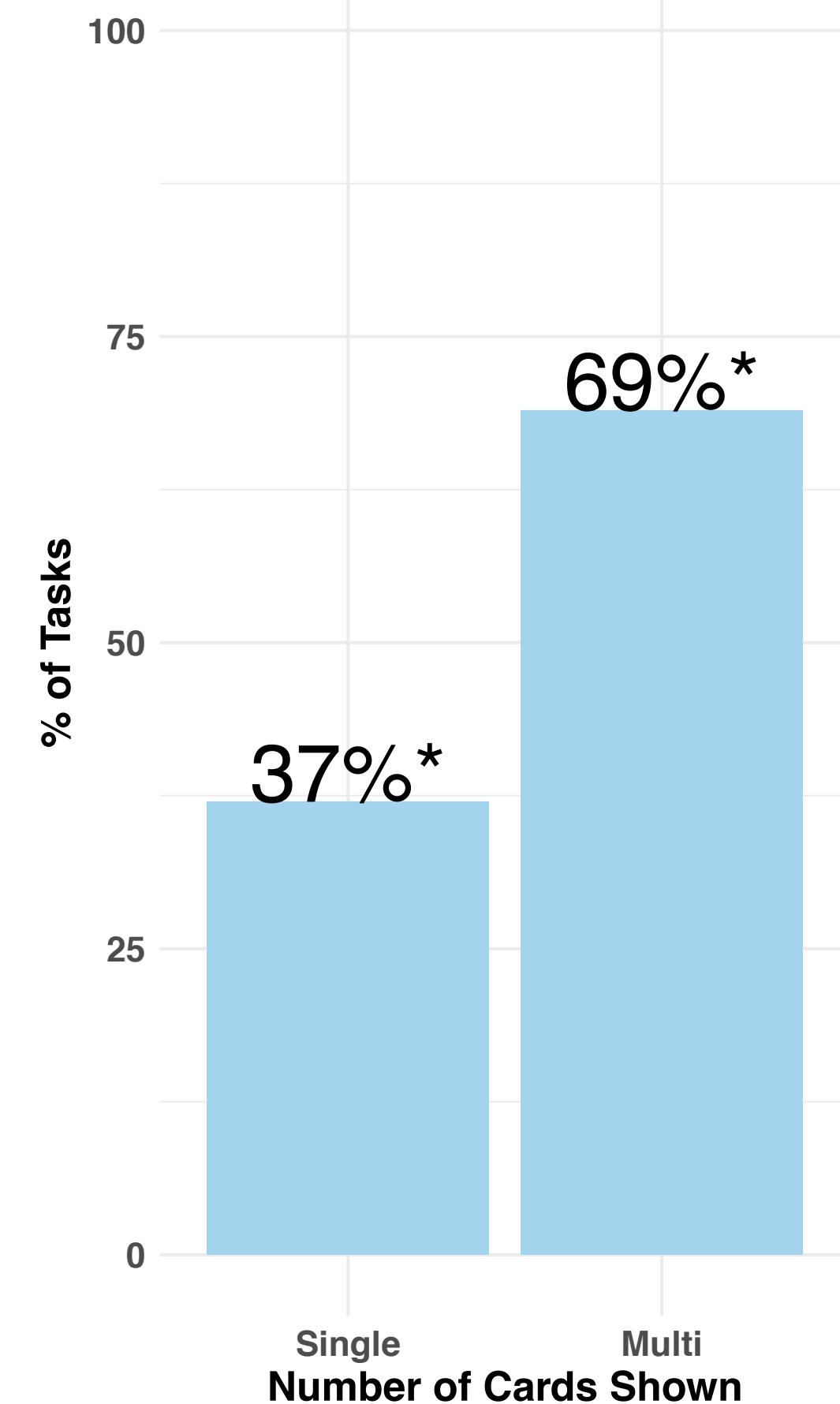


Source(s) ...
Stiff neck? Blinding headache? HELP!!? | Yahoo Answers
<https://answers.yahoo.com/question/index?qid=20070920182229AA3y7R>
Last night I felt the worst because I was vomiting, had violent cramps and a fever. Should I ... show more For the past four days I have had a terrible headache. My head pounds and it gets worse with lights ... COULD STILL BE A MIGRAINE BUT WITH A STIFF NECK AND THE OTHER SYMPTOMS YOU NEED TO GO TO THE E.R. RIGHT NOW AND BE CHECKED ...

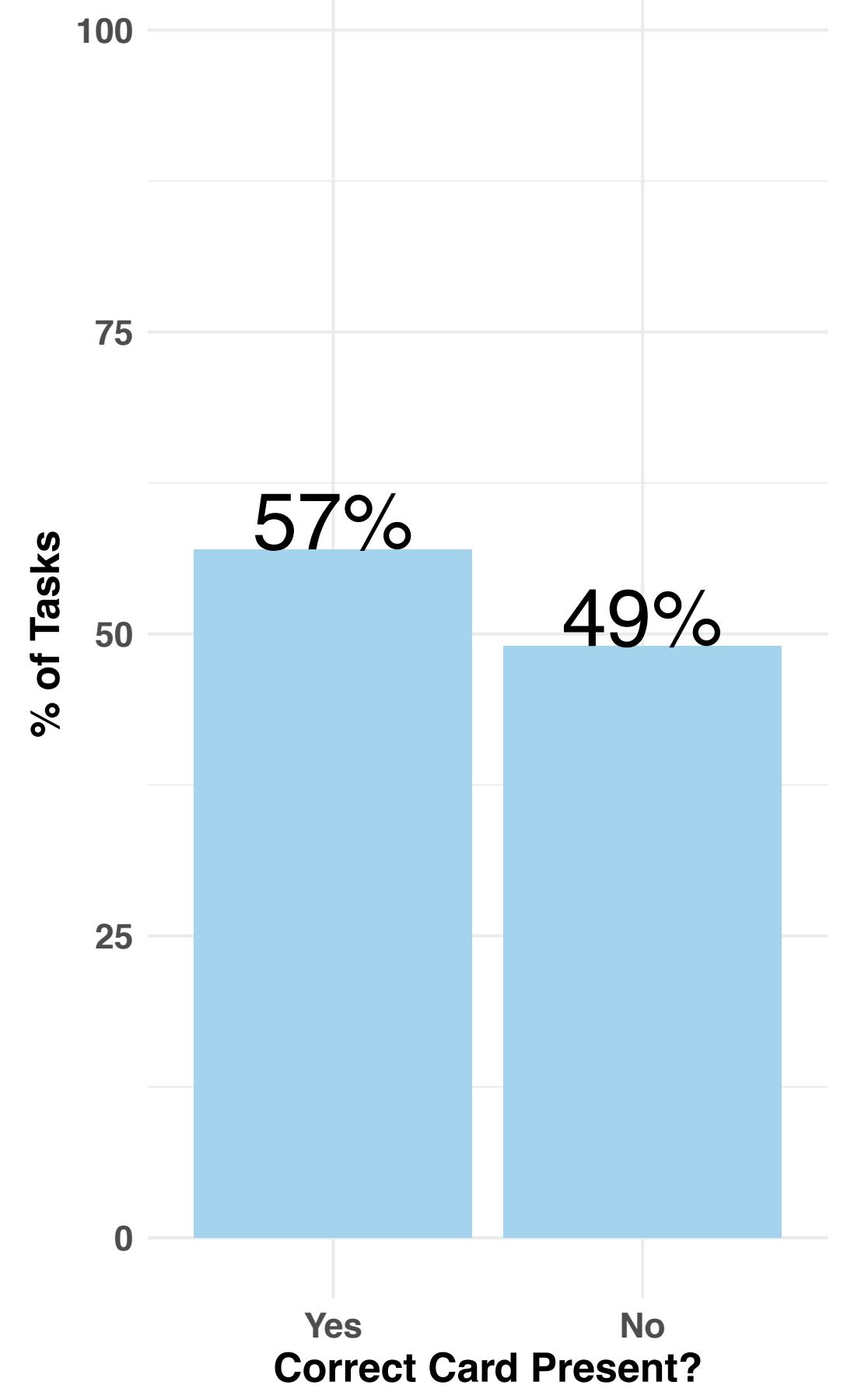
Impact of single and multi-cards?

(1) Health cards selected as a source of information

Information from **multi-cards** selected **significantly more frequently** than from single-card



Information from health cards selected **regardless of the correctness** of health card

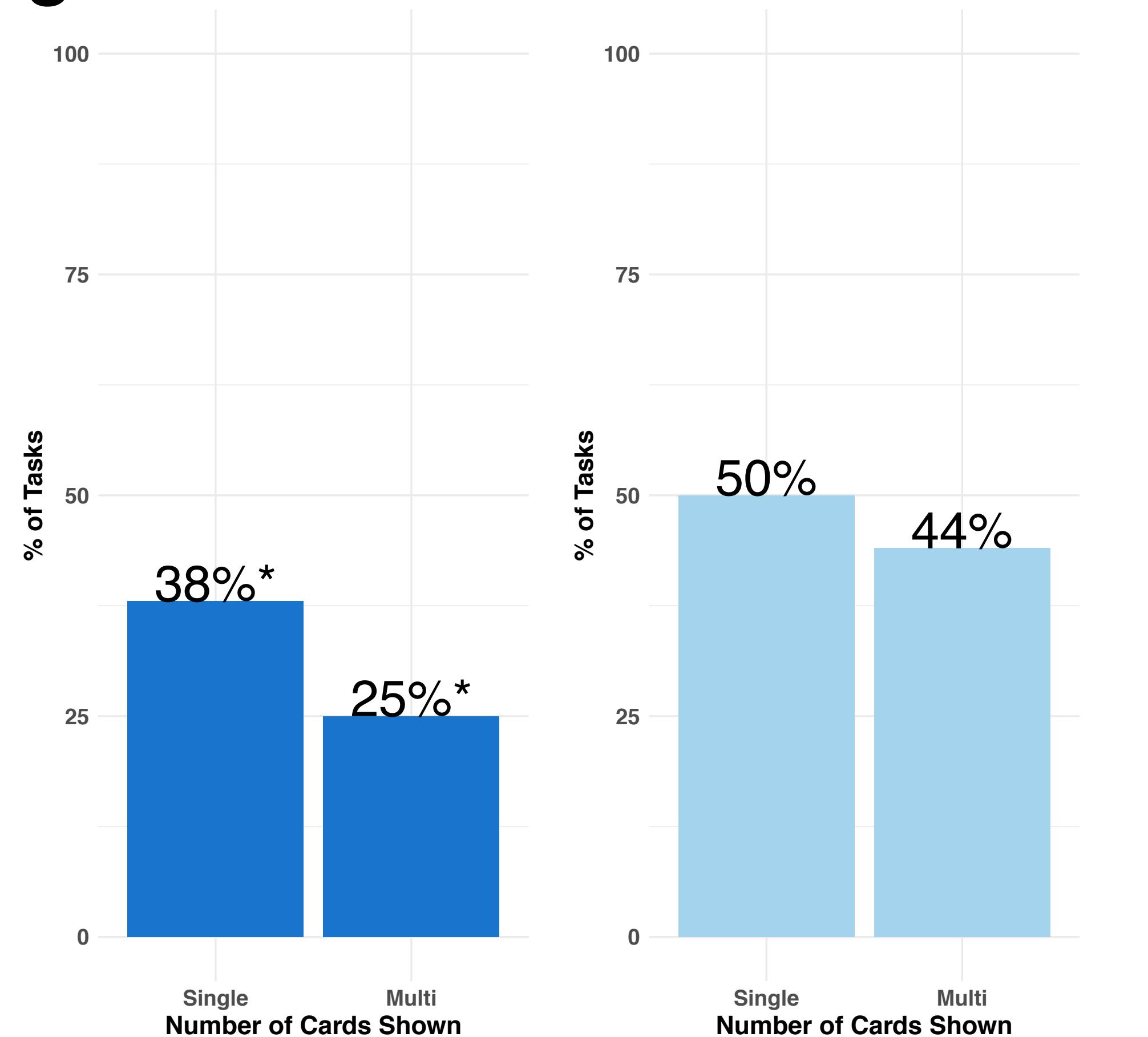




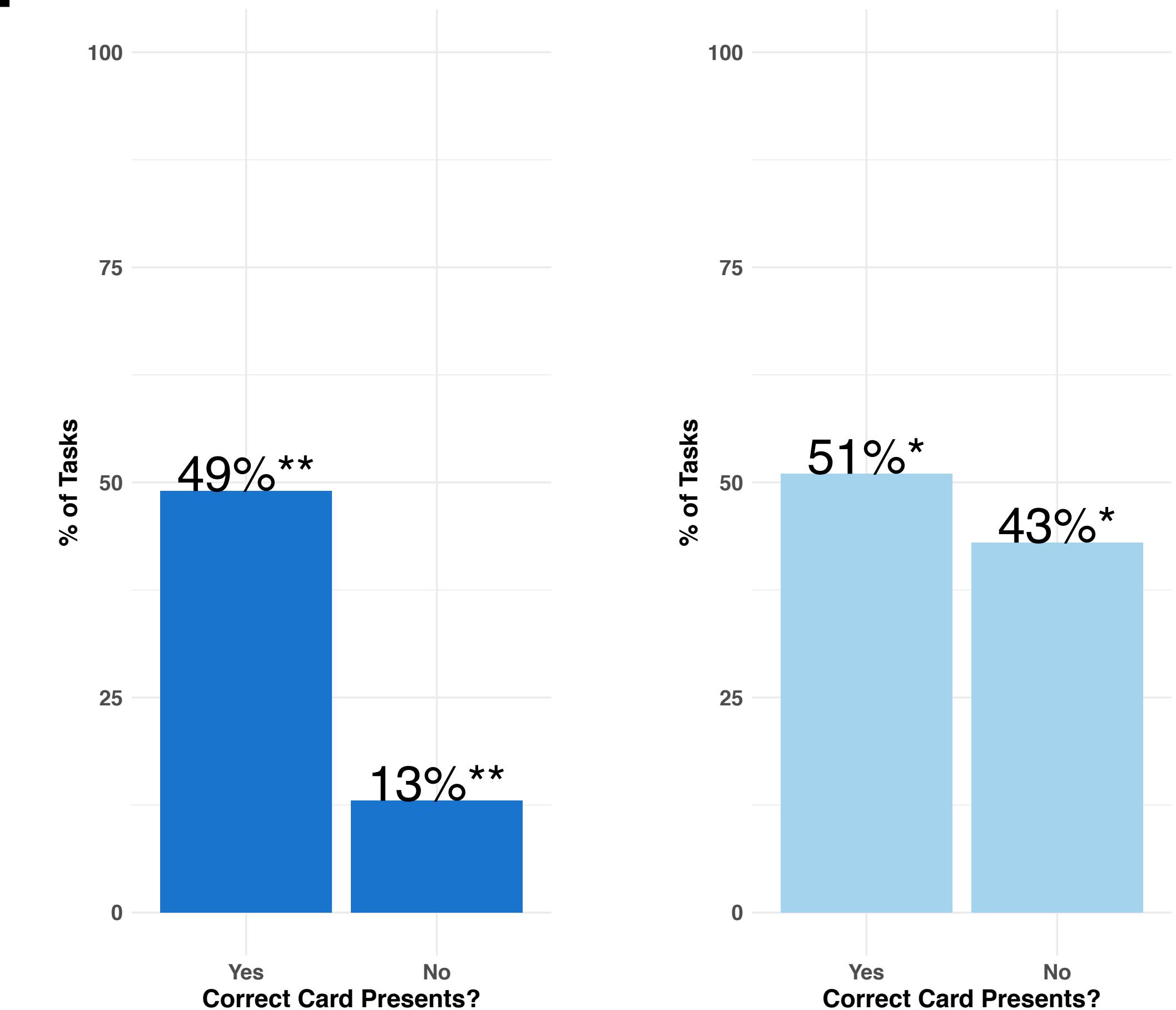
Impact of single and multi-cards?

(2) Correctness of diagnosis and level of urgency

Single card lead to significantly
higher number of correct
diagnosis than multi-cards



Correctness of **decisions** is
dependent on correctness of
presented health card





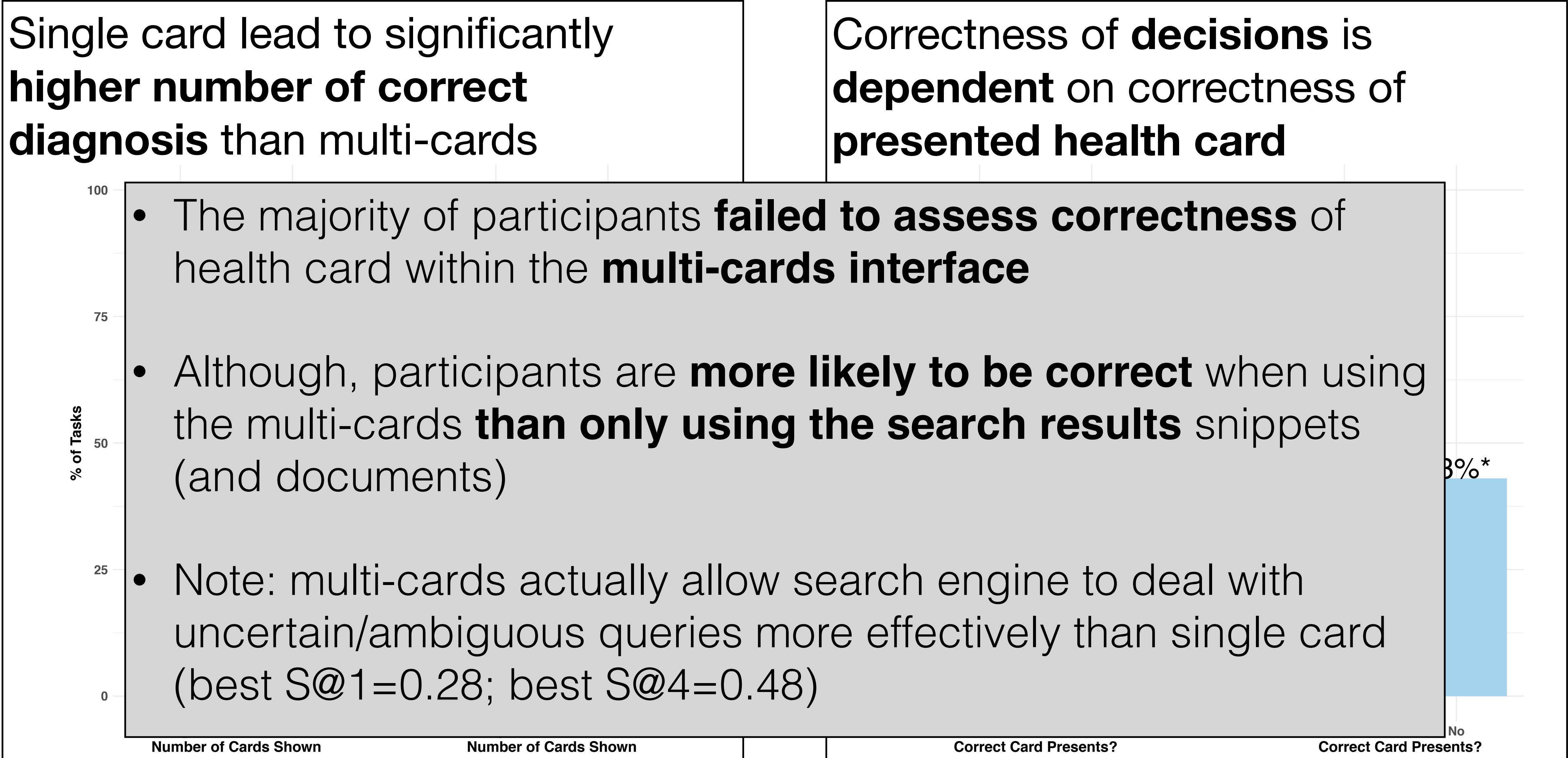
Impact of single and multi-cards?

(2) Correctness of diagnosis and level of urgency

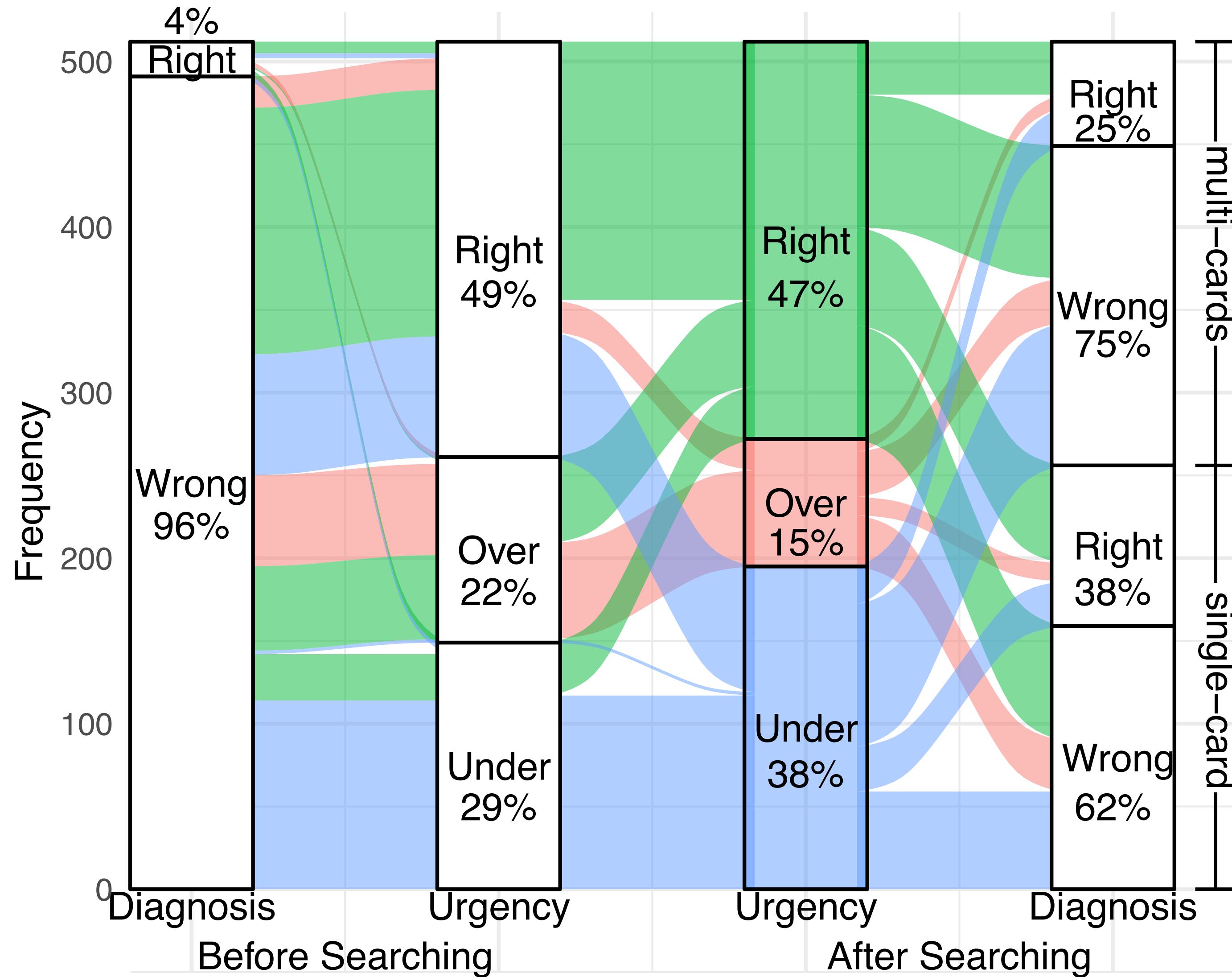
Single card lead to significantly
higher number of correct
diagnosis than multi-cards

Correctness of **decisions** is
dependent on correctness of
presented health card

- The majority of participants **failed to assess correctness** of health card within the **multi-cards interface**
- Although, participants are **more likely to be correct** when using the multi-cards **than only using the search results** snippets (and documents)
- Note: multi-cards actually allow search engine to deal with uncertain/ambiguous queries more effectively than single card (best S@1=0.28; best S@4=0.48)



How does Search Impact Health Decision Making?



- Search activities (with health cards) lead to **significant increase of correct diagnosis**
- Search activities did **NOT** lead to **better level of urgency** correctness
- **43%** of **correct diagnosis** had an **incorrect level of urgency** (73% of them were **underestimated**)

What have we learnt from this study?

- Presenting **health cards** helps people to make more **correct diagnosis**
- People **struggle to infer the level of urgency** based on search results. In fact, people tend to **underestimate** urgency across search interfaces.
- This is problematic as making the correct diagnosis is not as important as identifying the correct level of urgency

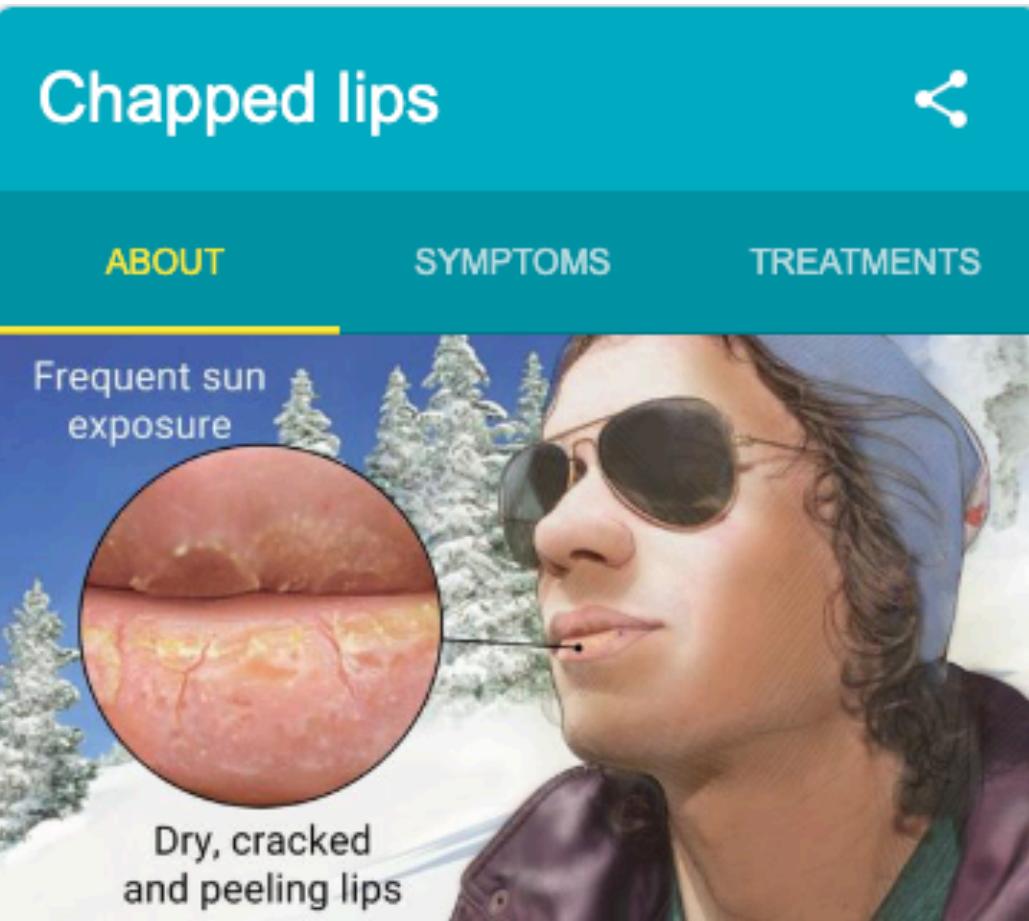
Diagnos	Urgency Level	Risk
→ Correct	Under-estimated	Harm well-being + additional health system costs
Correct	Over-estimated	Waste medical resources
Incorrect	Correct	None to minimum

- Future work needs to **better assist people in deciding the urgency level** (what to do)

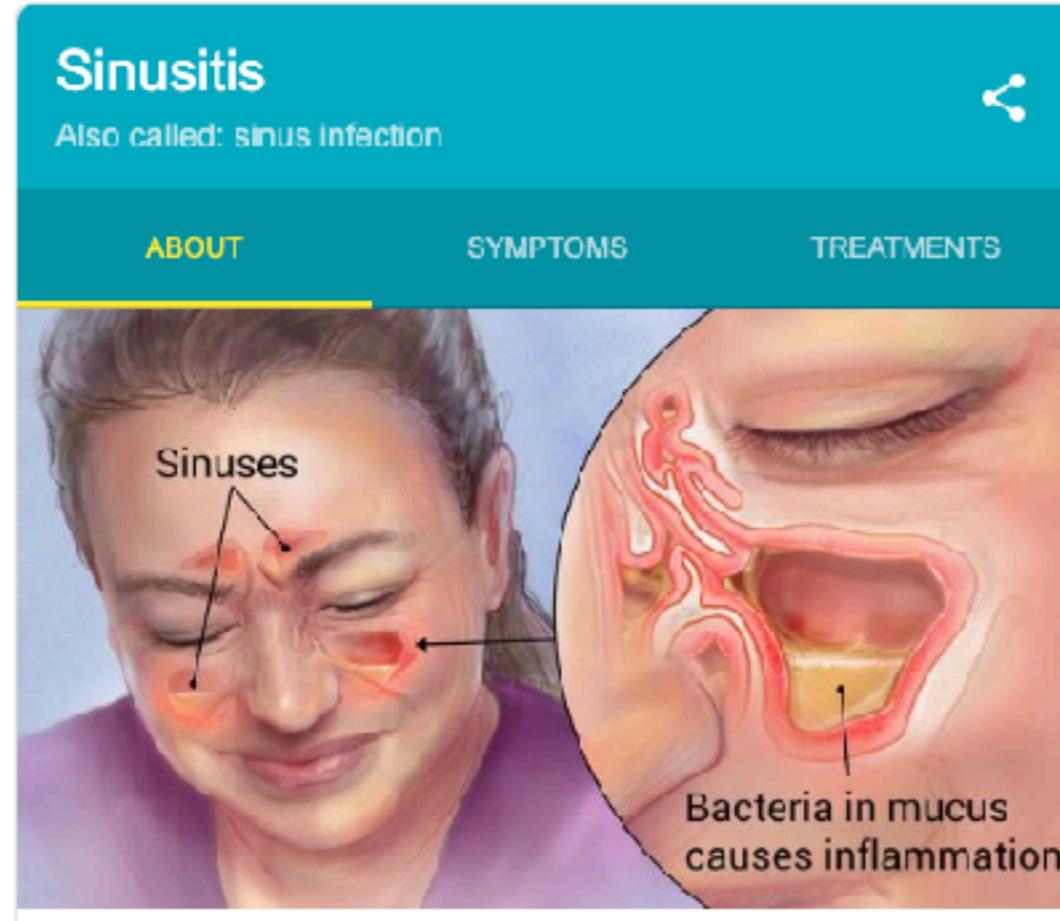
Example of an attempt to address these findings

Google Health Cards Before Our Study

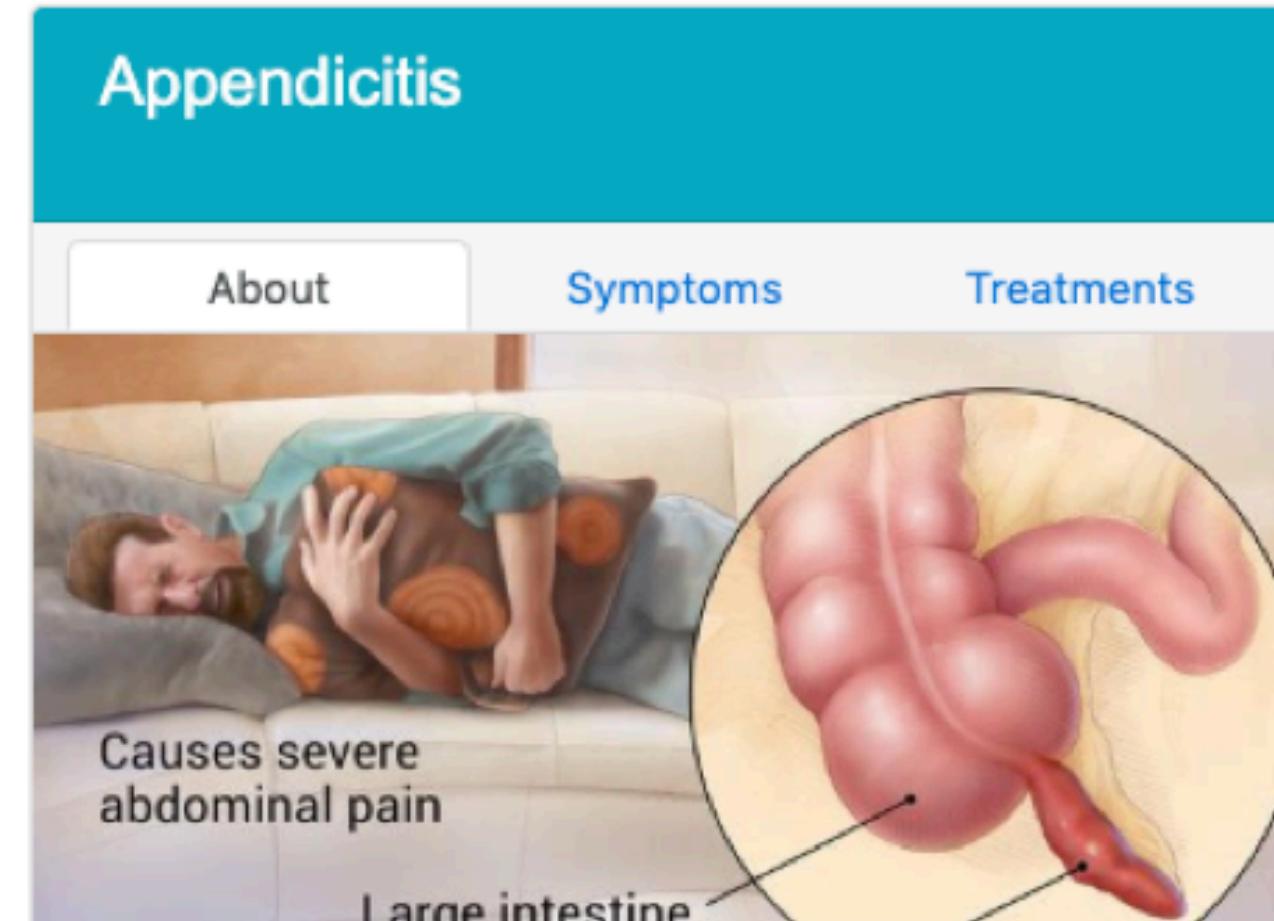
Self-treat



Seek Medical Attention



Emergency



Now

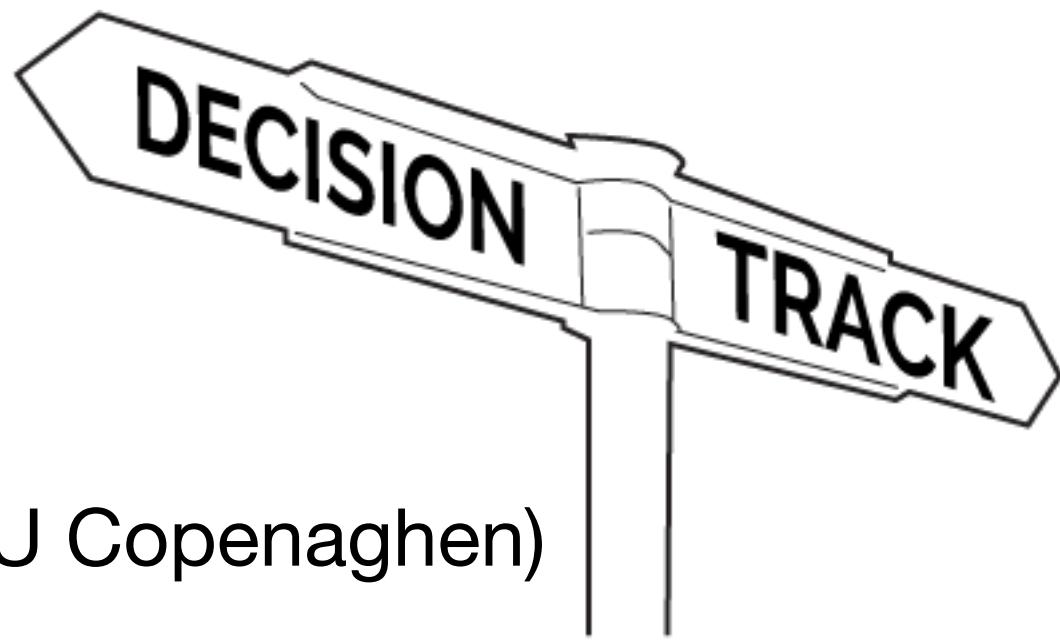
Emergency



The appendix is a pouch on the colon that has no known purpose. Appendicitis begins with pain near the belly button and then moves to the right side. This is often accompanied by nausea, vomiting, poor appetite, fever and chills. Appendicitis is usually treated with surgery and antibiotics. If untreated, the appendix can rupture and cause an abscess or systemic infection (sepsis). Consult a doctor for medical advice

Sources: Mayo Clinic and others. Learn more

The TREC Decision Track



- Goals of the track:
 - **Foster research** on retrieval methods that **promote better decision making with search engines**
 - Develop **new online and offline evaluation** methods to predict decision quality induced by search results
- It ran first time in 2019 with an ad hoc search task. In 2020:
 - Task 1: (Offline, ad hoc) devise search technologies that **promote correct information** over incorrect information, with the assumption that correct information can better lead people to make correct decisions
 - Task 2: (Online) Given a query, a document ranking (results list) and interaction/behavioural data, **predict the decisions** users will take at the end of the search process
 - Task 3: (Offline) Given a query, a document ranking (results list), and assessments (for relevance and correctness), **predict the decisions** users will take at the end of the search process
- The task provides assessments of relevance, credibility, correctness + decisions and interactions

Questions?



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

ie
lab