Health Search

From Consumers to Clinicians

Slides available at

https://github.com/ielab/afirm2019-health-search

Dr. Guido Zuccon

University of Queensland

g.zuccon@uq.edu.au

www.ielab.io

Outline

Slides, references and auxiliary material available at
https://github.com/ielab/afirm2019-health-search

- In this lecture: **Health Information, End Users & Tasks**
- Lecture derived from full day tutorial on health search. Other topics include:
    - Techniques and methods
    - Hands-on with health semantic IR methods
    - Evaluation, open challenges and future directions
- You can find more slides and material at https://ielab.io/health-search-tutorial/

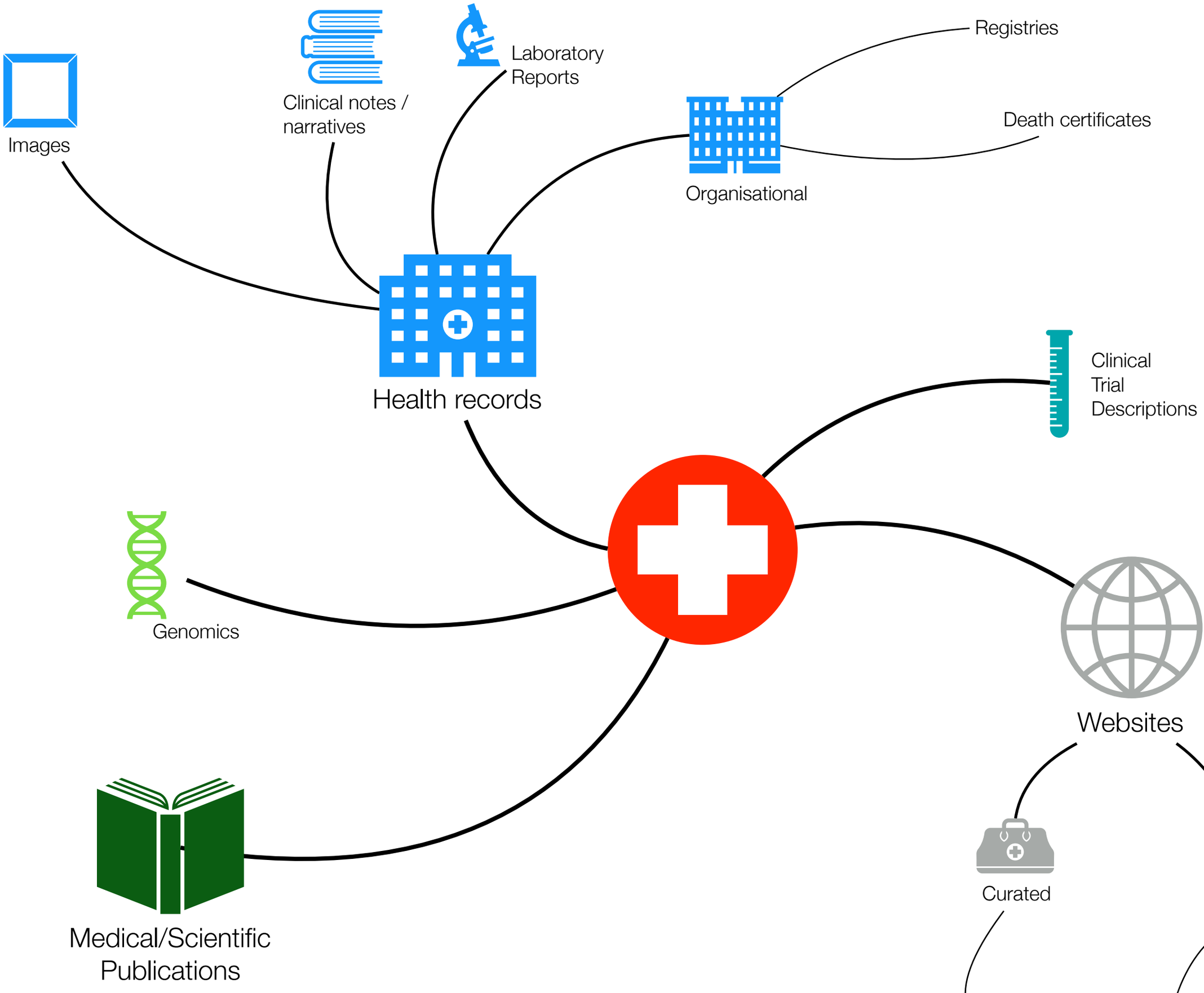
We **separately discuss tasks and methods** because:

- Some methods have been applied across tasks
- Some tasks are affected by the underlying same problems

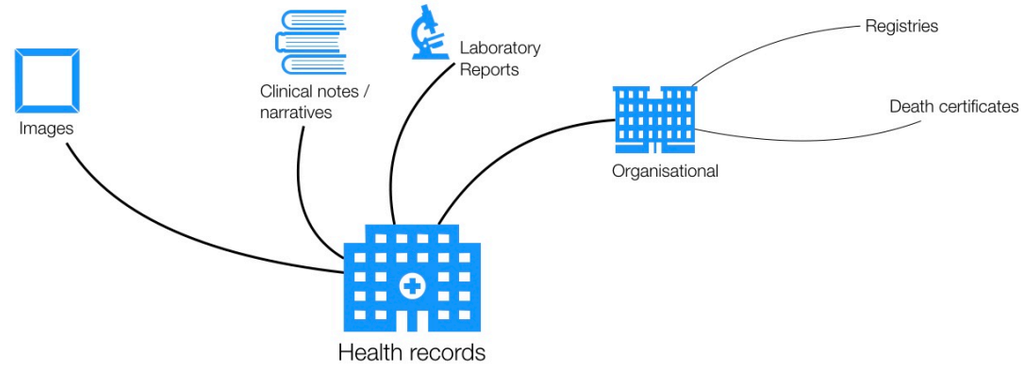
Why health search?

- Large societal impact
    - Advances in health search, could potential translate in better health/ society/economy
    - Good field for attracting research funding
- Fundamental problems are the same/similar to other area of IR, just exacerbated
    - Semantic gap
    - Query formulation
    - Result understanding
    - Cognitive biases, incorrect information fake news, etc

The myriad of health information

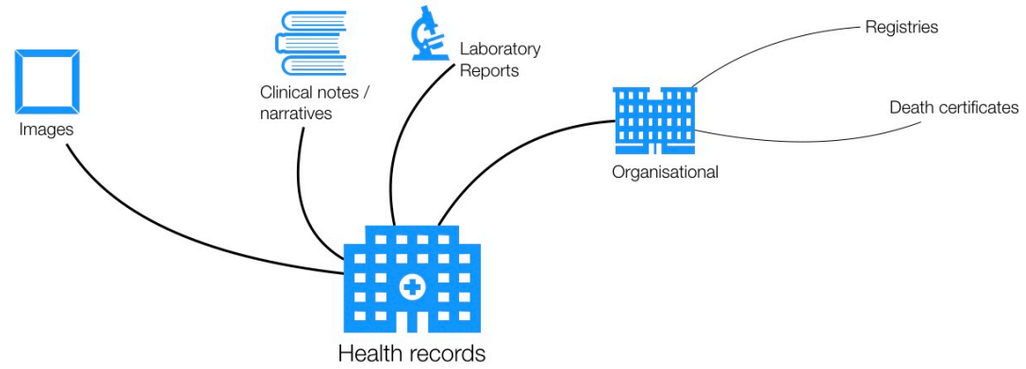

Images
Clinical notes / narratives
Laboratory Reports
Registries
Death certificates
Organisational
Health records
Clinical Trial Descriptions
Genomics
Websites
Curated
Un-curated
Health portals
Social media
Forums
Medical/Scientific Publications

Health Records: Clinical Notes



- Main **purpose** of health records: to communicate information between clinicians
- Often notes contain instructions from one person to another; e.g. from doctor to nurse
  - written by both physicians and nurses
- **record events** during a patient's care
  - to compare past status to current status,
  - to communicate findings, opinions and plans between physicians/nurses
  - for retrospective review of case details

Health Records: Clinical Notes



Samuel J. Smith

1234567-8

4/5/2006

HISTORY OF PRESENT ILLNESS: Mr. Smith is a 63-year-old gentleman with coronary artery disease, **hypertension**, **hypercholesterolemia**, **COPD** and tobacco abuse. He reports doing well. He did have some more knee pain for a few weeks, but this has resolved. He is having more trouble with his sinuses. I had started him on Flonase back in **December**. He says this has not really helped. Over the past **couple weeks** he has had significant congestion and thick discharge. **No fevers or headaches** but does have diffuse upper right-sided teeth pain. He **denies any chest pains**, palpitations, **PND**, **orthopnea**, edema or syncope. His breathing is doing fine. **No cough**. He continues to smoke about half-a-pack per day. He plans on trying the patches again.

CURRENT MEDICATIONS: Updated on CIS. They include **aspirin, atenolol, Lipitor, Advair, Spiriva, albuterol** and will add **Singulair** today.

ALLERGIES: Sulfa caused a rash.

SOCIAL HISTORY: Smokes as above.

REVIEW OF SYSTEMS: CONSTITUTIONAL: Weight stable. GI: No abdominal pain or change in bowel habits.

PHYSICAL EXAMINATION:

VITAL SIGNS: Weight is **217 lbs**, blood pressure **131/61**, pulse **63**.

HEENT: TMs clear bilaterally, mild maxillary sinus tenderness on the right, nasal mucosa boggy with moderate discharge, teeth in good repair with no erythema or swelling

LUNGS: Clear, even with forced expiration.

health specific terms

acronyms

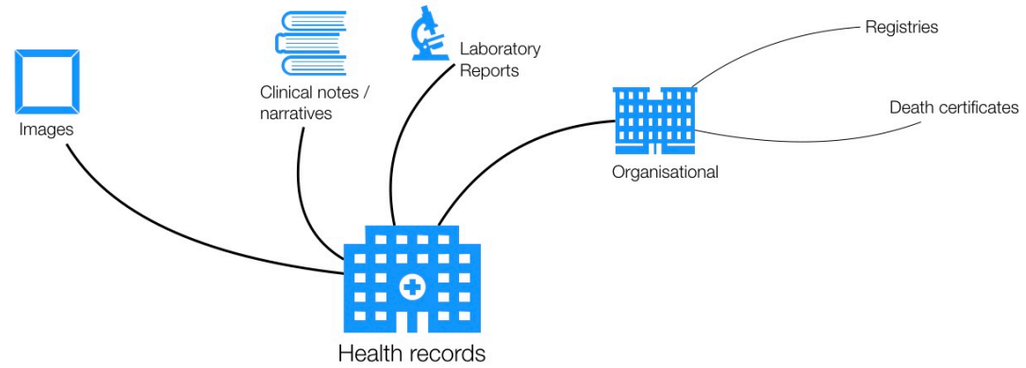
negated terms

temporal

quantities/measurements

brand name vs medication

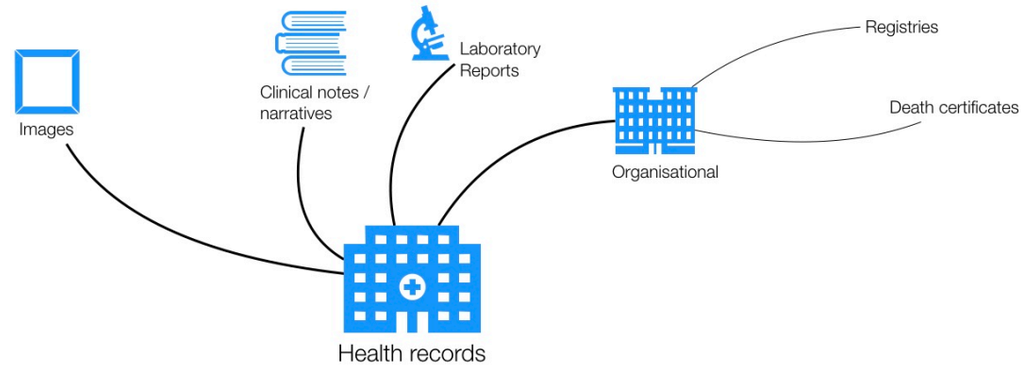
Health Records: Clinical Notes



Clinical notes often noisy:

- **Acronyms** often cannot be told apart:
  - "ARF" could mean "Acute Renal Failure" or "Acute Rheumatic Fever"
- Not consistent **headings** among notes
  - HISTORY OF PRESENT ILLNESS vs HPI
  - MEDICATIONS vs CURRENT MEDICATIONS
- **Temporal** aspects: PAST MEDICATIONS, 2 weeks, etc
- **Negations**: No fever, denies pain, etc…

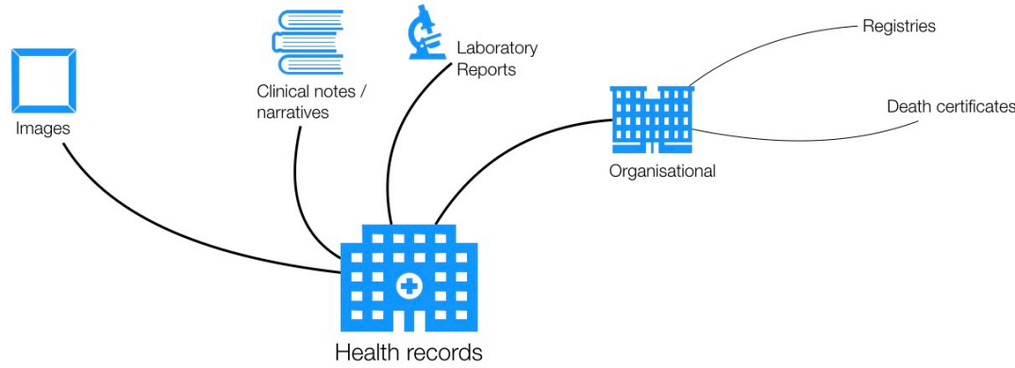
Health Records: Clinical Notes



Clinical notes often noisy:

- **Quantities & measurements** require specific parser and interpretation:
  - blood pressure 131/61: is it high? low?
- **Brand name vs medication**: requires domain knowledge
  - Atorvastatin [medication] vs Lipitor [brand name] vs Statins [medication class]
- **Health specific terms & synonyms**, requires understanding of relations
  - High blood pressure VS hypertension

Health Records: Laboratory Reports



SURGICAL PATHOLOGY REPORT

Diagnosis
Skin, left axilla, punch biopsy-
axillary granular parakeratosis.

Test, Pathologist Pathologist (Electronic Signature)

PT 02/29/2008

Microscopic Examination
Sections show parakeratotic confluent scale containing an abundance of prominent keratohyalin granules. The underlying epidermis shows psoriasiform hyperplasia without acantholysis. The histology defines axillary granular parakeratosis.

Gross Examination
Punch biopsy of skin: left axillary
Size: 0.4 x 0.4 cm
Excision depth: 0.5 cm
Specimen is bisected and entirely submitted in 1 cassette for microscopic examination.

PT /PT

Specimen
From left axillary

Pertinent History
Hailey Hailey

Often reports quantities, in tabular form (thus difficult to machine-read)

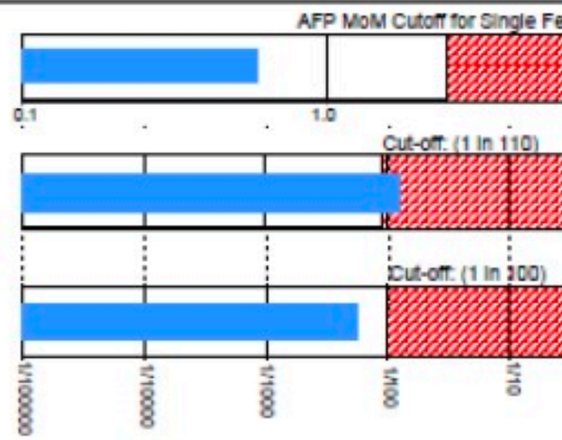
Often come with comments/observations

Client Order ID:

Patient Information Used In Risk Calculations:		Marker	Measurement	MoM
Maternal Age at Delivery:	36.5 yrs	AFP	20 ng/mL	0.59
Estimated Due Date	June 30, 2010	hCG	30000 IU/L	1.13
Gestational Age at Draw:	16 Weeks 1 Day(s)	uE3	0.50 ng/mL	0.53
Maternal Weight:	145 lbs	Inhibin A	300 pg/mL	1.69
Maternal Race:	White	PAPP-A	800 mIU/L	0.54
Number of Fetuses:	Singleton	NT	4.00 mm	3.51
Family History of neural tube defects:	No			
Patient is medication-dependent diabetic:	No	Sonographer Name:	Amie Healy	
Crown Rump Length:	5.30 cm	Sonographer Cert #:	P00943	
		Ultrasound Date:	December 14, 20	

Interpretation:

Open Neural Tube Defects	**Normal**	Risk Before Test: 1 in 900 Risk After Test: <1 in 10000
Down Syndrome	**Abnormal**	Risk Before Test: 1 in 210 Risk After Test: 1 in 80 *
Trisomy 18	**Normal**	Risk Before Test: 1 in 2100 Risk After Test: 1 in 180



Comments:

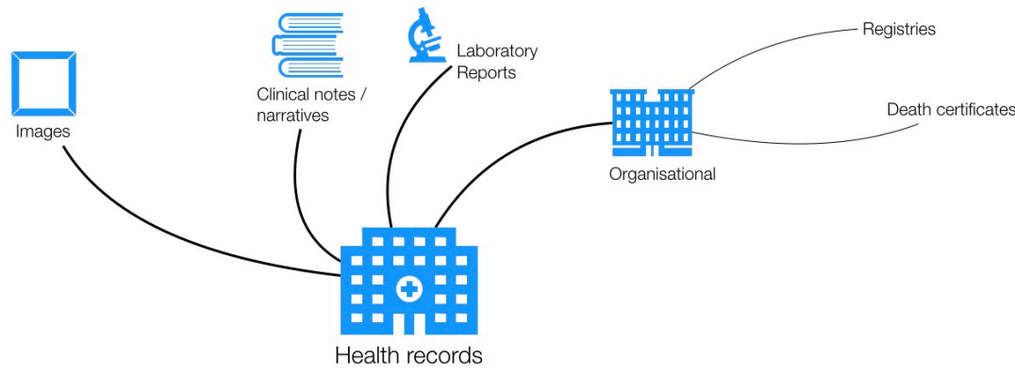
Assuming the patient information listed is correct, this maternal screen is ABNORMAL. Other possible outcomes of abnormal screens include: normal pregnancy, intrauterine fetal demise or missed abortion. If you have questions regarding this screen, please call Genetics at 800-242-2787 ext. 2020.

This is a screening test for Down syndrome, trisomy 18 and open neural tube defects. It will not detect all cases of these disorders, and its ability to identify other chromosome disorders has not been established.

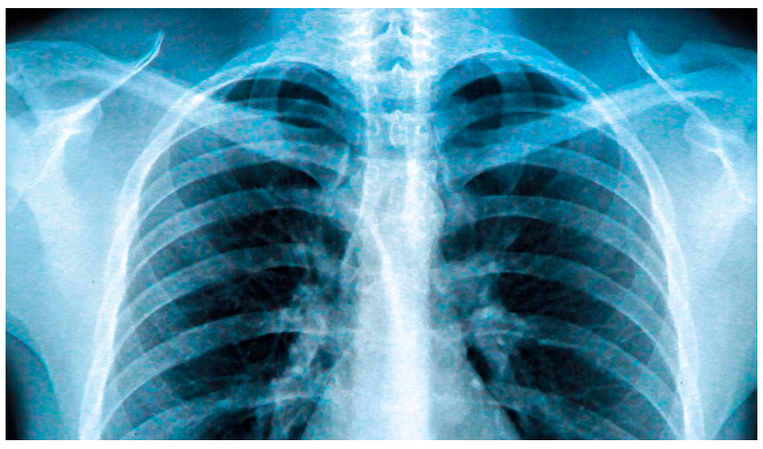
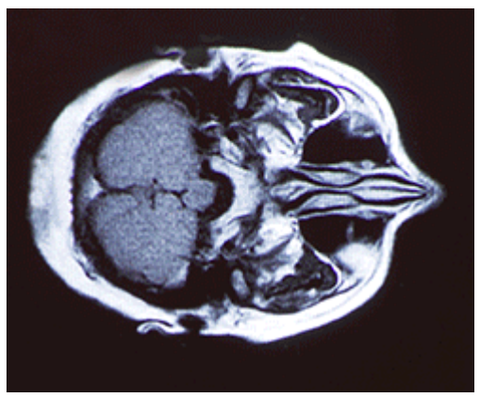
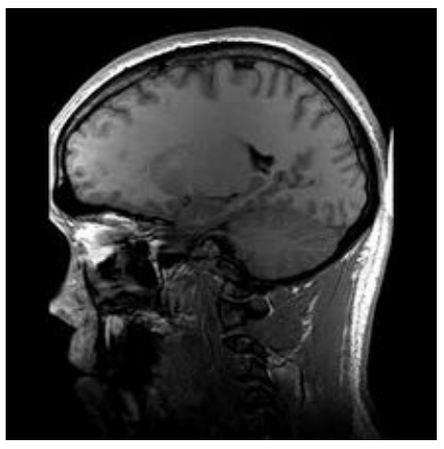
The PAPP-A test uses a kit designated by the manufacturer as "for research use, not for clinical use." The performance characteristics of this test were validated by ARUP Laboratories. The U.S. Food and Drug Administration (FDA) has not approved or cleared this test. The results are not intended to be used as the sole means for clinical diagnosis or patient management decisions. ARUP is authorized under Clinical Laboratory Improvement Amendments (CLIA) and by all states to perform high-complexity testing.

Risk estimates determined using Integrated Test Technology under license from Intema Ltd, UK.

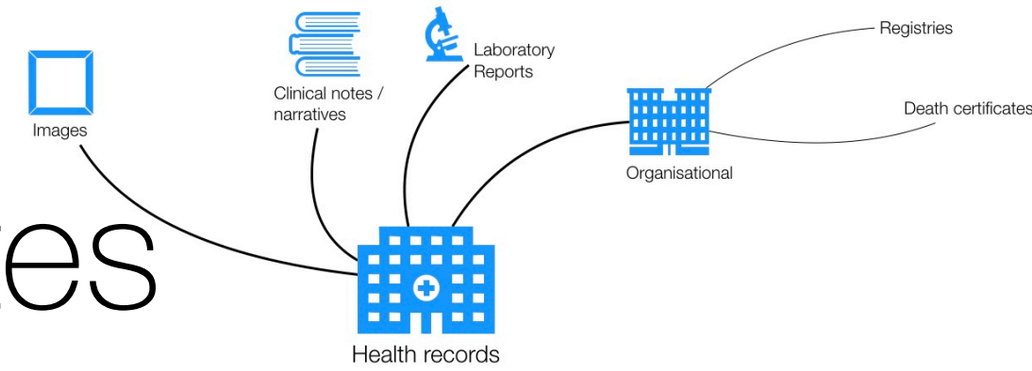
Health Records: Images



- Part of laboratory testing
- X-ray images, CT scans, MRIs, ultrasound imaging
- Sometimes images come along with textual comments/ interpretations: e.g. x-ray reports
  - Interesting for many multimodal information access tasks
- We do not discuss problems in medical image retrieval here. Plenty of work done from the community, both TBIR and CBIR. Have a look at relevant ImageCLEF tasks

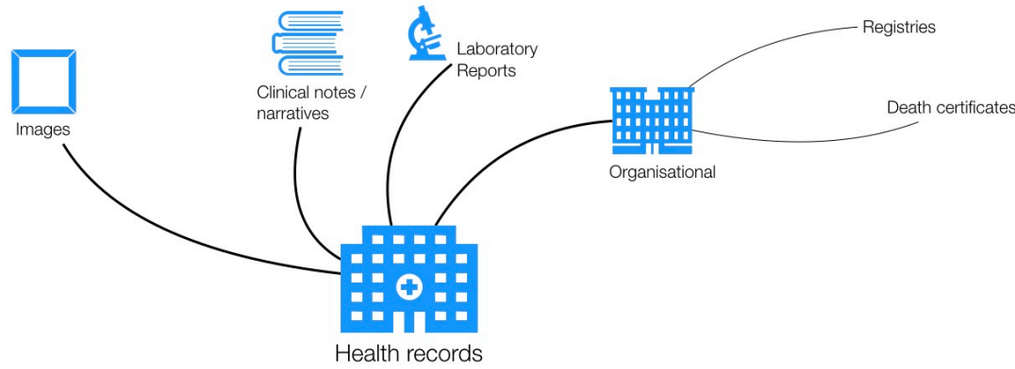





Health Records: Registries & Certificates



- Authorities collect medical data for surveillance and statistical purposes (more on these tasks later)
- Records that are collected are usually:
  - Laboratory tests and reports
  - Death certificates
  - Entries completed through forms
- Collected at population level, into purpose-built databases

Health Records: Death Certificates



Very structured: follow set template, with specific rules and meaning

Contain domain specific terminology

DEATH CERTIFICATE EXAMPLE

Name of deceased **Samuel Clock**

Date of death as stated to me **4th** day of **July** **2018** Age as stated to me **75**

Place of death **Elizabeth Infirmary, Newtown, NE3 4SA**

Last seen alive by me **3rd** day of **July** **2018**

1 The certified cause of death takes account of information obtained from post-mortem.

2 Information from post-mortem may be available later.

3 Post-mortem not being held.

④ I have reported this death to the Coroner for further action.

Please ring appropriate digit(s) and letter

ⓐ Seen after death by me.

b Seen after death by another medical practitioner but not me.

c Not seen after death by a medical practitioner.

CAUSE OF DEATH

I (a) Disease or condition directly leading to death **COMMUNITY ACQUIRED PNEUMONIA**

(b) Other disease or condition, if any, leading to I(a) **PLEURAL MESOTHELIOMA**

(c) Other disease or condition, if any, leading to I(b)

II Other significant conditions CONTRIBUTING TO THE DEATH but not related to the disease or condition causing it **ISCHAEMIC HEART DISEASE, TYPE 2 DIABETES MELLITUS**

Approximate interval between onset and death: **14 days**

The death might have been due to or contributed to by the employment followed at some time by the deceased ☑

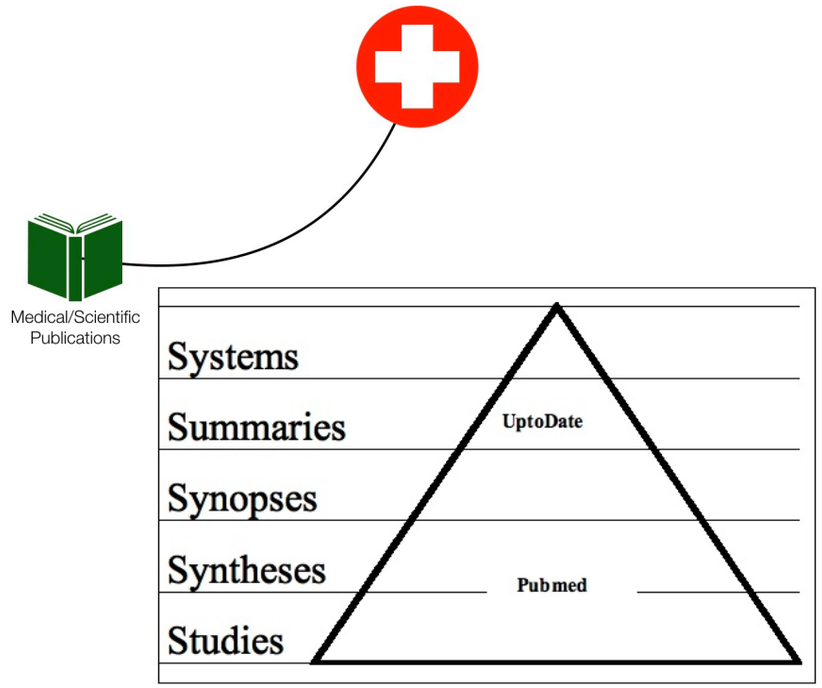
I certify that this death certificate is accurate

Signature *MSmith* **Dr Michael Smith** Qualifications **MBBS (Medicine & Surgery) - GMC 4939**

Residence **Ward 32, Elizabeth Infirmary, NE3 4SA** Date **4/7/18**

Consultant responsible for the above-named patient **Dr Tyvand**

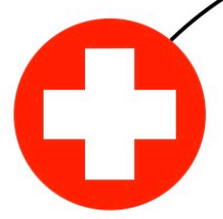
Medical Scientific Publications



[Haynes, 2007; Hoogendam et al., 2008]

- Classification of scientific publications
- **Primary** research:
    - **Published** in journals conference proceedings, technical reports, books, etc.
    - Includes **re-analysis**, e.g., meta-analysis and systematic reviews
    - e.g. PubMed/Medline; often available as title+abstract, not full text
        - Pubmed is an interface used to search Medline, as well as additional biomedical content.
- **Secondary** research:
    - reviews, condensations, synopses of primary literature
    - **textbooks** and **handbooks**
    - **Guidelines** important for normalising care and measuring quality

Clinical Trial Descriptions



- Clinical trials are experiments/observations done in clinical research
- Designed to answer specific questions about biomedical or behavioral interventions, including treatments and interventions
- Clinical trial protocol (description): document used to define and manage the trial.
    - prepared by panel of experts
    - describes scientific rationale, objective(s), design, population, methodology, statistical considerations and organization of the trial
    - Contains inclusion/exclusion criteria of participants
- Clinical trials descriptions are also used to advertise and recruit participants for the trial

Clinical Trial Descriptions

Study Description

Go to

Brief Summary:

Surgery to the shoulder may be performed with patients seated upright in a position known as the "Beach Chair Position (BCP)." This position has certain advantages compared to alternative surgical positions (e.g. side lying) in some situations. However, it has been found that surgery in the BCP can temporarily decrease the amount of oxygen in the brain as a result of the combined effects of gravity and anaesthesia. This can result in complications following surgery such as some memory loss and confusion. Rarely, more serious complications have been reported in the past including death and stroke.

Due to these reported complications the use of "cerebral oximetry" during shoulder surgery in the BCP has become more common. Before and during surger monitor placed on the patients forehead measures the amount of oxygen present in the brain to help control this to an acceptable level. A number of monitors are now commercially available. Two monitors are commonly discussed in the literature; the INVOS™ 5100 and th actual relationship between the supply of oxygen to the brain during surgery and the chance of later developing p "post operative cognitive decline" - POCD) is not clear. It is also not known if one monitor is more accurate than a

Therefore, the main aim of this study is to examine the relationship between cerebral oxygen levels during shoulde problems with memory and thinking). A second aim is to compare the INVOS™ 5100 and FORE-SIGHT® monitor cerebral desaturation events (CDEs) as well as the importance of other key clinical variables (e.g. blood pressure,

Condition or disease	Intervention/treatment
Cognitive Dysfunction	Device: Dual-monitoring

Detailed Description:

PURPOSE OF THE INVESTIGATION The purpose of this investigation is to generate evidence about cerebral oxyg incidence of POCD. Currently, evidence relating to POCD following surgery is conflicting and relates mostly to out strong need to explore this relationship in the specific context of shoulder surgery in the BCP.

INTERVENTION GROUPS This study will involve a single prospective cohort. Patients who meet the selection crit

Eligibility Criteria

Go to

Information from the National Library of Medicine

Choosing to participate in a study is an important personal decision. Talk with your doctor and family members or friends about deciding to join a study. To learn more about this study, you or your doctor may contact the study research staff using the contacts provided below. For general information, Learn About Clinical Studies.

Ages Eligible for Study:	18 Years to 99 Years (Adult, Older Adult)
Sexes Eligible for Study:	All
Accepts Healthy Volunteers:	No

Criteria

Inclusion Criteria:

- Receiving treatment primarily by, but not restricted to, one of the Primary investigators for a shoulder condition that requires surgery in the BCP.
- Over 18 years of age
- Able to read and speak English

Exclusion Criteria:

- Under 18 years of age
- Pregnant women
- Pre-operative Mini-Mental State Examination (MMSE) < 24
- Pre-existing cerebrovascular disease as reported by the assessing medical consultant and recorded in patient charts
- Orthostatic hypotension
- American Society of Anaesthesiologists (ASA) physical status III, IV and V*
- History of drug and/or alcohol abuse

https://clinicaltrials.gov/ct2/show/NCT03036345

Websites

- **Curated websites**:
  - **Health portals**: webmd, mayoclinic, medlineplus, uptodate, medscape, everydayhealth, etc
    - Often from govt, company, edu
  - Generalist knowledge bases: **Wikipedia** (EN: 4.8 billion pageviews in 2013) and other wikis (https://en.wikipedia.org/wiki/List_of_medical_wikis)
  - **Symptom checkers**: provide diagnoses and triaging based on Q&A interaction
    - E.g. https://symptoms.webmd.com
- Provide carefully collated health information, reliable, clearly written
- Sometimes inconclusive, e.g. “consult a doctor”
- Symptom checkers often incorrect, or inconclusive
  - [Semigran et al, 2015]: 23 symptom checkers studied:
    66% of cases misdiagnosis; 43% of mis-triaged

Websites

- **<u>Un-curated websites:</u>**
  - **promotional**: attempt to promote a service/treatment/etc
  - **experiential**: reporting on the experience with a disease/ treatment/service provider
  - **informational**: provide info about a product/service
  - Often from company, individual (doctor, health advocate, patient), news
- Widely vary in quality, trustworthiness and ease of understanding
- Often forcefully driving to a specific choice/solution

Websites

- **<u>Un-curated websites:</u>**
  - **Forums**: reddit AskADoctor (et al), PatientsLikeMe, HealthTap, patient.info
    - Often connect patients with doctors
    - Of varying quality and control, e.g. Reddit VS HealthTap
  - **Social media**: increasing use of Facebook, Twitter for sharing health content [Benetoli et al., 2017]
    - Healthcare promotion, but also promotion of products/services
    - Asking/sharing health advice among personal network, personal experiences

Quality of health information online

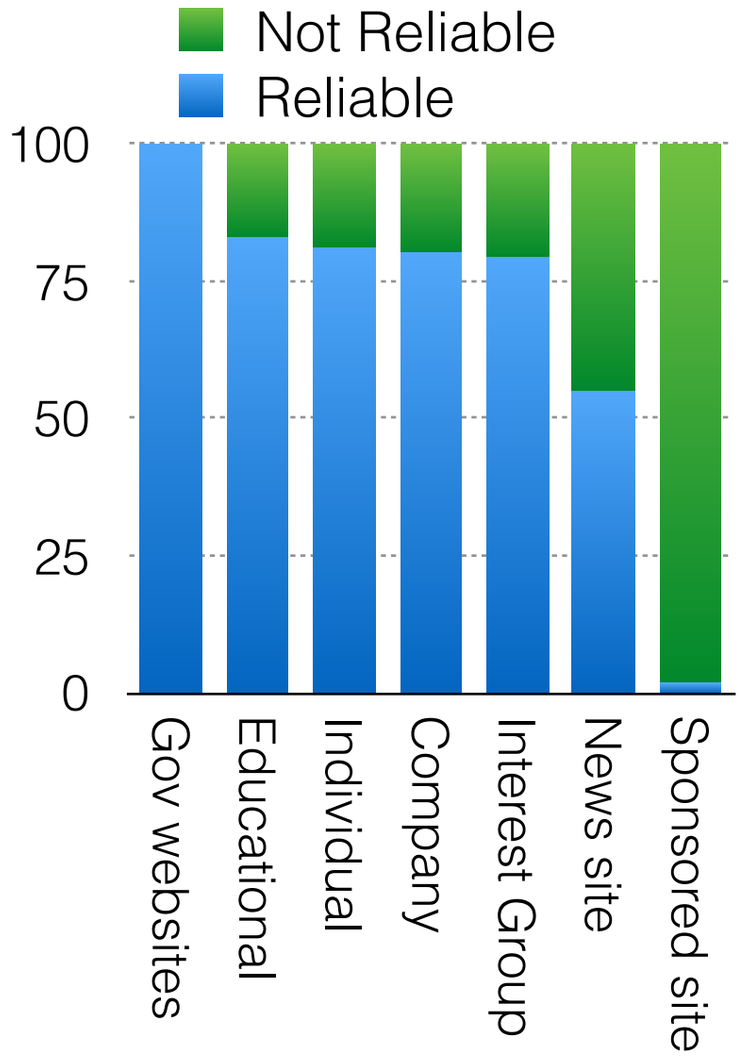
[Zhang et al., 2015]: systematic review of literature on quality of online health information (N=165). Literature has measured

1. **substance** of content: **accuracy** and completeness
2. **formality** of content: currency, **credibility** (**trustwortiness**), **readability** (**understandability**)
3. **design** of platforms: accessibility, aesthetics, navigability, interactivity, privacy, cultural sensitivity

- quality of health information **varied across medical domains and websites**
- overall **quality is problematic** (55.2% negative, 6.1% positive)
- most analysed work has not used “real” queries

Trustworthiness of health information online

- [Scullard et al., 2010]: evaluated first 100 search results for 5 paediatric web queries
- 39% gave correct information; 11% were incorrect and 49% failed to answer the question
- Correctness varied across topics, gov sites gave uniformly accurate advice

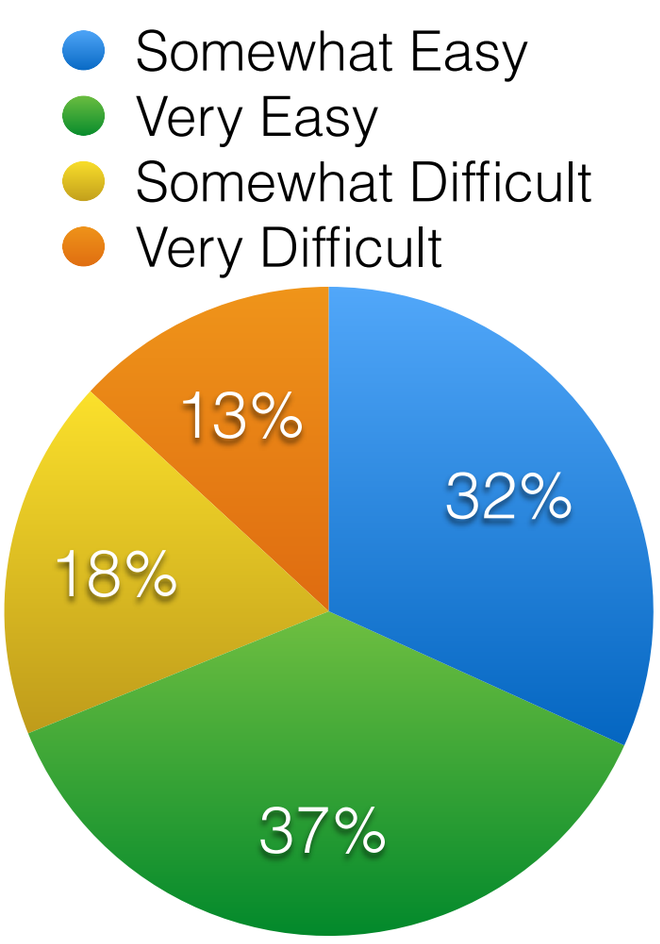


Trustworthiness of health information online

- [Rains et al., 2009]: studies what **influence credibility** of health web pages (N=86, students)
  - **structural features of pages** and **message characteristics** related to perceptions of credibility
    - **Credible** websites have: **navigation menus, links to external web sites, organisation's physical address**, **statistics, references&quotes, and identification of authorship**
- [Sbaffi&Rowley, 2017]: review of literature on health web pages trust (N=73)
  - Positive effect on trust: ease of use, content, website design, clear layout, interactive features, authority of owner/author
  - **Negative** effect on trust: **advertising**

Readability of health information online

- Many studies on readability/understandability of health web pages
- Based on **measures** of readability, e.g. [Hutchinson et al., 2016]:
  - Used Flesch Kincaid Grade Level, Gunning Fog Score, SMOG index, Coleman Liau Index, Automated Readability Index
  - Top Google results hard to understand for grade <9; NIH recommendation grade 6-7.
- Based on **assessments**:
  - [Palotti et al., 2015] analysis of CLEF 2015 CHS qrels: people believe they well understand only ~40%



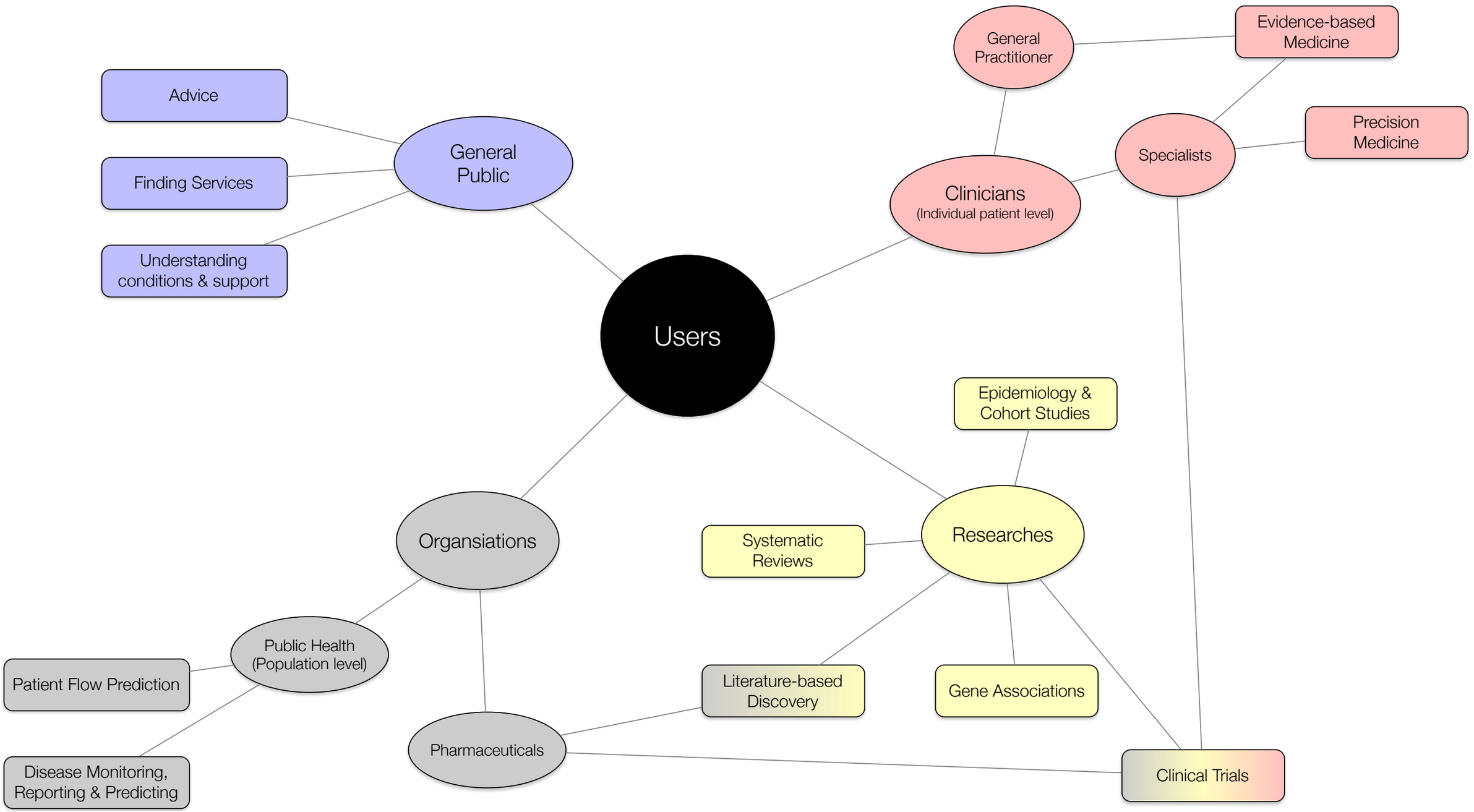
High quality health webpages: HON Guidelines

- Health On the Net (HON): organisation that promotes transparent and reliable health information online
- HON **guidelines** for web pages: https://www.hon.ch/HONcode/Guidelines/guidelines.html
- This could be used as features to determine quality of page:
  - Indication of authorship (if collaborative platform: whether moderated)
  - Purpose of website
  - Confidentiality & privacy
  - Referencing and dating
  - Justification of claims, all brand names identified
  - Website contact details/contact form
  - Disclosure of funding sources
  - Advertising policy

Users and tasks

User Task

Users & Tasks



What do clinicians search for?

[Ely et al., 2000]: created a **taxonomy of clinical questions**

- Analysed ~1400 questions -> 64 generic question types. Top 10:
  - What is the **drug** of choice for condition x? (11%)
  - What is the **cause** of symptom x? (8%)
  - What **test** is indicated in situation x? (8%)
  - What is the dose of drug x? (7%)
  - How should I treat condition x (not limited to drug treatment)? (6%)
  - How should I manage condition x (not specifying diagnostic or therapeutic)? (5%)
  - What is the cause of physical finding x? (5%)
  - What is the cause of test finding x? (5%)
  - Can drug x cause (adverse) finding y? (4%)
  - Could this patient have condition x? (4%)
- These are questions asked by clinicians in primary care, **not queries** to a search system

What do clinicians search for?

[Del Fiol et al., 2014]: systematic review focusing on **clinicians questions**

- 0.57 questions per patient
- 34% of questions concerned **drug treatment**; 24% concerned potential **causes** of a symptom, physical finding, or diagnostic test finding
- Only **51% of questions are pursued**
  - Why not: (A) **lack of time** (B) doubt that a **useful answer exists**
  - Makes a case for **just-in-time access** to **high-quality evidence** in the context of patient care decision making
- Found **answers to 78% of those pursued** (not just through search)
  - Note answers may not be correct!

How do Clinicians Search?

Queries:

- [Meats et al., 2007] analysed TRIP database queries:
  - most **single term**; ~12% **Boolean** operator (11%"AND" + 0.8% "OR")
  - PICO elements: **population** was most commonly used; lesser use of intervention. Comparator and outcome rarely used
  - top 20 terms related to disease, condition, or problem; fewer terms related to treatment, intervention, or diagnostic test
  - users interested in conducting effective/efficient searches but **do not know how**
- [Tamine et al., 2015]: examined clinical queries from TREC (Genomics, Filtering, Medical Records) and imageCLEF
  - language **specificity level varies** significantly across **tasks** as well as **search difficulty**

How do Clinicians Search?

Queries:

- [Palotti et al., 2016]: analysed HON+TRIP+others logs
  - **2.91 terms** per query / 3.24 queries per session
  - Disease queries more prevalent than treatment
- [Koopman et al., 2017]: analysed query behaviour of a clinicians (N=4)
  - **Number** of queries a clinician would issue depend on: **topic** & **clinician**
  - **Verbose querier** (avg-len: 5.1-6.6 terms) vs **concise querier** (avg-len: 2.8-3.5 terms)
  - Verbose querier enters on average **less queries** per topic (1.37-1.59); concise querier enters on avg **more queries** (2.54-2.81)

How do Clinicians Search?

Time:

- [Hoogendam et al., 2008]: < 5 minutes
- [Westbrook et al., 2005]: ~8 minutes
- [McKibbon et al, 2006]: ~13 minutes
- [Palotti et al., 2016]: ~4.5 minutes
    - medical experts more persistent, interact longer with search engine than consumers

Clinicians' Search Tasks

- **Evidence based medicine**: searching **literature** to answer a clinical question (diagnosis/test/treatment) [Roberts et al., 2015]
  - Clinicians expected to seek and apply the best evidence to answer their clinical questions
  - Large reliance on secondary literature: guidelines, handbooks, synthesised information (57% of clinicians prefer secondary literature [Ellsworth et al., 2015])
  - Primary literature of interest: re-analyses

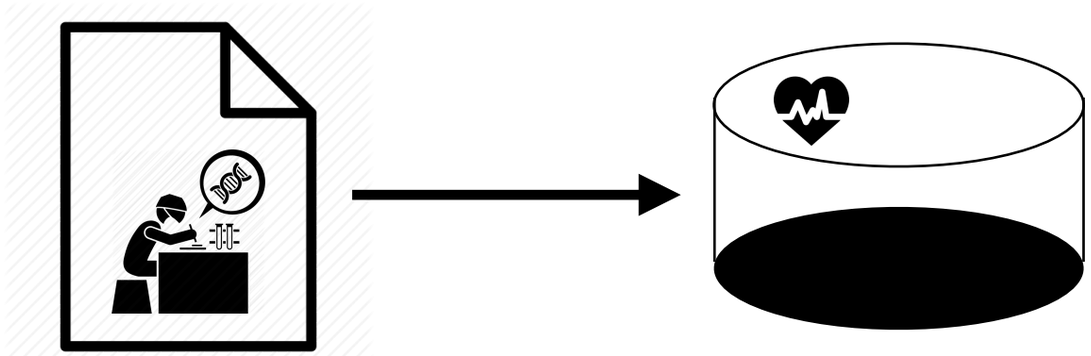
(Note, TREC CDS considers only primary literature)

- **Precision Medicine**: akin to EBM, but no "one size fits all": proper treatment depends upon genetic, environmental, and lifestyle [Roberts et al., 2017]
  - use detailed patient information (genetic information) to identify the most effective treatments
  - huge space of treatment options: difficulty in keeping up-to-date & hard to determine the best possible treatment

(Note, TREC PM also considers clinical trials as a fall-back)

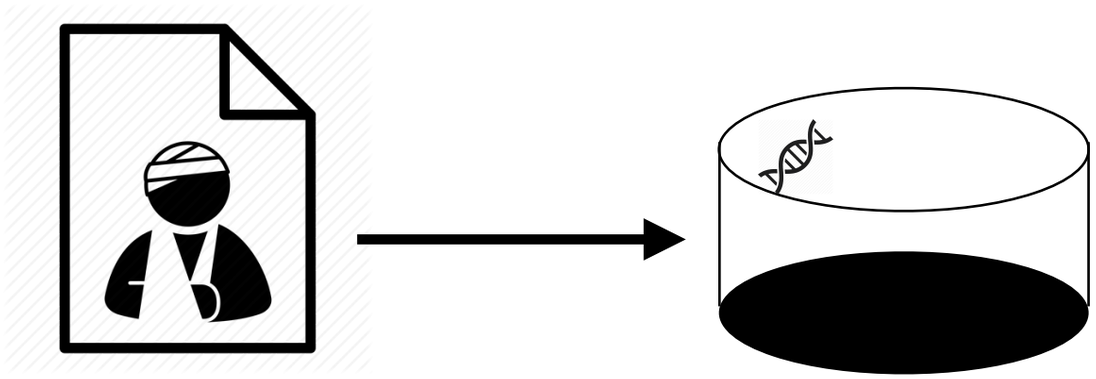
Medical Researchers' Search Tasks

- **Clinical Trials**:
  - MR/Org: leverage health records to **identify** potential **participants** [Voorhees, 2013]



Clinical Trial EHR Repository

  - Clinician: given a patient, **identify** clinical **trials** the patient could be eligible for [Koopman&Zuccon, 2016]



Patient's EHR Trials Repository

Different Users Search Differently for Clinical Trials

"A 51-year-old woman is seen in clinic for advice on osteoporosis. She has a past medical history of significant hypertension and diet-controlled diabetes mellitus. She currently smokes 1 pack of cigarettes per day. She was documented by previous LH and FSH levels to be in menopause within the last year. She is concerned about breaking her hip as she gets older and is seeking advice on osteoporosis prevention."

Automatic system on GP computer thing to match health record with a trial

"51-year-old smoker with hypertension and diabetes, in menopause, needs recommendations for preventing osteoporosis."

GP searching

- peripheral arterial disease
- cardiovascular disease
- peripheral vascular disease and possible therapies to prevent ischaemic limb
- calf Pain Exercise History of Myocardial infarct Hypertension polypharmacy
- peripheral vascular disease trial
- lower limb claudication trial
- peripheral arterial disease trial

Medical specialist performing ad-hoc search

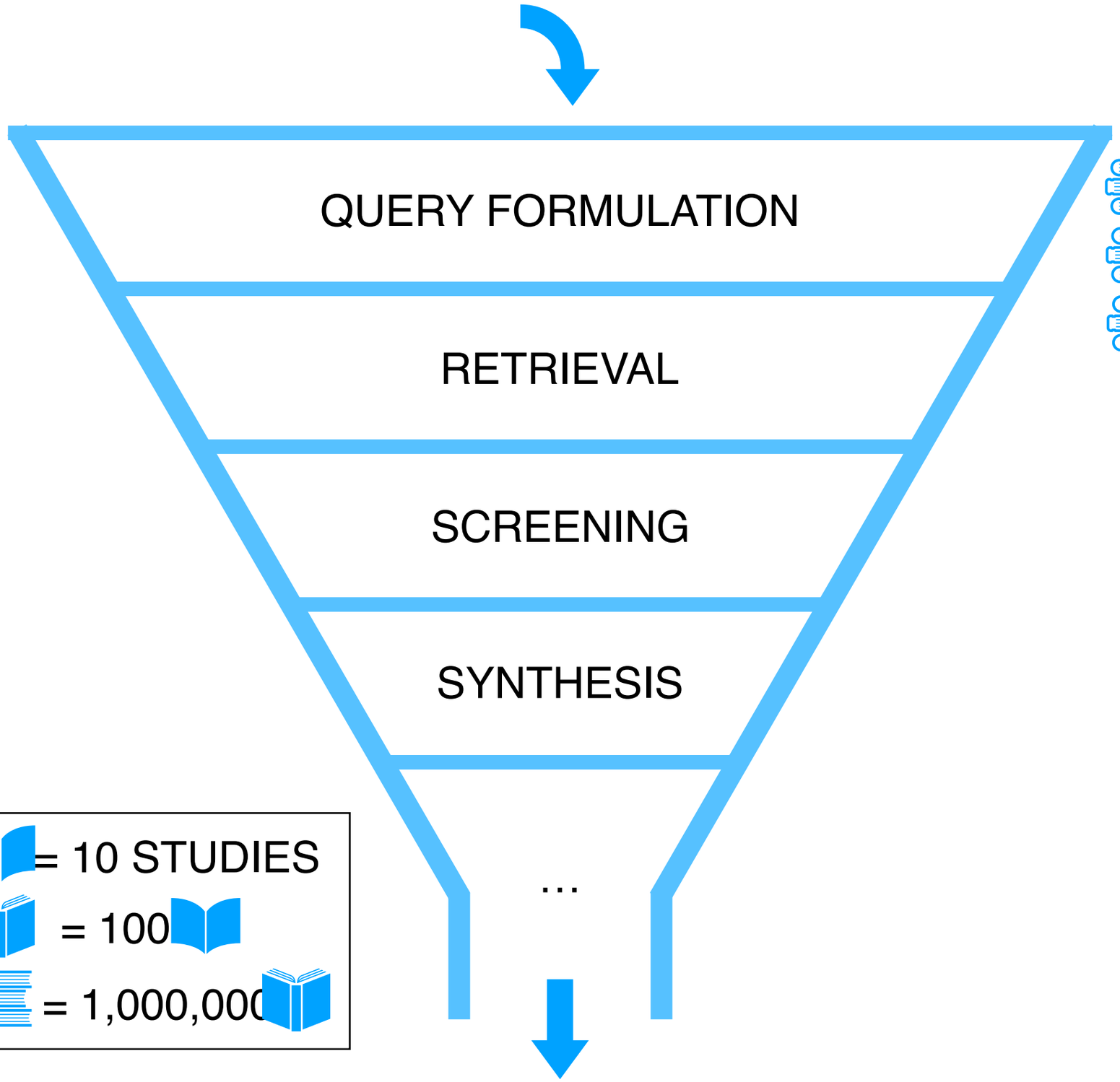
[Koopman&Zuccon, 2016]

Medical Researchers' Search Tasks

- **Systematic Reviews**: identify literature to screen for inclusion in a systematic review [Scells et al., 2017; Kanoulas et al., 2017]
  - Systematic review is a focused literature review
    - Synthesises all relevant documents for a particular research question; following protocol (which defines a boolean query)
  - Guide clinical decisions and inform policy
    - Cornerstone of evidence based medicine

RESEARCH QUESTION: ARE CARDIO SELECTIVE BETA-BLOCKERS…

Research question created



26 million citations in PubMed

4 million citations retrieved

278 citations screened as potentially relevant

22 studies chosen to be included

= 10 STUDIES

= 100

= 1,000,000

RECOMMENDATION: BETA-BLOCKER TREATMENT REDUCES MORTALITY…

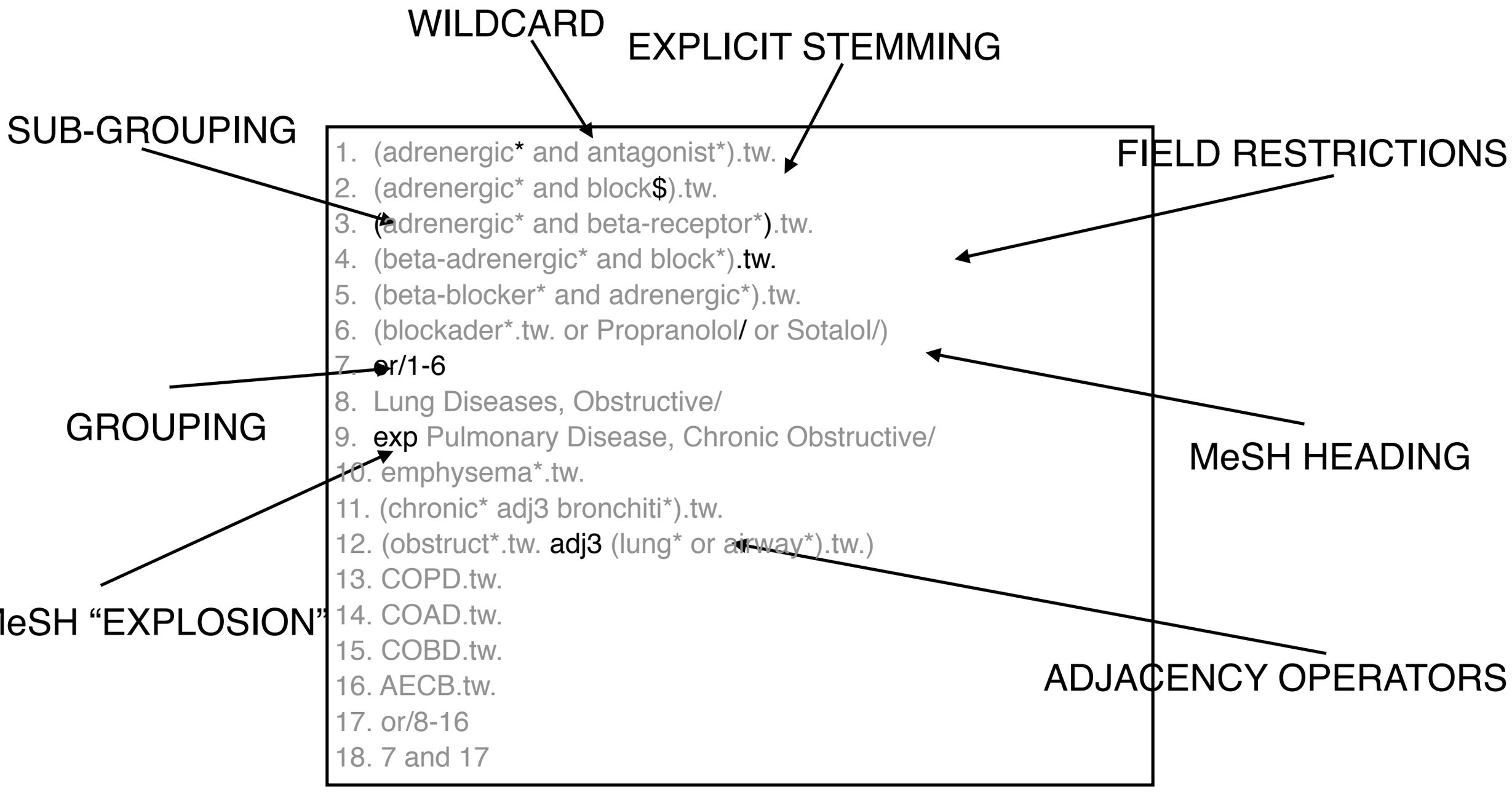
Studies synthesised to produce recommendation

Queries in Systematic Reviews

THESE AREN'T YOUR NORMAL BOOLEAN QUERIES

1. (adrenergic* and antagonist*).tw.
2. (adrenergic* and block$).tw.
3. (adrenergic* and beta-receptor*).tw.
4. (beta-adrenergic* and block*).tw.
5. (beta-blocker* and adrenergic*).tw.
6. (blockader*.tw. or Propranolol/ or Sotalol/)
7. or/1-6
8. Lung Diseases, Obstructive/
9. exp Pulmonary Disease, Chronic Obstructive/
10. emphysema*.tw.
11. (chronic* adj3 bronchiti*).tw.
12. (obstruct*.tw. adj3 (lung* or airway*).tw.)
13. COPD.tw.
14. COAD.tw.
15. COBD.tw.
16. AECB.tw.
17. or/8-16
18. 7 and 17

Anatomy of a Systematic Review Query



Why improving search within systematic reviews is important

- A majority of reviews require >1,000 hours to complete [Allen&Olkin, 1999]
- Can cost upwards of a quarter of a million USD [McGowan&Sampson, 2005]
- [McGowan&Sampson, 2005]: Most **expensive** and **laborious** phases **prior to eligibility**

Consumers searching for Health Advice on the Web

- **People seek health advice online**, often through search engines
  - 1/3 Americans [Fox&Duggan, 2013]
  - 65-95% of people across different countries [McDaid&Park, 2010]
- Many consumers reported being **unable to find** satisfactory information when performing a specific query [Zeng et al., 2004]
  - information found was **not new**
  - information found was too **general**
  - **confusing** interface or organization of website
  - information **overload** (too much information was retrieved)
- Vast differences in **comprehension**, **searching abilities**, and **levels of information needs**

The dark side of searching for health advice on the Web

- **Cyberchondria**: **unfounded escalation** of concerns about **common symptomatology**, based on the review of **search** results and literature on the Web [White&Horvitz, 2009]
  - log-based study + survey of 515 search experiences
  - escalation associated with
    - **amount** and **distribution** of medical content viewed by users,
    - presence of **escalatory terminology** in pages visited
    - user's **predisposition** to escalate versus to seek more reasonable explanations
- [Pogacar et al., 2017]: search engine results can significantly **influence people** taking **positive/negative** decisions based on **correct/incorrect** health information
  - User study (n=60) with biased search results towards correct or incorrect information regarding treatment
  - more incorrect decisions when interacting with results biased towards incorrect information

What do consumers search for?

- [Schwartz et al., 2006] surveyed ~1400 families
- Search topics: diseases/conditions (79%), medications (53%), nutrition&exercise (48%), providers (35%), prevention (34%), alternative therapies (25%)
- Subtasks in consumer health search:
    - Finding health **advice** (to support health decision)
    - **Understand** condition, treatments, etc
    - Find health **provider**

How do consumers search?

- [Eysenbach&Köhler, 2002]:
  - 65% of queries are **single keyword**; 3.5% contain a phrase.
  - **Rarely** look **beyond first** SERP
  - Spend about **6 minutes** searching
- [Zeng et al, 2006]: ~60-70% queries are one to two words
  - **difficulty** in **understanding and use medical terminology**.

How do consumers search?

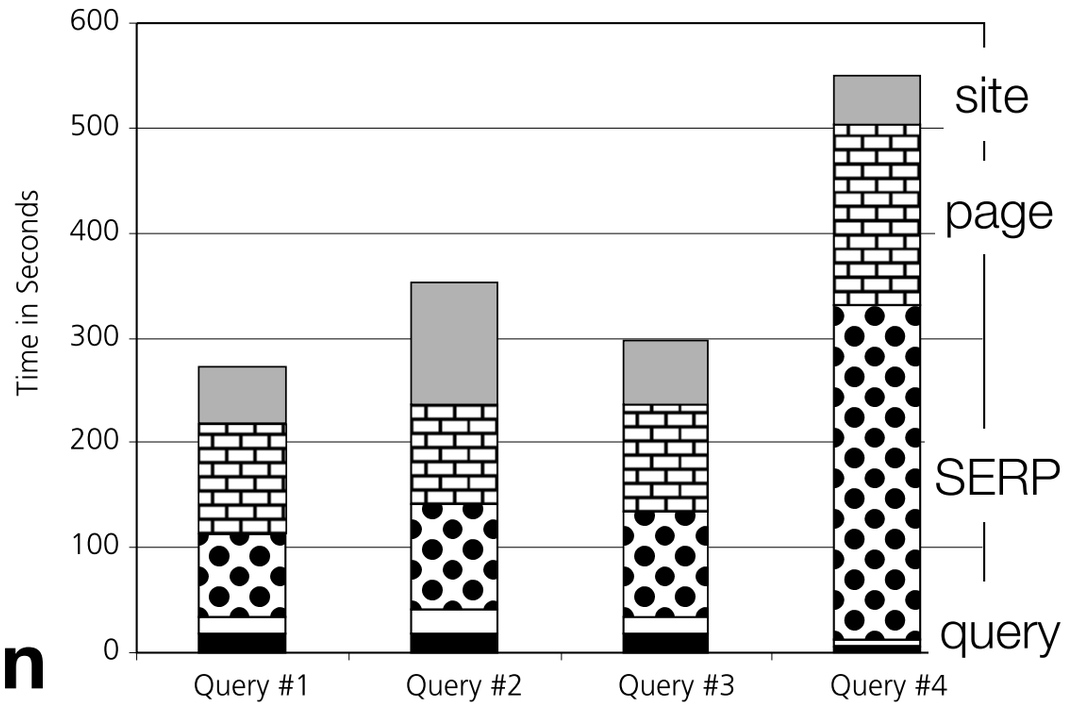
- [Toms&Latter, 2007] examined search behaviour of 48 consumers on 4 health search tasks

  - Analysed transaction logs, video screen capture, retrospective verbal protocols, self-reported questionnaires

    - ~1.3 **queries** per search **task**.
    - query **length** ~ 4.2 keywords (3.2 stopwords)
    - ~ 5.4 **SERPs** examined

- significant problems in **query formulation** and in making **efficient selections** from SERP

Image from [Toms&Latter, 2007]



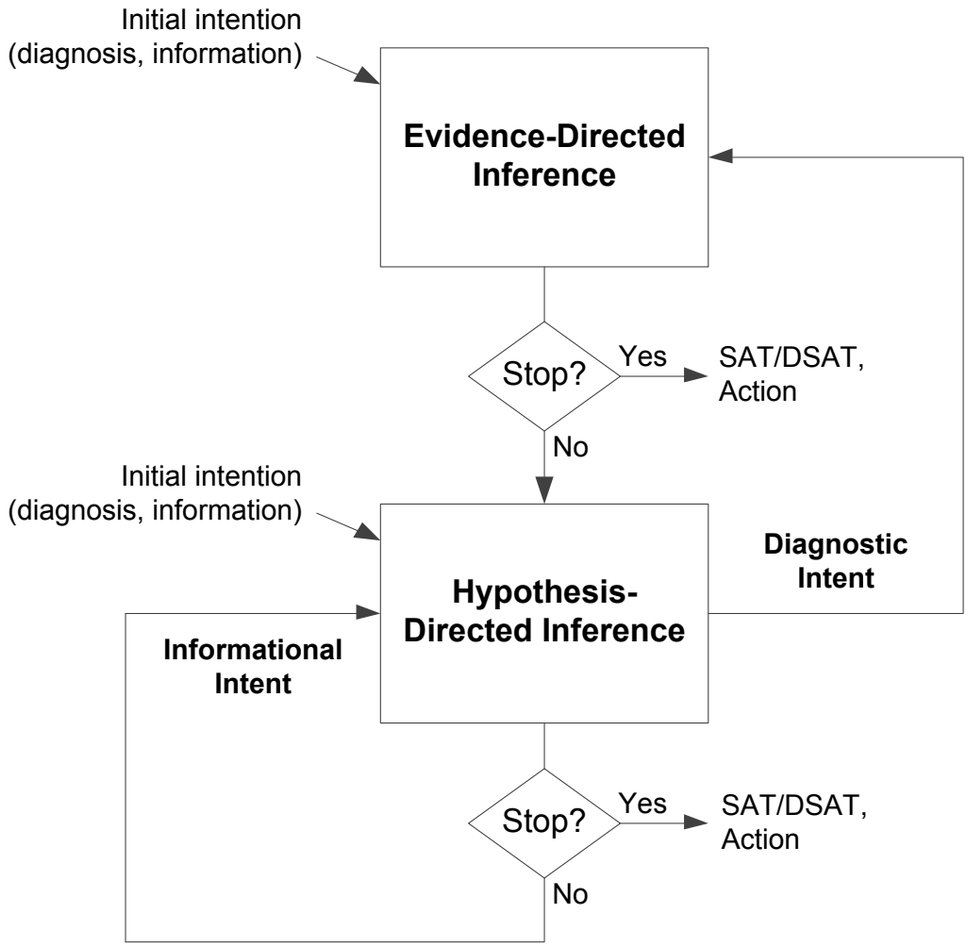
- 4.5–9 minutes per task.
- Time spent on SERP ~ time spent on webpage

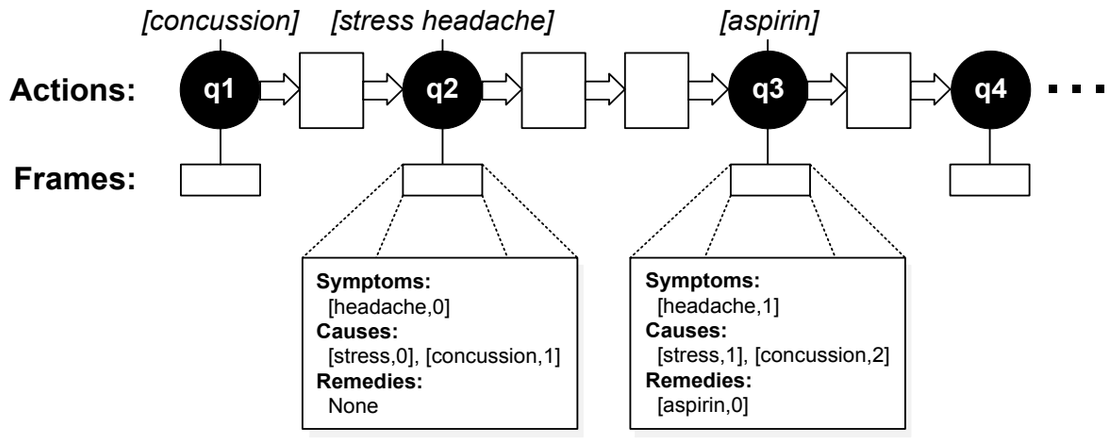
Exploratory Behaviour in CHS

- [Cartright et al., 2011] argue that a portion of health-directed searches are **exploratory** in nature. These could be divided into **two iterative phases**

- **evidence-directed**: findings are fused to construct a list of potential explanatory diagnoses ranked by likelihood

- **hypothesis-directed**: list of diagnoses used to guide collection of additional evidence, to validate/choose hypotheses.





Images from [Cartright et al., 2011]

How do consumers search? Querying…

What would be your query to Google if you have this on your skin?

[Zuccon et al., 2015]

How do consumers search?
Querying…

What would be your query to Google if you have this on your skin?

q: “Crater type bite mark”

q: “Ring wound below wrinkled eyelid”

[Zuccon et al., 2015]

How do consumers search?
Querying…

What would be your query to Google if you have this on your skin?

q: “Crater type bite mark”

q: “Ring wound below wrinkled eyelid”

What Bit Me? Mystery Bug Bites Solved | SafeBee
www.safebee.com › Outdoors ▾
Jun 16, 2015 - What it's **like**: You may feel a sharp **sting** when you're **bitten** or nothing at all. ... The brown recluse has a violin-shaped **mark** on its back that isn't ... six weeks to go away, and the **bite** can leave a large **crater** and scarring.

[Zuccon et al., 2015]

Cognitive bias when search for health information

- **Web searchers** exhibit their **own biases** and are also subject to **bias from search engine** [White, 2013], e.g. favour positive information over negative
- [Lau&Coiera, 2007]: 75 clinicians + 227 students; studied influence on decision post-search of different biases:
  - prior belief (**anchoring**): p 0.001
  - documents **order effect**: clinicians p 0.76; students p 0.026
  - documents processed for different lengths of time (**exposure effect**): clinicians p 0.27; students p 0.0081
  - **reinforcement through repeated exposure** to a document: no significant impact (clinician p 0.31; students p 0.81)
- [Lau&Coiera, 2006] proposed bayesian model to predict the impact of search results on health decision, with cognitive biases
- [Lau&Coiera, 2009] proposed mechanisms to de-bias search (mostly to do with search result presentation)

Summary of Problems in CHS

- **Query formulation**
    - Vocabulary mismatch b/w layman and professional language
    - Describing rather than naming (circumlocutory queries): use of medical terminology
- **Result appraisal** (both SERP and document)
    - Understanding medical language/resources
    - Ability to tell correct from incorrect advice (credibility)
    - Cognitive biases

Summary of Problems when Clinicians Search

- Mostly centred around the **semantic gap problem** [Koopman 2014]
    - the difference between the raw (medical) data/evidence and the way a human being might interpret it [Patel et al., 2007]
- **Vocabulary** mismatch
    - *hypertension vs. high blood pressure*
- **Granularity** mismatch
    - *Malaria vs. Plasmodium*
- Conceptual **implication**
    - *Dialysis Machine → Kidney Disease*
- **Inferences** of similarity
    - *Comorbidities (Anxiety and Depression)*
- Other problems: use of **negation**, **temporality** and **quantities**, age/gender, levels of evidence (e.g. discharge summary VS lab test; study VS systematic review)

Summary of Problems when Clinicians Search

- Mostly centred around the **semantic gap problem** [Koopman 2014]
    - the difference between the raw (medical) data/evidence and the way a human being might interpret it [Patel et al., 2007]
- **Vocabulary** mismatch
    - *hypertension vs. high blood pressure*
- **Granularity** mismatch
    - *Malaria vs. Plasmodium*
- Conceptual **implication**
    - *Dialysis Machine → Kidney Disease*
- **Inferences** of similarity
    - *Comorbidities (Anxiety and Depression)*

Note semantic gap problems occur also for CHS, with vocabulary mismatch being the most prevalent

- Other problems: use of **negation**, **temporality** and **quantities**, age/gender, levels of evidence (e.g. discharge summary VS lab test; study VS systematic review)

Pointers to Methods, Evaluation, Resources

Pointers to: Methods in Health Search

- Dealing with the **semantic gap**: exploiting the semantics of medical language
    - concept based search & inference, query expansion, learning to rank, embeddings, neural networks
    - Implicit VS explicit semantics
- Dealing with the nuances of **medical language**
    - negation, family history, understandability
- Understanding and aiding **query formulation**
    - query variations, query reformulation, query clarification, query suggestion, query intent, query difficulty, task-based solutions

Implicit VS Explicit Semantics

- Explicit semantics: structured human representation of knowledge and its concepts
    - e.g., medical terminologies
- Implicit Semantics: draw representation of words/concepts from data
    - e.g., distributional/latent semantic models

ICD

International Statistical Classification of Diseases and Related Health Problems (ICD)

Diagnosis classification from World Health Organisation

Used extensively in **billing**

International Statistical Classification of Diseases and Related Health Problems 10th Revision

Chapter	Blocks	Title
I	A00–B99	Certain infectious and parasitic diseases
II	C00–D48	Neoplasms
III	D50–D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00–E90	Endocrine, nutritional and metabolic diseases
V	F00–F99	Mental and behavioural disorders
VI	G00–G99	Diseases of the nervous system
VII	H00–H59	Diseases of the eye and adnexa
VIII	H60–H95	Diseases of the ear and mastoid process
IX	I00–I99	Diseases of the circulatory system
X	J00–J99	Diseases of the respiratory system
XI	K00–K93	Diseases of the digestive system
XII	L00–L99	Diseases of the skin and subcutaneous tissue
	M00–	Diseases of the musculoskeletal system

Unified Medical Language System (UMLS)

- UMLS is a compendium of many controlled vocabularies in the biomedical sciences

- **Combined many terminologies under one umbrella**
- UMLS concept grouped into higher level semantic types
  - Concept: *Myocardial Infarction* [C0027051] of type *Disease or Syndrome* [T047]
  - https://uts.nlm.nih.gov//metathesaurus.html

An important note

- These resources contain information that can help characterise medical language
  - Synonyms of a term
  - Relationship between terms/concepts
- Rarely do these resources contain information that directly answers questions like
  - What is the drug of choice for condition x?
  - What is the cause of symptom x?
  - What test is indicated in situation x?
  - How should I treat condition x (not limited to drug treatment)?
  - How should I manage condition x (not specifying diagnostic or therapeutic)?
  - What is the cause of physical finding x?
  - What is the cause of test finding x?
  - Can drug x cause (adverse) finding y?
  - Could this patient have condition x?
- That is, they **do not directly resolve the clinical questions** presented in [Ely et al., 2000] taxonomy
- They capture truisms/**universal facts**, not subjective knowledge/things that could change over time

Implicit Medical Concept Representations: Word Embeddings

- [Pyysalo et al., 2013]: word2vec and random indexing on very large corpus of biomedical scientific literature. http://bio.nlplab.org
- [De Vine et al., 2014]: word2vec on medical journal abstracts (embedding for UMLS)
  - Learns embedding of a concept, from co-occurrence with concepts
- [Zuccon et al., 2015, b]: word2vec on TREC Medical Records Track. http://zuccon.net/ntlm.html
- [Choi et al., 2016]: word2vec on medical claims (embedding for ICD), clinical narratives (embedding for UMLS) https://github.com/clinicalml/embeddings

Implicit Medical Concept Representations: Word Embeddings

- [Beam et al., 2018]: cui2vec (variation of word2vec) on 60M insurance claims + 20M health records + 1.7M full text biomedical articles. https://figshare.com/s/00d69861786cd0156d81

- [Miftahutdinov et al., 2017]: word2vec trained on online user-generated drug reviews (e.g., askapatient.com, amazon, webmd, etc): https://github.com/dartrevan/ChemTextMining/tree/master/word2vec

- Nuances of medical word embeddings:

    - [Chiu et al., 2016]: bigger corpora do not necessarily produce better biomedical word embeddings

Pointers to: Evaluation in Health Search

- Specific **evaluation challenges**: relevance and beyond
    - **relevance hard to asses**: vocabulary mismatch, temporality of relevance, dependent aspects, expertise influence perception of relevance
    - **dimensions of relevance** of key importance in certain health search tasks: understandability, trustworthiness.
- Evaluation campaigns, **collections** and resources (see table next)

Task	Dataset
Matching patient to clinical trials or trials to patients	1. TREC Medical Records Track [Voorhees&Hersh, 2012] 2. Clinical Trials Test Collection [Koopman&Zuccon, 2016] 3. MIMIC-III: dataset of patient records [Johnson et al., 2016]
Consumer Health Search	1. CLEF eHealth Consumer Health Search Task [Zuccon et al., 2016] 2. FIRE 2016 Consumer Health Information Search
Evidence-based Medicine & Clinical Decision Support (CDS)	1. TREC Genomics Track 2. TREC Clinical Decision Support [Simpson et al, 2014] 3. TREC Precision Medicine Track [Roberts et al., 2017]
Compilation of systematic reviews	1. Systematic review test collection [Scells et al., 2017] 2. CLEF eHealth Technology Assisted Review 2017 [Kanoulas et al., 2017]
Image Retrieval	ImageCLEF [Muller et al., 2010]
Identifying concepts from free-text	1. Annotated “problems”, “tests” & “treatments” 2. Annotated SNOMED concept

Good lessons from evaluation campaigns

- **Retrieval of health records for cohort selection** (TREC Medical Records [Edinger et al., 2012])
  - Both **precision** and **recall errors** due to **incorrect lexical representations and lexical mismatches**
    - Non-relevant visits were most often retrieved because they contained a non-relevant reference to the topic terms
    - Relevant visits were most often infrequently retrieved because they used a synonym for a topic term
- Other issues: time factors, negation detection, overlap in terminology between conditions or procedures (hearing loss vs hearing aid)

Good lessons from evaluation campaigns

- **Retrieval of evidence based medicine**
  (TREC CDS [Roberts et al., 2016], analysing 2014 results)

  - How to best to use **concept extraction** system such as MetaMap of key importance: can easily become a red herring

  - **Negation and attribute extraction** (age, gender, etc.) intuitively important, but best systems did not use them
    If negation extraction, soft-matching strategy best

  - **article preference** to identify appropriate articles for Diagnosis, Treatment, and Test (fundamental mismatch b/w irrelevant articles and clinical important attributes)

    - Methods tried did not work: specialised lexicons, MeSH terms, and machine learning classifiers

Good lessons from evaluation campaigns

[Karimi et al., 2018] provides **platform** to facilitate experimentation and hypothesis testing

- Can tease-out which components provide improvements
  - query and document expansion (UMLS), word embeddings, negation detection/removal, LTR
- Main findings on TREC CDS
  - **Articles body** contributes to retrieving over 50% of relevant results
  - adding UMLS concepts does not improve retrieval using titles only
  - concepts in abstracts slightly improved retrieval for queries built using Desc and Sum, but not Note
  - **PRF** works well, also in combination with **word embeddings**; but **LTR** can outperform all these

Closing remarks

Open challenges

- Ethics and sharing of data — privacy concerns vs need for large scale evaluation
- Integration of data driven and symbolic representations
- Inference with knowledge graphs
- Query understanding
- Results presentation

} require personalisation, context understanding, better user understanding

- Translation of IR for impact on health

Where to go for help?

- Content from this lecture: https://github.com/ielab/afirm2019-health-search
- Content from previous versions of this tutorial (full day):, e.g. https://ielab.io/russir2018-health-search-tutorial/
- Bibliography of all literature mentioned here
- Hersh's book: "Information Retrieval: A Health and Biomedical Perspective"