SHARPS

Strategic Healthcare IT Advanced Research Projects on Security

STRATEGIC HEALTH IT ADVANCED RESEARCH PROJECTS ON SECURITY

sharps.org

Stanford

UNDERSTANDING THE CHALLENGES WITH MEDICAL DATA SEGMENTATION

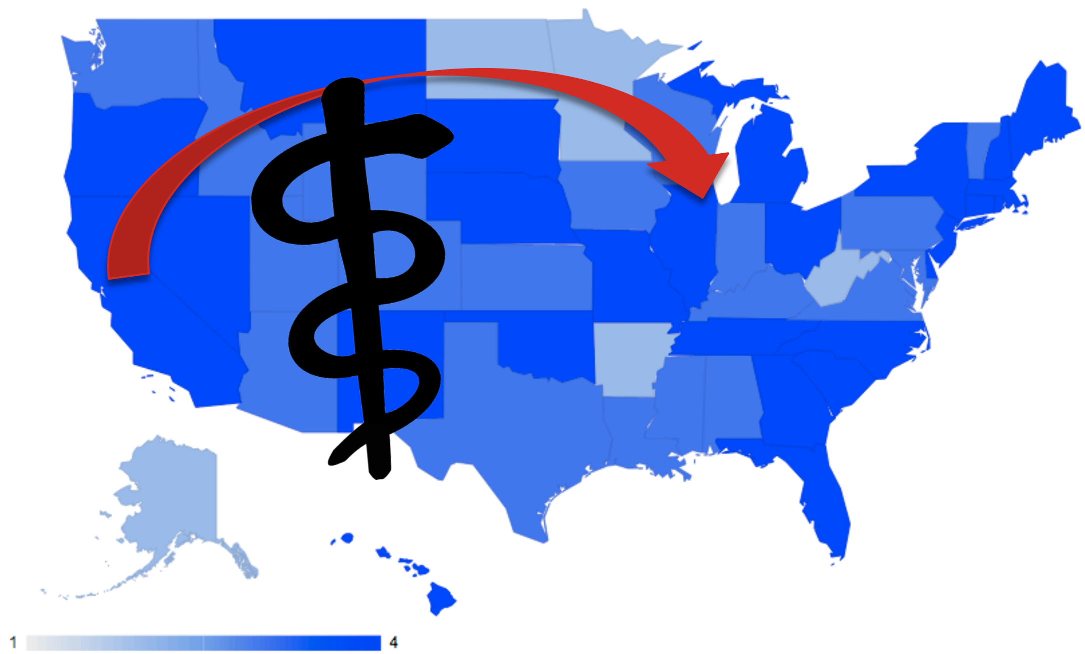
Ellick M. Chan, Peifung E. Lam and John C. Mitchell

8/12/13

Health Information Exchange (HIE)

SHARPS

- Federal
  - HIPAA
  - HITECH
- State laws on
  - Mental Health
  - Substance Abuse
  - STDs
  - Genetic testing
- Organizational



Health Information Exchange Cloud

Compliance approaches

SHARPS

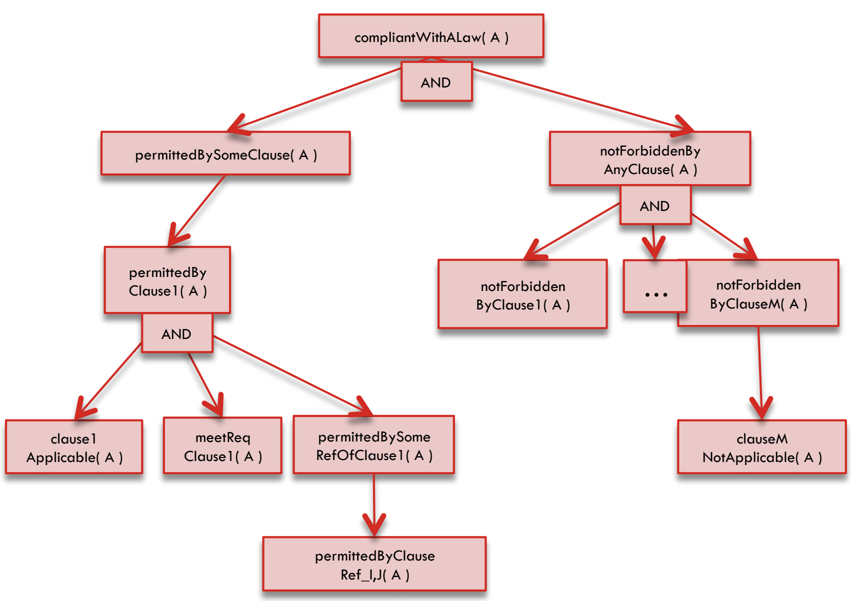
Automated Policy

HIPAA Law

§ 164.502 Uses and disclosures of protected health information: general rules.
(a) Standard. A covered entity may not use or disclose protected health information, except as permitted or required by this subpart or by subpart C of part 160 of this subchapter.
(1) Permitted uses and disclosures. A covered entity is permitted to use or disclose protected health information as follows:
(i) To the individual;
(ii) For treatment, payment, or health care operations, as permitted by and in compliance with §164.506;
(iii) Incident to a use or disclosure otherwise permitted or required by this subpart, provided that the covered entity has complied with the applicable requirements of §164.502(b), §164.514(d), and §164.530(c) with respect to such otherwise permitted

```
>|
%%Standard rules for "uses and disclosures"
permitted_by_164_502_a(A) :-
  is_from_coveredEntity(A),
  is_phi(A),
  (permitted_by_160_C(A);
  permitted_by_164_502_a_1(A);
  required_by_164_502_a_2(A)).

permitted_by_164_502_a_1(A):-
  permitted_by_164_502_a_1_i(A);
  permitted_by_164_502_a_1_ii(A);
  permitted_by_164_502_a_1_iii(A);
  permitted_by_164_502_a_1_iv(A);
  permitted_by_164_502_a_1_v(A);
  permitted_by_164_502_a_1_vi(A).
```



Data segmentation

Health Record

- Medications
- Previous diagnoses
- Labs

Sensitive conditions

According to research by the California HealthCare Foundation, 15 percent of patients who know their information will be shared would hide information from their doctor, and another 33 percent would consider hiding information[1].

Example

Medications

- 1. Tylenol
- 2. Sudafed
- 3. AZT
- 4. Bactrim

Problem List

- 1. Headache
- 2. Sinus Infection
- 3. HIV positive
- 4. UTI

Letter
I hope you and your partner had a great weekend in Provincetown and that the thrush has improved with the mouthwash sample I gave you.

Adapted from J. Halamka, 2012

Example

Medications

- 1. Tylenol
- 2. Sudafed
- 3. AZT
- 4. Bactrim

Letter
I hope you and your partner had a great weekend in Provincetown and that the thrush has improved with the mouthwash sample I gave you.

Hide HIV/AIDS ICD-9 code 042

Problem List

- 1. Headache
- 2. Sinus Infection
- **3. ~~HIV positive~~**
- 4. UTI

Adapted from J. Halamka, 2012

Example

Medications

- 1. Tylenol
- 2. Sudafed
- **3. ~~AZT~~** +side effects
- 4. Bactrim

Problem List

- 1. Headache
- 2. Sinus Infection
- 3. ~~HIV positive~~
- 4. UTI

Letter
I hope you and your partner had a great weekend in Provincetown and that the thrush has improved with the mouthwash sample I gave you.

Zidovudine (INN) or azidothymidine (AZT) is a type of antiretroviral drug used for the treatment of **HIV/AIDS**.

Side effects: anemia, neutropenia, hepatotoxicity, cardiomyopathy, and myopathy

Adapted from J. Halamka, 2012

Example

SHARPS

Medications

- 1. Tylenol
- 2. Sudafed
- 3. ~~AZT~~ +side effects
- **4. Bactrim**

Problem List ?

- 1. Headache
- 2. Sinus Infection
- 3. ~~HIV positive~~
- 4. UTI

Letter

I hope you and your partner had a great weekend in Provincetown and that the thrush has improved with the mouthwash sample I gave you.

Trimethoprim/sulfamethoxazole or **co-trimoxazole** (abbreviated SXT, TMP-SMX, TMP-SMZ or TMP-sulfa) is a sulfonamide antibiotic combination of trimethoprim and sulfamethoxazole, in the ratio of 1 to 5, used in the treatment of a **variety of bacterial infections.**

Adapted from J. Halamka, 2012

Example

Medications

- 1. Tylenol
- 2. Sudafed
- 3. ~~AZT~~ +side effects
- 4. Bactrim

Problem List

- 1. Headache
- 2. Sinus Infection
- 3. ~~HIV positive~~
- 4. UTI

Letter

I hope you and your partner had a great weekend in Provincetown and that the thrush has improved with the mouthwash sample I gave you.

Prophylaxis (preventative med) for immunocompromised patient?

Patient has urinary tract infection (UTI), plausibly deniable case.

Adapted from J. Halamka, 2012

Example

SHARPS

Medications

- 1. Tylenol
- 2. Sudafed
- 3. ~~AZT~~ +side effects
- 4. Bactrim

Problem List

- 1. Headache
- 2. Sinus Infection
- 3. ~~HIV positive~~
- 4. UTI

Letter

I hope you and your **partner** had a great weekend in **Provincetown** and that the thrush has improved with the mouthwash sample I gave you.

Highest rate of same-sec couples in Provincetown, MA.

Karen Christel Krahulik, *Provincetown: From Pilgrim Landing to Gay Resort*, NYU Press, 2007, p. 51.

Adapted from J. Halamka, 2012

Example

SHARPS

Medications

- 1. Tylenol
- 2. Sudafed
- 3. ~~AZT~~ +side effects
- 4. Bactrim

Problem List

- 1. Headache
- 2. Sinus Infection
- 3. ~~HIV positive~~
- 4. UTI

Letter
I hope you and your partner had a great weekend in Provincetown and that the **thrush** has improved with the **mouthwash** sample I gave you.

Candidiasis (thrush) - **Candidiasis** or **thrush** is a fungal infection (mycosis). Commonly causes mouth yeast infections, which manifest as white patches in the mouth. **15% of immuno-compromised** patients may develop this.

Adapted from J. Halamka, 2012

Example

Medications

- 1. Tylenol
- 2. Sudafed
- 3. ~~AZT~~ +side effects
- 4. Bactrim

Letter

I hope you and your partner had a great weekend in Provincetown and that the thrush has improved with the mouthwash sample I gave you.

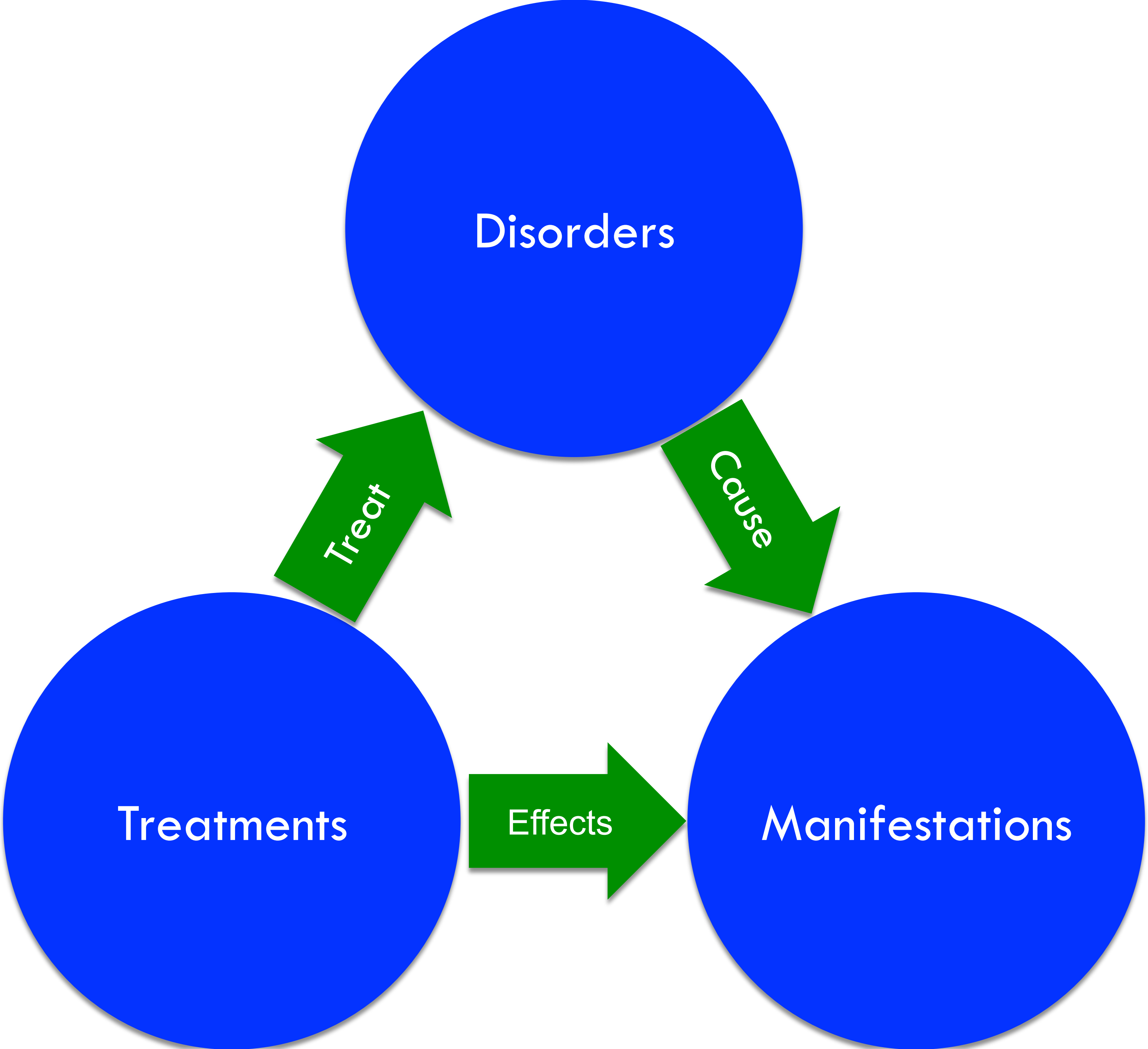
Headaches & HIV: 24/535 patients – 4.5%
CDC NHDS 2010 dataset. 115,000 patients.

Mononucleosis-like symptoms

Problem List

- 1. Headache
- 2. Sinus Infection
- 3. ~~HIV positive~~
- 4. UTI

Adapted from J. Halamka, 2012



Disorders
Treat
Cause
Treatments
Effects
Manifestations

Threat Model

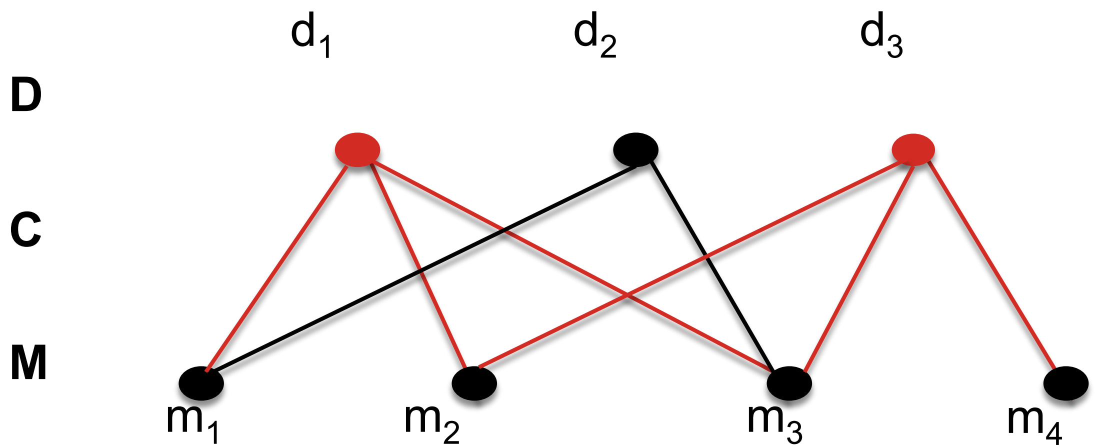
- Attacker has direct access to redacted health record, medical literature
- Attacker does not have the computational capability to circumvent security mechanisms that protect the primary sensitive codes

Treatments

Concept	Description	Links	Notes
Risperidone	Treats schizophrenia, bipolar disorder, and autism.	schizophrenia, bipolar disorder, autism, weight gain, insomnia, alopecia	Use of Risperidone usually implies treatment of a mental health disorder.
Carbamazepine	Anti-convulsant and mood-stabilizing drug. Treats epilepsy and bipolar disorder.	epilepsy, bipolar disorder, headaches, drowsiness	Primarily used to treat mental health disorders. Could be used off-label to treat Complex regional pain syndrome(ICD9: 337.21)
Citalopram	Primarily used as an SSRI to treat depression. Can also be used to treat hot flashes.	depression, hot flashes, anorgasmia, nausea, diarrhea	Can treat both sensitive and non-sensitive conditions.
Lamotrigine	Primarily used as an anticonvulsant drug to treat epilepsy and bipolar disorder. Can also treat migraines.	epilepsy, bipolar disorder, migraines	Can be used to treat mental health disorders or migraines.

Formal model

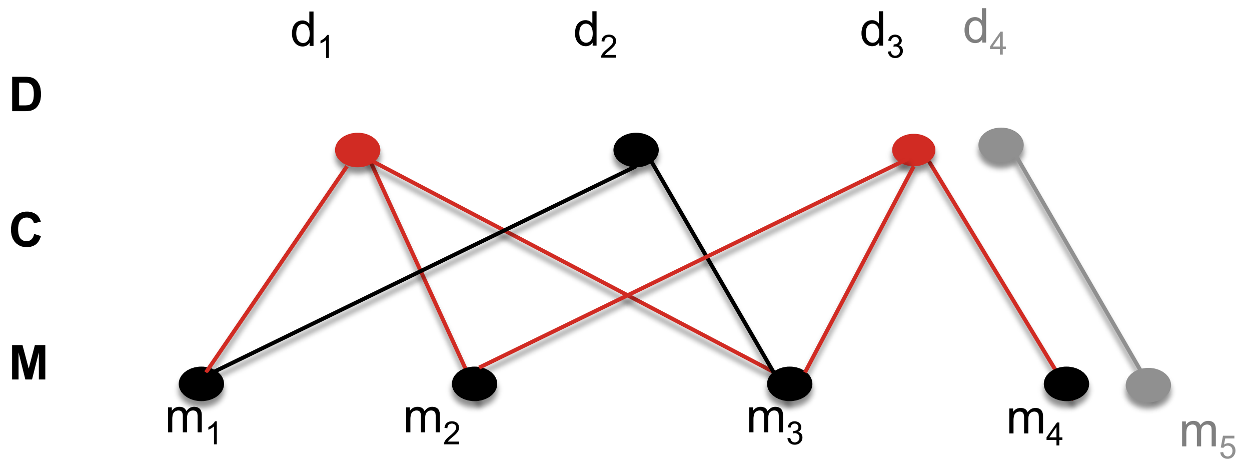
Hypothesis
$\{d_1, d_3\}$
$\{d_2, d_3\}$
$\{d_1, d_2, d_3\}$
$\{d_1, d_2\}$



Reggia's set cover model
- Plausibility – set cover
- Likelihood – Occam's razor and fitness

Formal model

Hypothesis
$\{d_1, d_3\}$
$\{d_2, d_3\}$
$\{d_1, d_2, d_3\}$
$\{d_1, d_2\}$



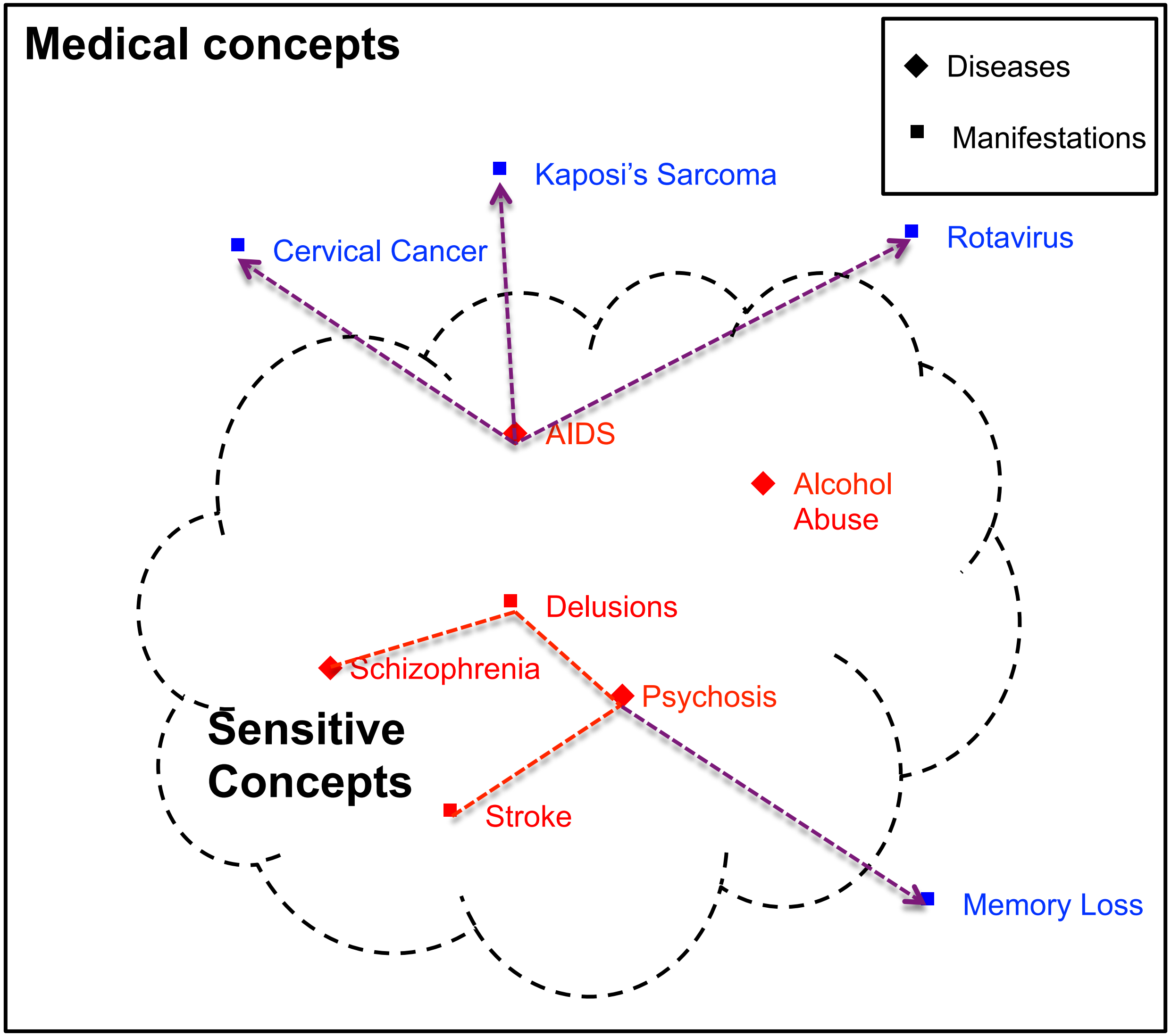
Reggia's set cover model

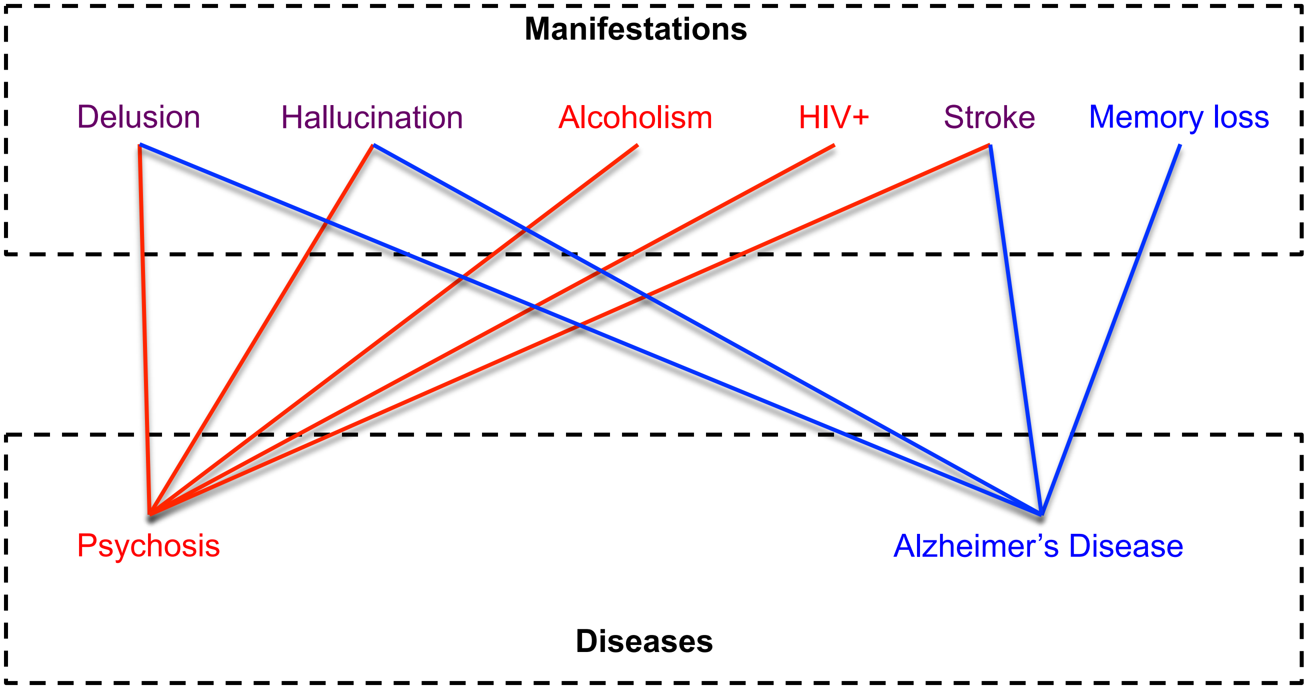
- Plausibility – set cover
- Likelihood – Occam's razor and fitness

Explanation of manifestations

- Best explanation E of manifestations:
  - Covers all observed manifestations M+
  - Is the simplest (parsimonious) definition

- Heuristics for “best cover”
  - Minimality - |E| is minimal
  - Criticism: minimal cardinality covers can be too restrictive
    - Occam’s razor vs Hickam’s dictum
  - Irredundancy – removing any disorder results in a non-cover of M+
  - Relevancy – Every *d* in D must be causally associated with some *m* in M+


Medical concepts
Diseases
Manifestations
Kaposi's Sarcoma
Cervical Cancer
Rotavirus
AIDS
Alcohol Abuse
Delusions
Schizophrenia
Psychosis
Sensitive Concepts
Stroke
Memory Loss



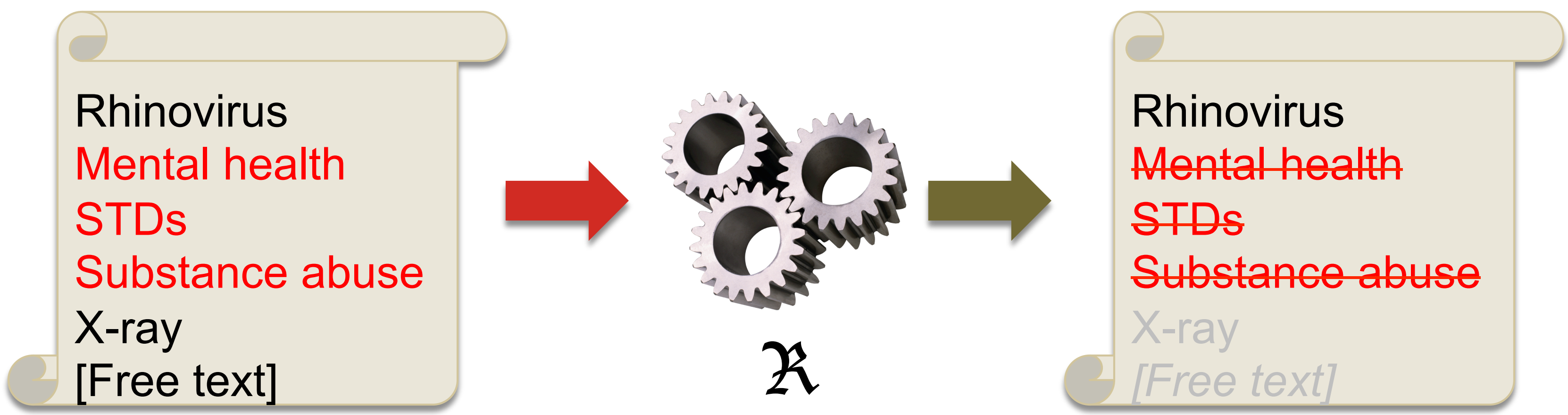
Source: PubMed, NIH.gov

Predicate-Reducer definition

SHARPS

A – Medical algorithm
π – Policy determines sensitive code s
M – Medical record
Predicate P(M, π) – Determines if $s \in M$
Reducer R(M, π) – Removes s from M

Ideal reducer $$A(m) = A(R_\pi(m))\ \forall m \in M$$



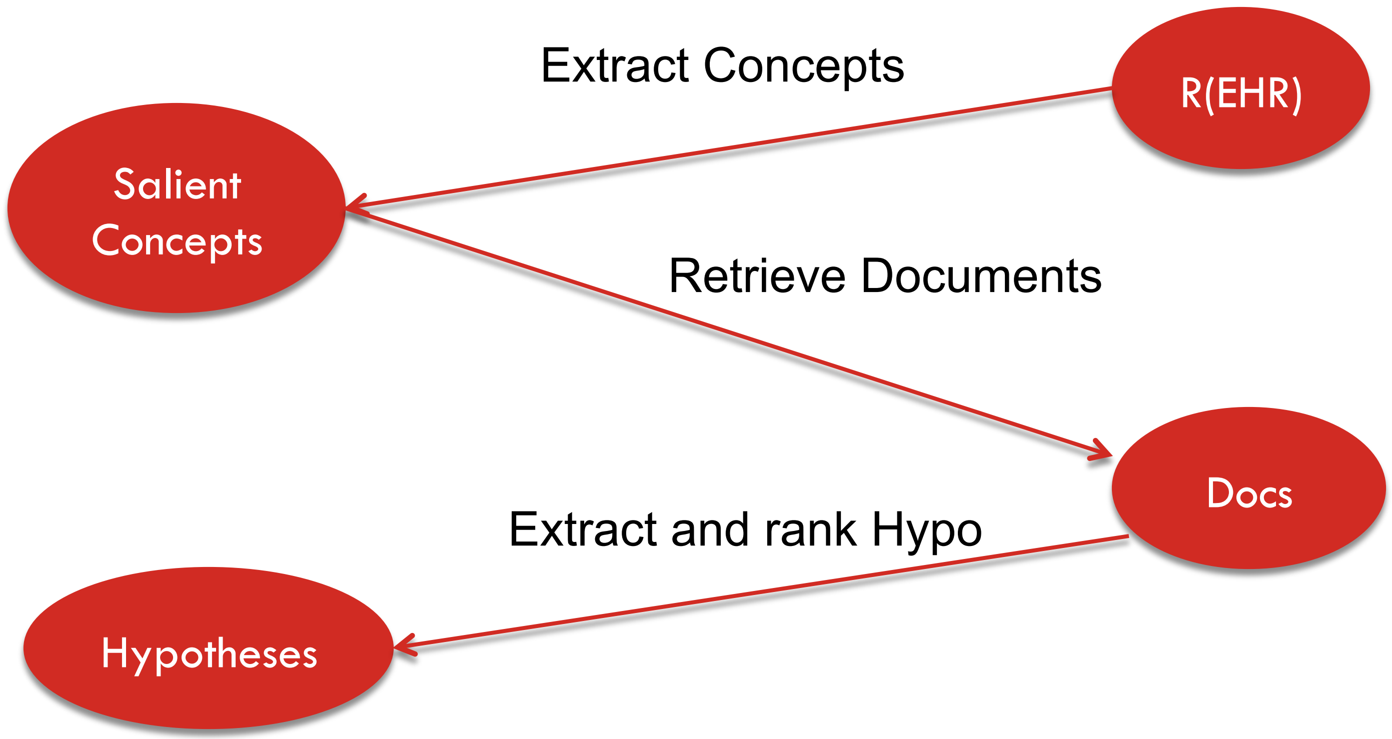
Inference approach

- Input: Reduce(Diseases U Manifestations U Treatments)
- Output: Inferred Diseases

- 1. For each input, evoke hypotheses
- 2. Evaluate hypotheses
- 3. Rank hypotheses according to fitness

- Hypothesis fitness
  - Competing hypotheses, e.g. d_1 or d_2

Algorithm overview



Algorithm overview


```
hypotheses ← ∅;
repeat
    query ← ∅;
    for j = 1 → numTerms do
        /* select a concept from the EHR using
           a probability distribution             */
        x ← select_concept(concept_probs, EHR)
        query ← query ⋃ x;
    end
    /* search for docs that contain the query
       terms                                      */
    sr ← search(query, knowledge_base) ;
    /* Identifies hypotheses from medical
       concepts in documents                      */
    hypotheses ← update_hyp(hypotheses, sr);
    /* Evaluates hypotheses according to
       plausibility criteria                      */
    results ← eval_hypotheses(hypotheses) ⋃ results;
until convergence;
rank(results);
```

Algorithm 1: Inference algorithm

Concept Support Index

SHARPS

Concept Support Index

Let $H \subseteq W$ be a set of concepts representing a hypothesis that the patient has had the medical manifestations, diseases, and treatments in H. Let $h \in H$ be a particular concept in H, then the Concept Support Index with respect to a medical knowledge document *doc* is defined as:

$$CSI(h, doc) = \frac{Count(h, doc)}{\sum_{w \in W} Count(w, doc)} \tag{1}$$

$$CSI(H, doc) = \sum_{h \in H} CSI(h, doc) \cdot w_h \tag{2}$$

, where $w_h \in [0, 1]$, $\sum_{h \in H} w_h = 1$, and $Count(h, doc)$ counts the number of occurrences of h in *doc*.

Hypothesis Fitness Index

SHARPS

$$HFI(H, Docs) = \sum_{doc \in Docs} CSI(H, doc) \cdot weight(doc, H) \tag{3}$$

where $weight(H, doc)$ is a weighting function. One such function could be BM25 [20, 30, 34], which is defined as

$$\mathrm{BM25}(D, Q) = \sum_{q_i \in Q} \mathrm{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{\mathrm{avgdl}})}, \tag{4}$$

where

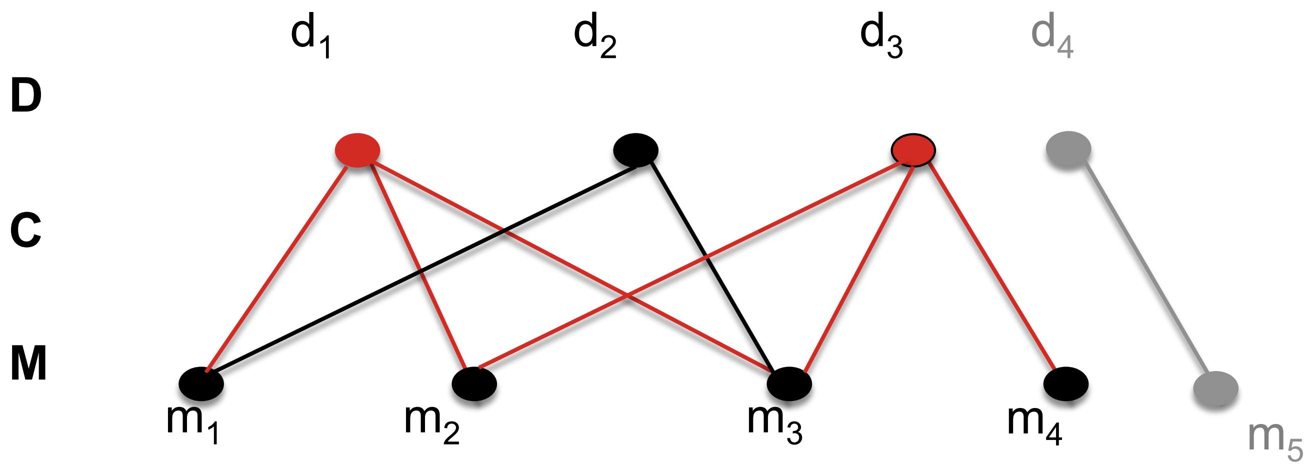
$$\mathrm{IDF}(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}, \tag{5}$$

$f(q_i, D)$ is the term frequency of q_i in D, $k_1 \in \mathbb{R}^+$, $b \in [0, 1]$, and $avgdl$ is the average document length of $Docs$.

Results

Condition	Query	Results	Medical codes	Notes
Rett Syndrome	“wringing” AND “female” AND “constipation” AND ”scoliosis”	3 articles suggest Rett Syndrome.	F84.2, R09.0, K59.0, 737.0	Pubmed
Rett Syndrome	“wringing” AND “female” AND “constipation” AND ”scoliosis”	1.73M results, 5 of top 10 results suggest Rett Syndrome, including NIH Medline.	F84.2, R09.0, K59.0, 737.0	Google
AIDS	"Toxoplasmosis" AND "Hepatitis B" AND "Encephalopathy" AND "Progressive multifocal leukoencephalopathy" AND "Cryptococcosis”	140,000 results. 5 of top 10 suggest AIDS.	130, 070.2, 348.30, 046.3, 117.5	Google
AIDS	…	18,000 results. >8 of top 10 suggest AIDS.	130, 070.2, 348.30, 046.3, 117.5	Bing

Possible defenses



- Deniability through relative strengths of hypotheses
  - Hide non-sensitive EHR as well
  - Enhance competing hypothesis, e.g. Citalopram
  - Association rule hiding

A message from our sponsors…

Strategic Healthcare IT
Advanced Research
Projects on Security

We thank:

- Carl Gunter and Mike Berry - predicate-reducer model
- James Reggia - formalization of the hypothetico-deductive model
- Brad Malin - helpful resources
- Ivan Handler - health information exchange level
- Fisayo Ositelu - medical insight.

Questions?

SHARPS

Ask your doctor!