HKU Med LKS Faculty of Medicine
Department of Orthopaedics & Traumatology
香港大學矯形及創傷外科學系

Understand clinical data science infrastructure on Electronic Health Records

Dr Teng Zhang

Digital Health Laboratory
Department of Orthopaedics and Traumatology
University of Hong Kong

1

Confidential information all rights belongs to Department of O&T, HKU, Hong Kong

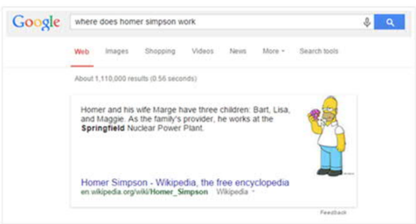
HKU Med LKS Faculty of Medicine
Department of Orthopaedics & Traumatology
香港大學矯形及創傷外科學系

Contents

2

What is Knowledge Graph

Knowledge Graph is a knowledge base that uses a graph-structured data model or topology to integrate data. Knowledge graphs are often used to store interlinked descriptions of entities – objects, events, situations or abstract concepts – with free-form semantics.



Knowledge Graph answer driven by the entity-based search associations.

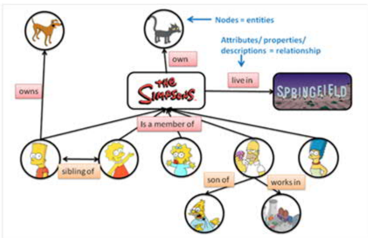


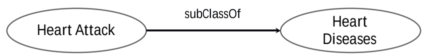
Illustration of a semantic taxonomy of entities (as nodes) and attributes (relationships).

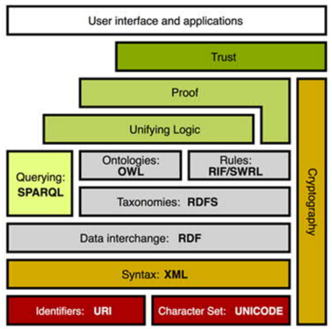
Knowledge Graph Representation

3

Knowledge Representation at the core of Semantic Web

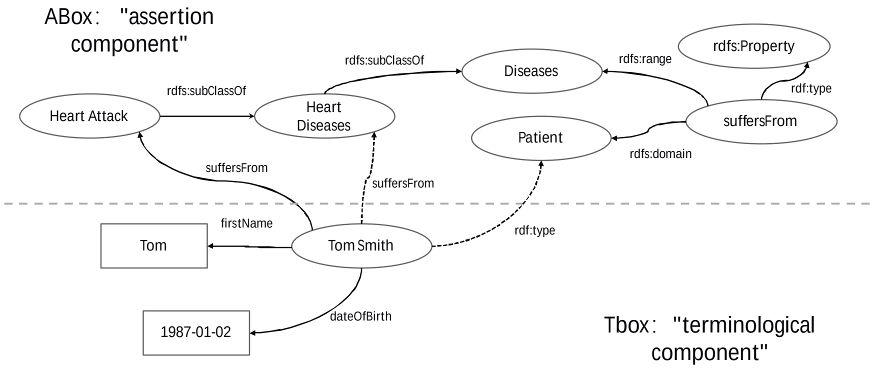
- Ontology: Describe the concepts and their relationships in a particular domain
- Computation: reasoning (set / rule)





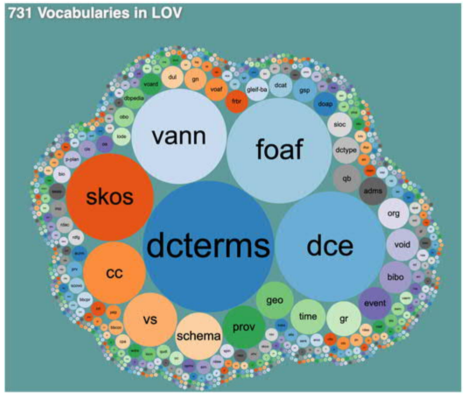
4

Knowledge Graph and Subclasses



5

Linked Open Vocabularies

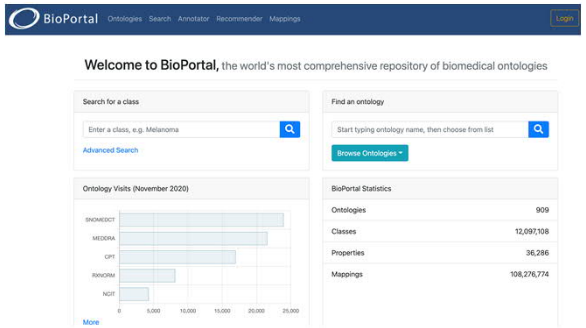


https://lov.linkeddata.es/dataset/lov

Open and linked TBox

6

BioPortal

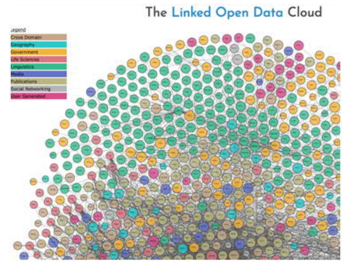


Biomedical TBoxes

https://bioportal.bioontology.org/

7

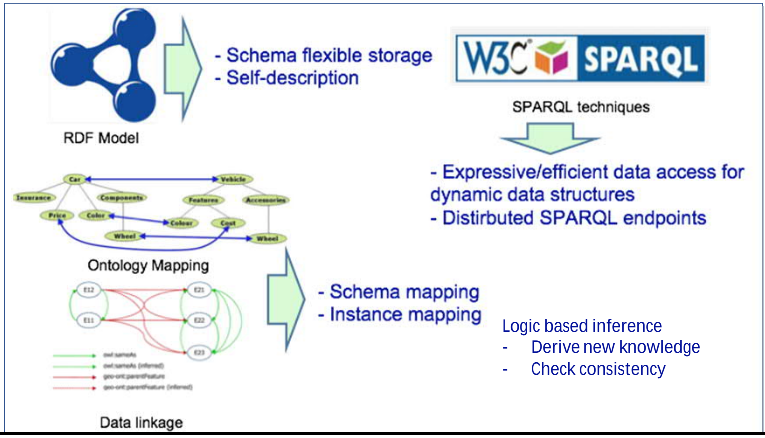
Open A-Boxes



The dataset currently contains 1269 datasets with 16201 links (as of May 2020)

8

Semantic Web for Big Data: technical contribution



9

HKU Med LKS Faculty of Medicine
Department of Orthopaedics & Traumatology
香港大學矯形及創傷外科學系

Contents

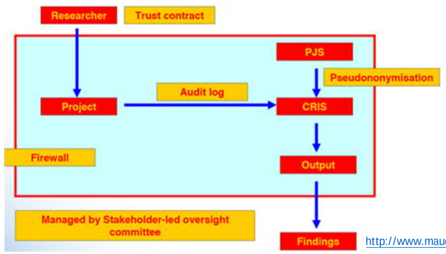
10

10

History: the Clinical Record Interactive Search (CRIS)

South London and Maudsley Psychiatric Hospital

- Oldest (founded in 1247)
- Largest EU Mental health provider



- assembles de-identified SLaM EHRs on 324k patients since 2006, containing over 25gb text fields and 19m rows
- provides authorised researchers with regulated, secure access
- ~100 published studies since 2009

http://www.maudsleybrc.nihr.ac.uk/about-us/core-facilities/clinical-record-interactive-search-cris/

11

The Width of Clinical Data

Health records

Diagnosis, procedures, outcomes

Images

MRI, CT, US, X-ray, Pathology

Drugs

OTC medication

Hospital Pharmacy

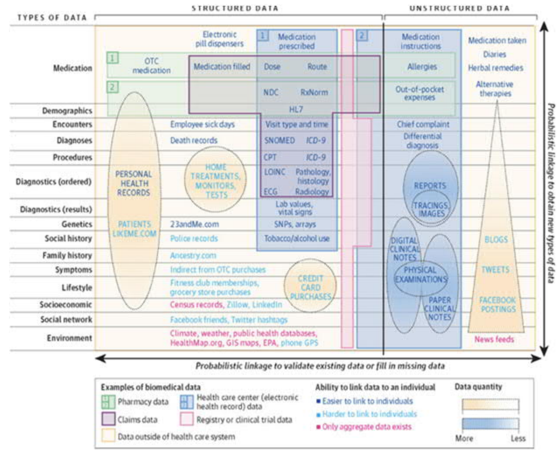
Costs

Income and costs per encounter/event

Medical devices

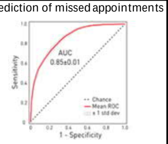
CRT, DBS, etc

Genomic Data



12

Clinical Data Science - Actionable Analytics



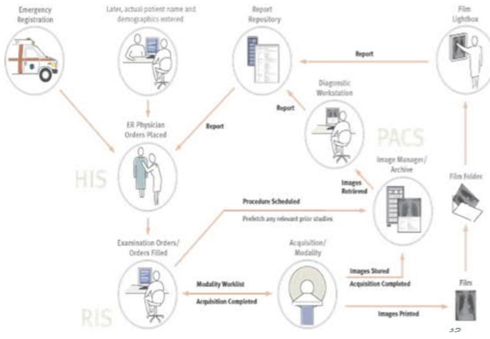
Improving Clinical Care
Extract measurement or predict outcomes given some input data
Baseline Diagnosis, Prognosis, Drug/Intervention Response

Clinical Efficiency
Automate labour-intensive tasks
Data Preparation, Quantification, Prioritisation, Integration

Value Based Care
Optimise the value (outcome/cost) of the full care pipeline
Optimise hospital-specific care pathways

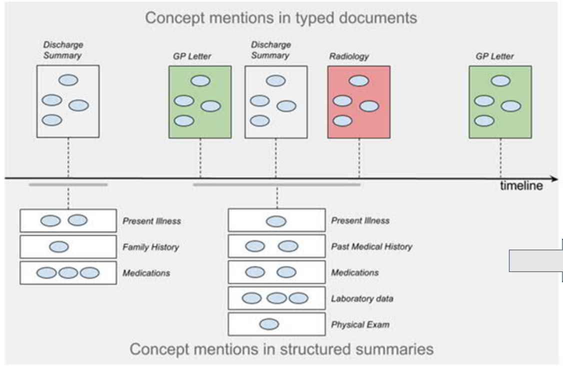
Clinical Auditing
Continuous assessment of quality of care
Achieve local CCG standards and performance targets

Operational Efficiency
Clinical data can inform administrative decisions
Optimise bed utilisation, cost codes, DNA, etc.



13

Patient Profiling



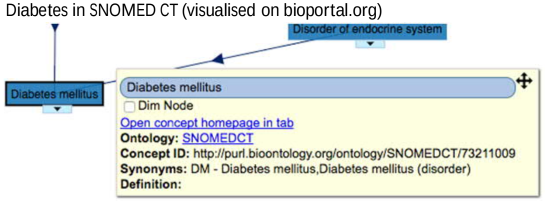
Vital Signs and other measurements

Lipase	47 ...	Leukocytes	10.8 ...	Erythrocytes	3.97 ...
Hematocrit	30.9 ...	HDL	4.3 ...	Cholesterol	4.3 ...
LDL	134	Calculated	134	Alanine Aminotrai	205 ...
serum AST	270 ...	Lactate Dehydroge	487	Phosphate	117
Bilirubin	0.6	Lactate	1.1 ...	Calcium	8.2 ...
Sanguis	47 ...	Mean cell hemoglc	25.1	MCHC	32.3
RDW	14.6	Primed lymphocyt	512	Blood Sugar	162
K 4	97	Bicarbonate	29	Amylase	55
Triglycerides	186	Norvasc	10	Prednisone	60
Blood culture	2200				

14

Globally unique identifiers

Diabetes in SNOMED CT (visualised on bioportal.org)



Global unique IDs for concepts (classes and relations)

- Techniques: IRI in Semantic Web; database IDs: ICD10CM:E08-E13, DOID:9351
- Functions: data integration

15

15

Domain vocabulary

has_exact_synonym	IDDM insulin-dependent diabetes mellitus type I diabetes mellitus
has_obo_namespace	disease_ontology
id	DOID:9744
label	type 1 diabetes mellitus
notation	DOID:9744
prefLabel	type 1 diabetes mellitus

Type 1 diabetes in Human Disease Ontology

- A set of labels with each concept
- Techniques: multiple labels, multilingual, synonyms, prefLabel
- Functions: Named Entity Recognition

16

16

Textual definitions and metadata

MEDLINEPLUS Definition:

Diabetes is a disease in which your blood glucose, or blood sugar, levels are too high. Glucose comes from the foods you eat. Insulin is a hormone that helps the glucose get into your cells to give them energy. With type 1 diabetes, your body does not make insulin. With type 2 diabetes, the more common type, your body does not make or use insulin well. Without enough insulin, the glucose stays in your blood. You can also have prediabetes. This means that your blood sugar is higher than normal but not high enough to be called diabetes. Having prediabetes puts you at a higher risk of getting type 2 diabetes.

Over time, having too much glucose in your blood can cause serious problems. It can damage your eyes, kidneys, and nerves. Diabetes can also cause heart disease, stroke and even the need to remove a limb. Pregnant women can also get diabetes, called gestational diabetes.

A blood test can show if you have diabetes. Exercise, weight control and sticking to your meal plan can help control your diabetes. You should also monitor your glucose level and take medicine if prescribed.

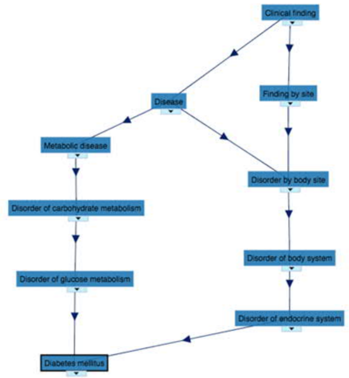
NIH: National Institute of Diabetes and Digestive and Kidney Diseases

- ➢ The provision of precise information for concept
- ➢ Techniques: textual attributes and other metadata
- ➢ Functions: clear and unambiguous description of concepts – allow computers to capture the semantics via word/phenotype embeddings.

17

17

Machine understandable semantics



Representing knowledge in a way that machines can "understand" and do inferences

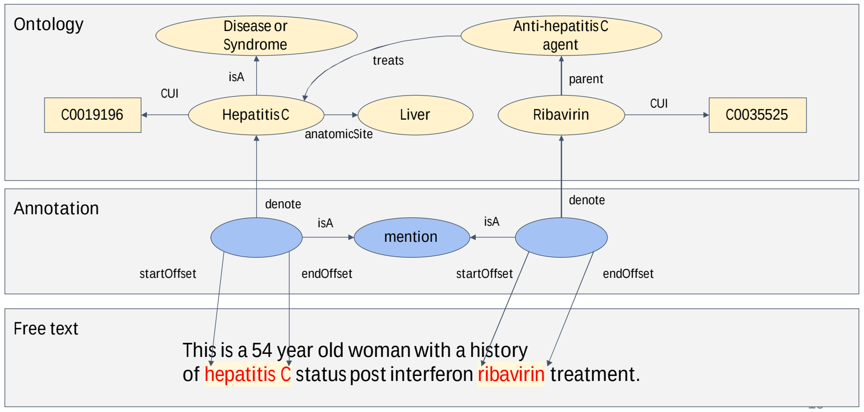
- Techniques: logic foundations (description logic), Semantic Web technique stack
- Functions: consistency checking, inference, query answering, graph analytics

Diabetes in SNOMED CT (visualised on bioportal.org)

18

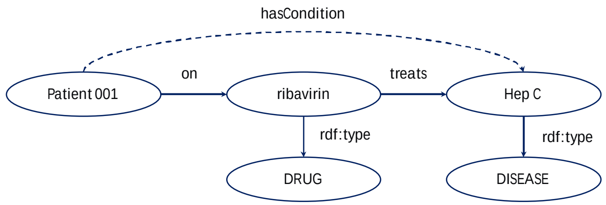
18

Semantic Annotations over EHR free texts



19

Beyond subClass inference...

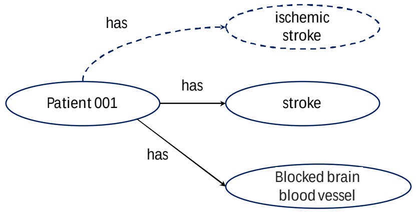


Rule1: if someone is on a DRUG that TREATS a condition, then he/she probably has that condition

20

20

Beyond subClass inference…



Rule2: if someone has a STROKE and also BLOCKED BRAIN BLOOD VESSEL, then the stroke is probably ischemic

21

21

HKU Med LKS Faculty of Medicine Department of Orthopaedics & Traumatology 香港大學矯形及創傷外科學系

Contents

22

22

Use case 1 - NHS Trust Clinical & Operational Support

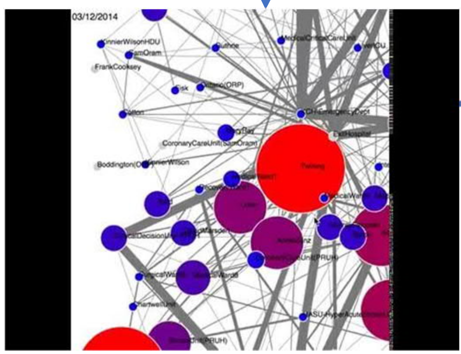
Real-time analysis and dashboarding

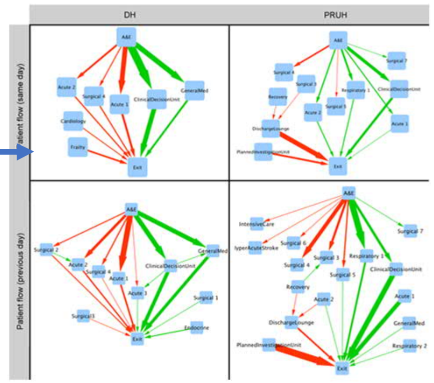
23

Patient flow

Key sub-networks for A&E performance

Patient id	From ward	To ward	Time
238492	A&E	Assessment	01/01/16 @ 10:23:04
238492	Assessment	Exit	01/01/16 @ 22:19:46
58204818	A&E	Stroke unit	01/01/16 @ 10:23:55





Bean et al. 2017 PLoS One

24

Current real-time clinical use cases

- e.g. Alerts for
  - Abnormal pathology results for rheumatology patients on Methothrexate
  - High-CCA antibodies suggesting pre-clinical Rheumatoid Arthritis
  - Patients being discharged on anticoagulants without being referred to Anticoagulation Clinic
  - Patients with recorded Atrial Fibrillation on their heart recorder text files
- Identifying frequent re-attenders with dementia
- Identifying high-cost drug use

How to alert?
- Email alerts
- Wi-fi Phone alerts

How about Closed-Loop Alerts?
(Alerts that arrive directly into clinical record)
Some vendor systems are closed proprietary systems impeding localised development.

25

Use case 2 – Support research using EHRs on the fly

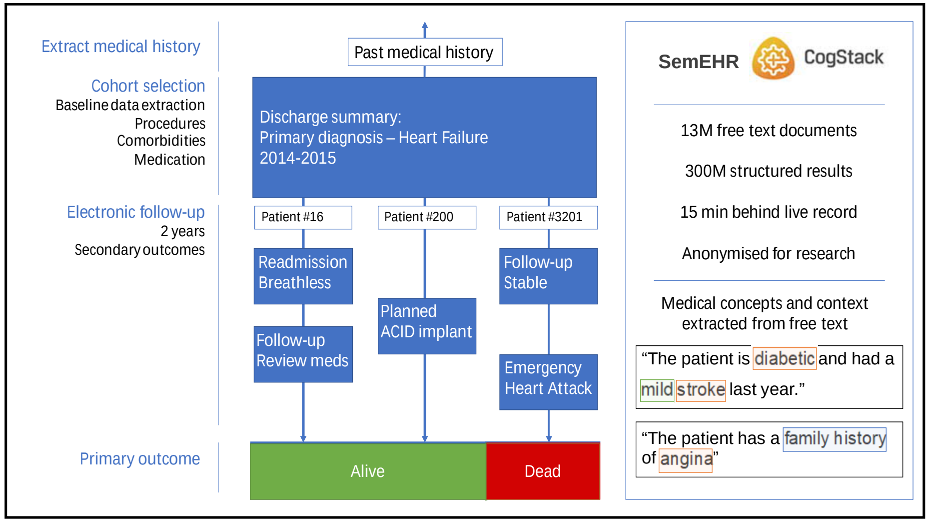
Very quick data exploration & analysis

26

Quick exploratory data analysis

- a cardiology grant application
- two outcomes (alive/dead) for all heart failure patients over a two-year period
- can be done in a few hours with proper access to a UK HER called CogStack

27



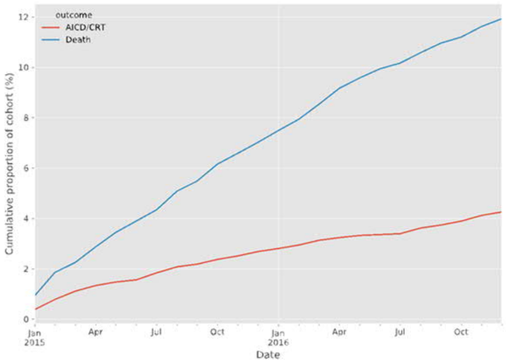
28

Heart failure outcomes

HF patients 2014-15
4625 patients

Follow up 2016/01–2017/12
10,709 documents
12 with AICD implanted (0.26%)
187 with CRT implanted (4.04%)
552 died (11.27%)

Unstructured routine data to cohort
Total time: 3 hours



29

Use case 3 – Clinical study: Prescribing Choice

Automated risk scores at scale

30

Clinical Motivation

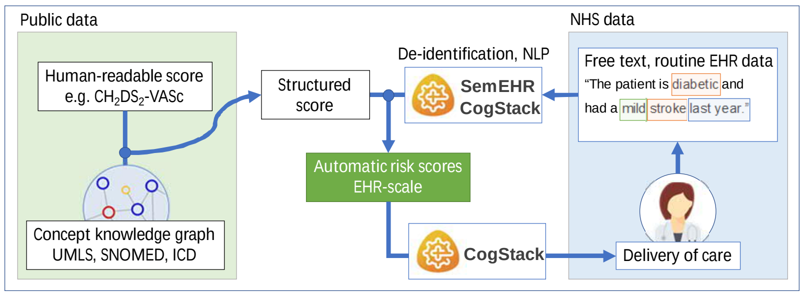
- Atrial Fibrillation (AF) is a heart condition causing an irregular heart rate
- Affects approx. 1m people in the UK
- AF patients have high stroke risk
  - Calculate risk with CH_2DS_2-VASc score
  - Prescribe based on threshold (1 for male, 2 for female)
- What proportion of AF patients are being prescribed an OAC
- Has that proportion been affected by the introduction of new drugs?
- Could the proportion be improved somehow?

31

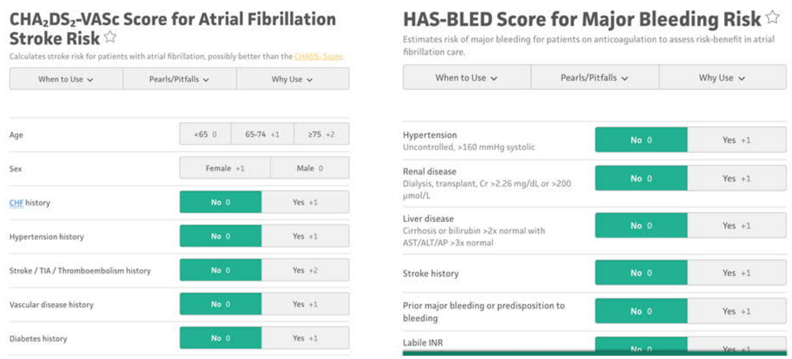
Risk scores at scale: pipeline

Dan Bean, Paul Scott, James Teo

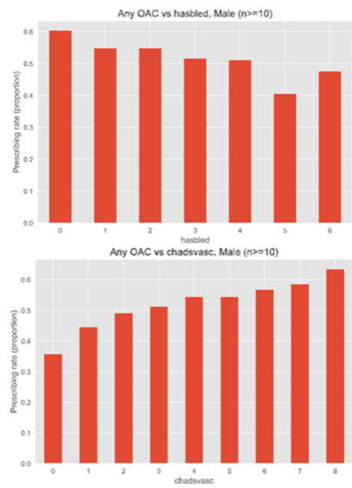
Atrial fibrillation at KCH
CHA_2DS_2-VASc
HAS-BLED
Frailty (Gilbert 2018)



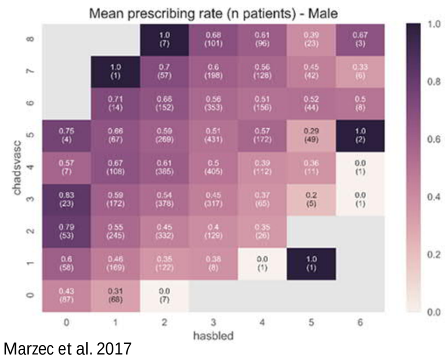
32



33



Prescribing increases with stroke risk and decreases with bleeding risk



Marzec et al. 2017
National Cardiovascular Data Registry PINNACLE registry
655,000 patients with nonvalvular AF and a CHA2DS2-VASc score of >1

34

Huayu Zhang, Amy Ferguson, Grant Robertson, Muchen Jiang, Teng Zhang, Cathie Sudlow, Keith Smith, Kristiina Rannikmae, Honghan Wu. Benchmarking network-based gene prioritization methods for cerebral small vessel disease. Accepted by Briefings in Bioinformatics, January 2021.

https://dx.doi.org/10.1007/978-3-319-45654-6

Thank you

35