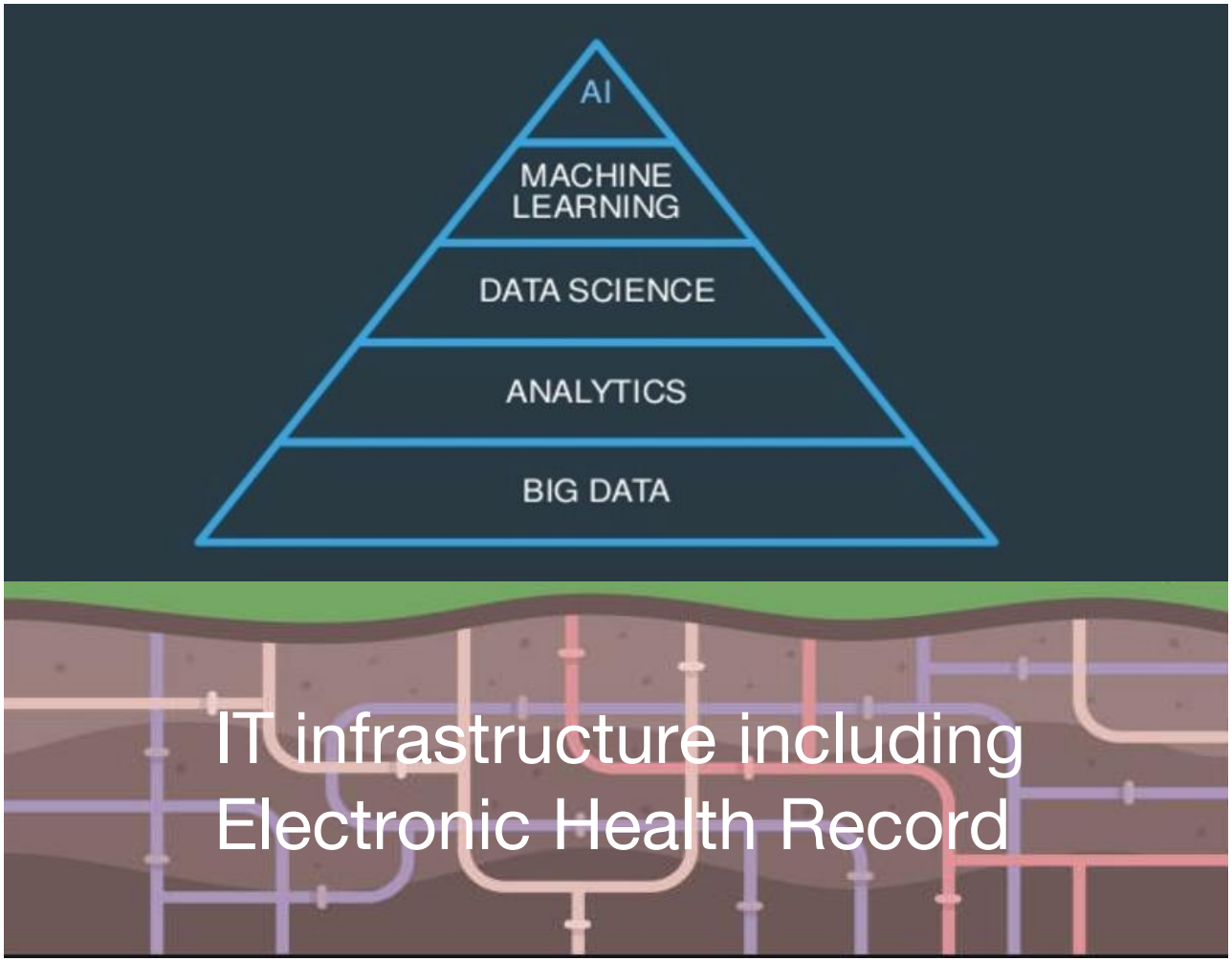
Big Data and Advanced Analytics in real-world NHS Hospitals

Dr James Teo
Clinical Director of Data & AI
Consultant Neurologist
Kings College Hospital
Guys & St Thomas Hospital

Hierarchy of Needs

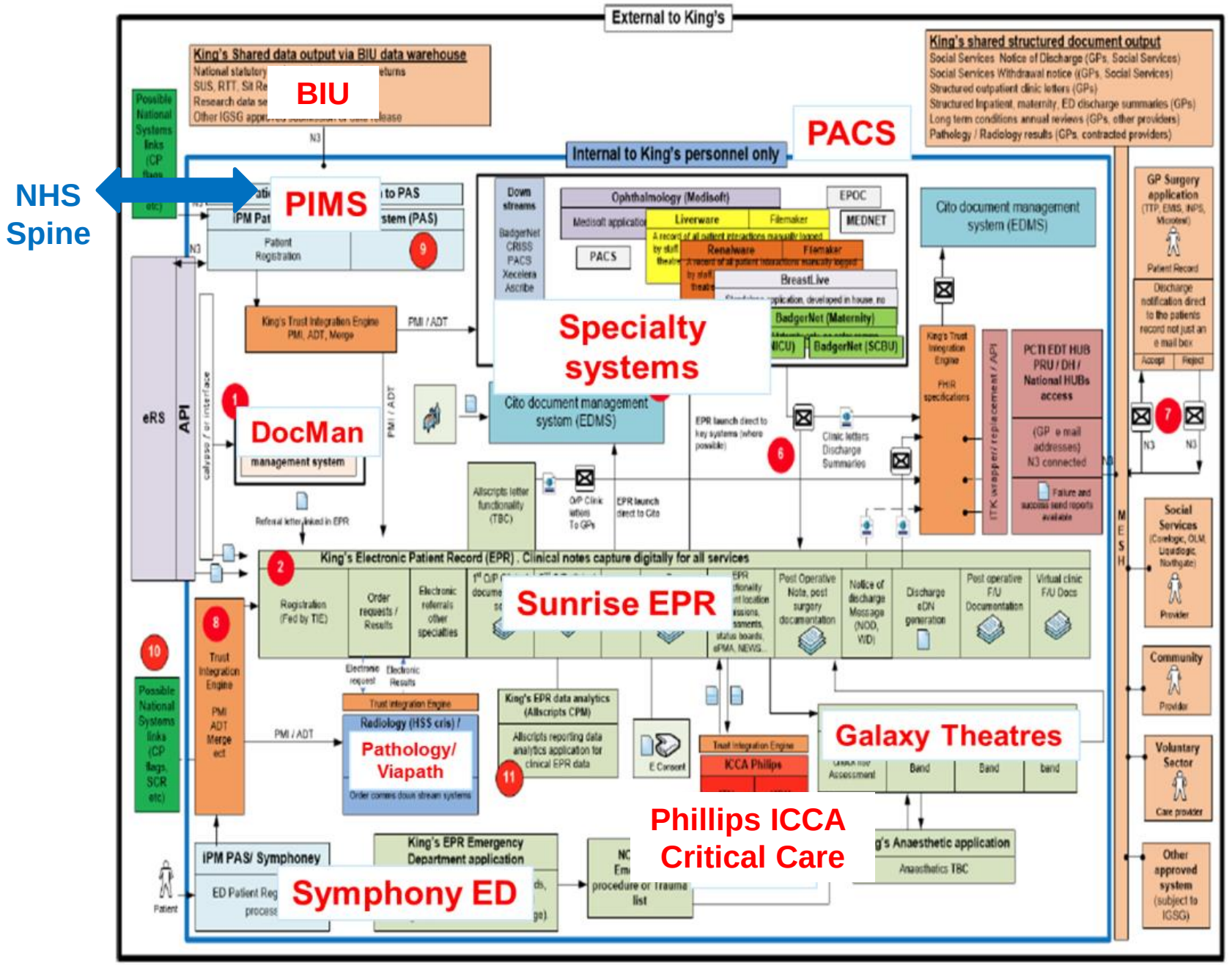



Health data

Problems

- Too much health data
- Health data is all over the place
- Health data is dirty
- Health data is unstructured
- Data is stored inflexibly

- Tools or analytics for understanding the volume of data is immature


NHS Spine
External to King's
King's Shared data output via BIU data warehouse
BIU
PACS
King's shared structured document output
Social Services Notice of Discharge (GPs, Social Services)
Social Services Withdrawal notice (GPs, Social Services)
Structured outpatient clinic letters (GPs)
Structured Inpatient, maternity, ED discharge summaries (GPs)
Long term conditions annual reviews (GPs, other providers)
Pathology / Radiology results (GPs, contracted providers)
Internal to King's personnel only
PIMS
Patient Registration
Specialty systems
Ophthalmology (Medisoft)
EPOC
MEDNET
Liverware
Filemaker
Renalware
BreastLive
BadgerNet (Maternity)
BadgerNet (SCBU)
PACS
Cito document management system (EDMS)
GP Surgery application
Patient Record
Discharge notification direct to the patients record not just an e mail box
Accept
Reject
King's Trust Integration Engine
PMI, ADT, Merge
PMI / ADT
eRS
API
DocMan
management system
Referral letter linked in EPR
Allscripts letter functionality (TBC)
O/P Clinic letters To GPs
EPR launch direct to Cito
EPR launch direct to key systems (where possible)
Clinic letters Discharge Summaries
King's Trust Integration Engine
FHIR specifications
ITK wrapper / replacement / API
PCTI EDT HUB PRU / DH / National HUBs access
(GP e mail addresses) N3 connected
Failure and success send reports available
N3
MESH
Social Services
Provider
King's Electronic Patient Record (EPR) . Clinical notes capture digitally for all services
Sunrise EPR
Registration (Fed by TIE)
Order requests / Results
Electronic referrals other specialties
Post Operative Note, post surgery documentation
Notice of discharge Message (NOD, WD)
Discharge eDN generation
Post operative F/U Documentation
Virtual clinic F/U Docs
Trust Integration Engine
PMI ADT Merge ect
Possible National Systems links (CP flags, SCR etc)
Electronic request
Electronic Results
Trust integration Engine
Radiology (HSS cris) /
Pathology/ Viapath
Order comms down stream systems
King's EPR data analytics (Allscripts CPM)
Allscripts reporting data analytics application for clinical EPR data
E Consent
Trust Integration Engine
ICCA Philips
Galaxy Theatres
Assessment
Band
Band
band
Phillips ICCA Critical Care
Community
Provider
Voluntary Sector
Care provider
Other approved system (subject to IGSG)
IPM PAS/ Symphony
ED Patient Reg
process
Patient
King's EPR Emergency Department application
Symphony ED
procedure or Trauma list
g's Anaesthetic application
Anaesthetics TBC

Unstructured Data

Easy for humans to input
Easy for humans to read
Contained in documents and variety of formats
Agnostic to ontologies and can capture non-health concepts
Particular to language

Problem: DIRTY & MESSY

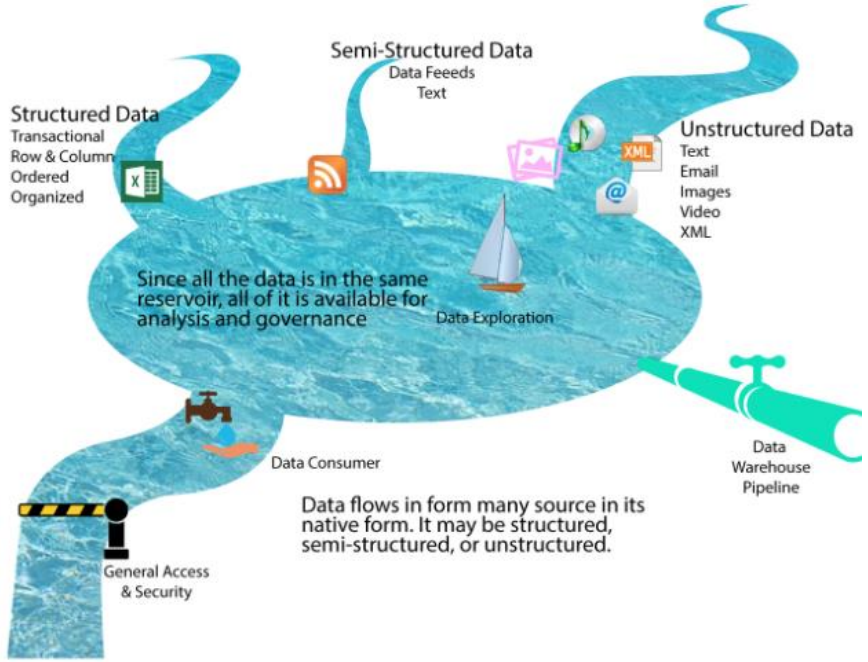
"Mrs Smith is a 65 year old woman with atrial fibrilation had a CVA in March. She had a past history of a #NOF and OA. She has a family history of breast cancer. She has been prescribed apixiban. She has no history of haemorrhage."

- Spelling / Typo
- Nomenclature
- Acronyms
- Negative terms
- Family history terms

Search

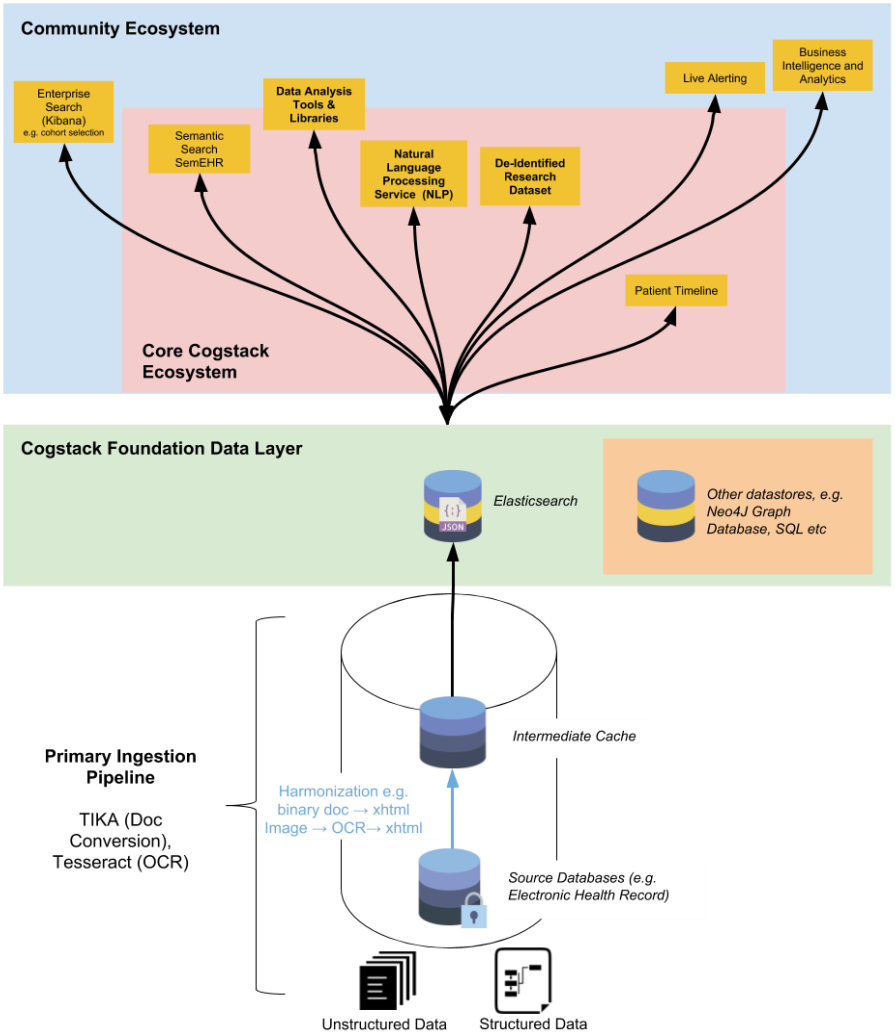
Database Silo's

Data Lake



But a lake is useless if you can't find what you are looking for

Cogstack Data Pipeline



Documentation (Wiki):
https://cogstack.atlassian.net/wiki/spaces/COGDOC/overview

GitHub:
https://github.com/CogStack/

CogStack

Watch an introduction to CogStack

HDRUK Health Data Research UK

King's College London

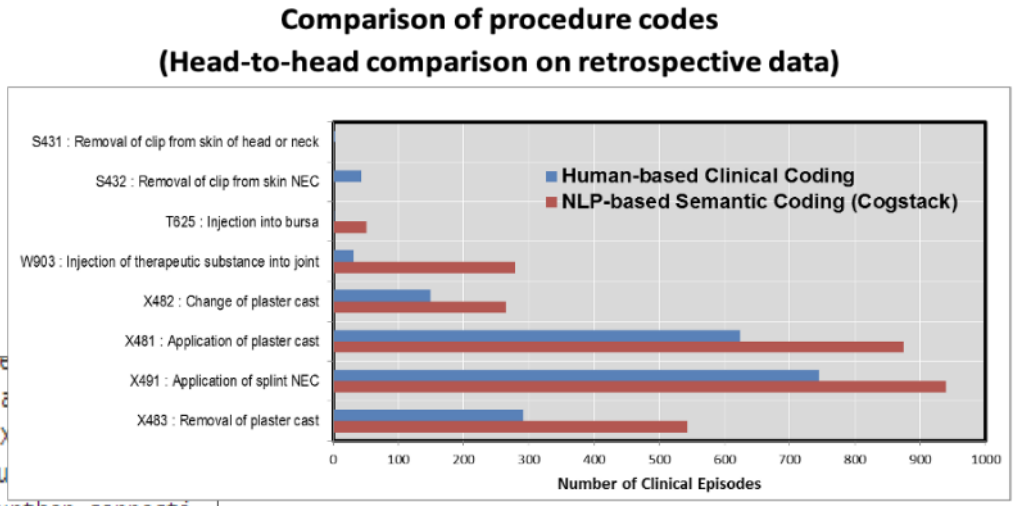
King's College Hospital NHS Foundation Trust

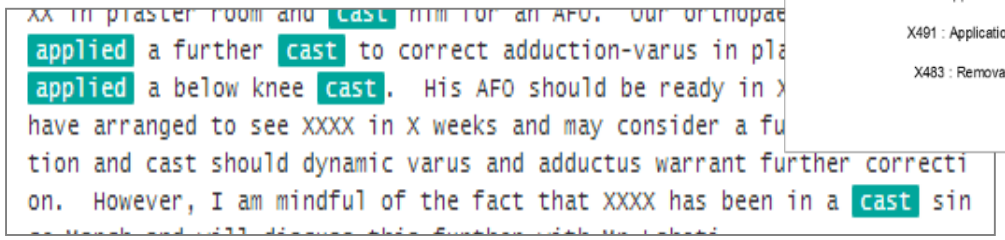
South London and Maudsley NHS Foundation Trust

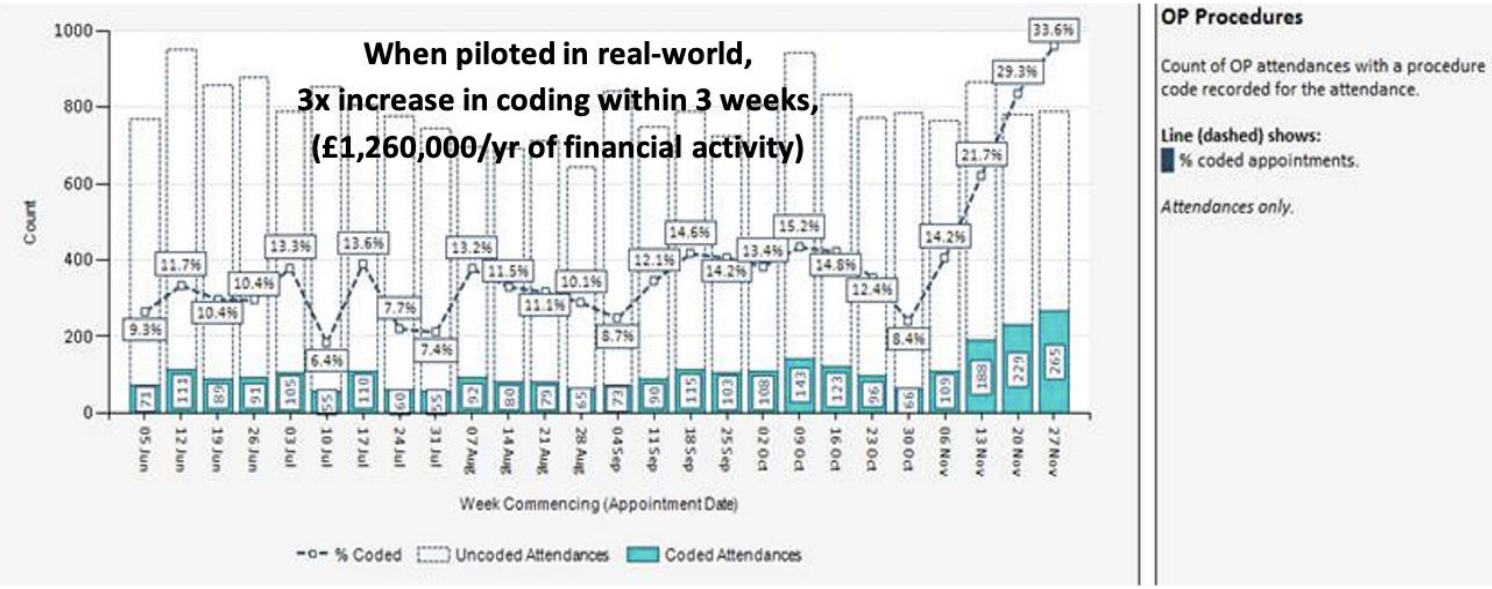
Keyword searching of entire clinical record

CLINICAL CODING
under-coding found in unstructured text

Mismatch of patients having OPD procedures documented in letters (RED) versus what is coded on appointment systems (BLUE)







Ophthalmology

What is the required data?

_Unique patient identifier eg hospital number_____

_____

For which patients?

1) Patients in whom MOG or Myelin oligodendrocyte glycoprotein is mentioned in notes
2) Patients with bilateral optic neuritis or bilateral inflammatory optic neuropathy
3) Patients who have (uveitis or iritis) and (MS or multiple sclerosis or demyelinating disease)

_____

Haemato-oncology

What is the required data?

__All patients who have had sCD25 requests from 2012 to 2019 irrespective of whether an actual sample was sent to Great Ormond Street for processing (and results should a sample have been sent)

For which patients?

_LITU or any liver wards_____

_____

For which period?

__2012 to current_____

_____

When is the data required? (Please allow at least 1 week from date of request)

__As soon as possible please_____

Gastroenterology

What is the required data?

Patients with inflammatory bowel disease (classified as inflammatory bowe
~~disease, IBD, ulcerative colitis or Crohn's disease) who have dysplasia found in~~
colonic biopsies taken at colonoscopy or sigmoidoscopy.

For which patients?

Patients with inflammatory bowel disease - Crohn's disease or ulcerative

asia can take years to develop, and is n
~~he period. This will help assess whethe~~
e given new therapies.

t least 1 week from date of request)
ailable

General Surgery

What is the required data?

The number of blood tranfusions that take place in patients that have appendectomies during their admission vs the number of group and screens they have ordered during their admission.

_____

For which patients?

Patients that have undergone appendectomies at King's College Hospital over the past 6 months

(NHS OPCS4.8 Codes H02.8 and H02.1)

(BUPA OPCS Codes and Category Grading H0280 and H0210)

For which period?

1/12/2018 to 1/6/2019

_____

Anticoagulation & Maternity

What is the required data?

List of patients seen in the thrombosis haem clinics, prescribed LMWH during ~~pregnancy and the postnatal period~~

For which patients?

Pregnant and postnatal patients prescribed LMWH seen in the clinics of Prof Arya, ~~Dr Raj Patel, Dr Lara Roberts, Dr Julia Czuprynska, Dr Preetan Dighe~~

For which period?

01-Jan-2009 to 31-Dec-2018

When is the data required? (Please allow at least 1 week from date of request)

Fri 27th September 2019

Critical Care/ ITU

What is the required data?

pull 3 separate lists of cases with diagnosis NORSE or super-refractory status or refractory status

For which patients?

All patients on our database for the last ten years

For which period?

Last ten years

When is the data required? (Please allow at least 1 week from date of request)

At the earliest

Gastroenterology

What is the required data?

Histology data , endoscopy data, cancer histology, imaging results : Specific data fields are: Histology reports with the following search terms listed: gastric intestinal metaplasia, atrophic gastritis, dysplasia, barretts oesophagus, columnar lined oesophagus with goblet cells.

For which patients?

Gastric intestinal metaplasia, Atropic gastritis, barretts oesophagus, neuroendocrine tumours

For which period?

2015-2019

When is the data required? (Please allow at least 1 week from date of request)

by end of November 2019

Anticoagulation

What is the required data?

We want to look at which anticoagulants are currently being used for management of LV thrombus (warfarin vs DOACs). A trial run of this was performed when I met with James Teo and we looked at the terms 'LV thrombus', ventricular thrombus,

For which patients?

Patients newly diagnosed with an LV thrombus at King's Denmark Hill Site.

For which period?

Calendar years 2016-2018 inclusive (i.e. 3 full years)

When is the data required? (Please allow at least 1 week from date of request)

As soon as possible - we have a clinical fellow for a limited period who will be able to look at the data for us.

Bariatric surgery / Diabetes

What is the required data?

The same data as requested by David Hopkins and Prof Rubino in November 2018.

The number (and percentage) of patients with diabetes who have undergone bariatric surgery in the last five years versus those who do not have diabetes

For which patients?

All patients who have had bariatric surgery (or are currently on the bariatric pathway)

For which period?

Last five years (2014/15 - to date)

When is the data required? (Please allow at least 1 week from date of request)

Non urgent

From: "JONES, Isla (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST)" <isla.jones@nhs.net>
Date: Thursday, 11 July 2019 at 19:32
To: "YORKE, Richard (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST)" <richard.yorke@nhs.net>
Cc: Oenone Williams <o.poole-wilson@nhs.net>
Subject: Re: BIMs in clinical notes on HAU June 2019

Hi Richard,

Thanks so much for this - this is a good [illegible] report

We were wanting to look at all cases in which BIMs / Best Interest Meetings were mentioned within the body of EPR clinical notes as before rather than documents. Would it be possible to pull this quickly on Friday?

Just need it for patients on Donne, Byron and Marjory Warren.

Thanks again Richard - I know you are very busy

Elderly Care/ Palliative Care

MUSTAFA, Omar (KING'S COLLEGE HOSPITAL NHS FOUNDATI Mark as unread
Tue 05/11/2019 08:50
Inbox

To: TEO, James (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST);
Cc: MU[illegible]TRUST);
WHITELAW, Benjamin (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST);
PRAGUE, Julia (IMPERIAL COLLEGE HEALTHCARE NHS TRUST);
Diab[illegible]NG'S COLLEGE HOSPITAL NHS FOUNDATION TRUST);
You replied on 05/11/2019 09:43.

Endocrinology Alerts

Dear James,

I wonder if you could check a couple of things about the Diabetes insipidus safety alert.

We are now doing a prospective phase to identify patients on daily basis. Will keep you posted on the results. There is some work nationally but so far not as advanced using Cogstack.

We noticed there are no alerts since Friday 1/11/2019. I am wondering whether there is no patients that fit the criteria or there is an issue with the code. We have spotted some patients

And does the query screen for the drug chart/EPMA and the new clinical notes system (via documents).

Thank you

Omar

From: NOTIFICATIONS, Kch (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST) <kch.notifications@nhs.net>
Sent: 01 November 2019 12:00
To: Diabetesinsipidus-Safety (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST) <kch-tr.diabetesinsipidus-safety@nhs.net>; TEO, James (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST) <jamesteo@nhs.net>
Subject: Diabetes Inspidus Alert

The following patients have the word **'insipidus'**, 'DDAVP' or 'desmopressin' mentioned during Inpatient or Emergency episodes in past 24hours:

From: YORKE, Richard (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST)
Sent: 05 October 2018 16:48
To: TEO, James (KING'S COLLEGE HOSPITAL NHS FOUNDATION TRUST)
Subject: CogStack

Hey James

I've now got access to Kibana

Neurosurgery/ Trauma

Any chance you can send me some queries to play with. Perhaps the one that you did for me that pulled Cervical #/Ondontoid peg #

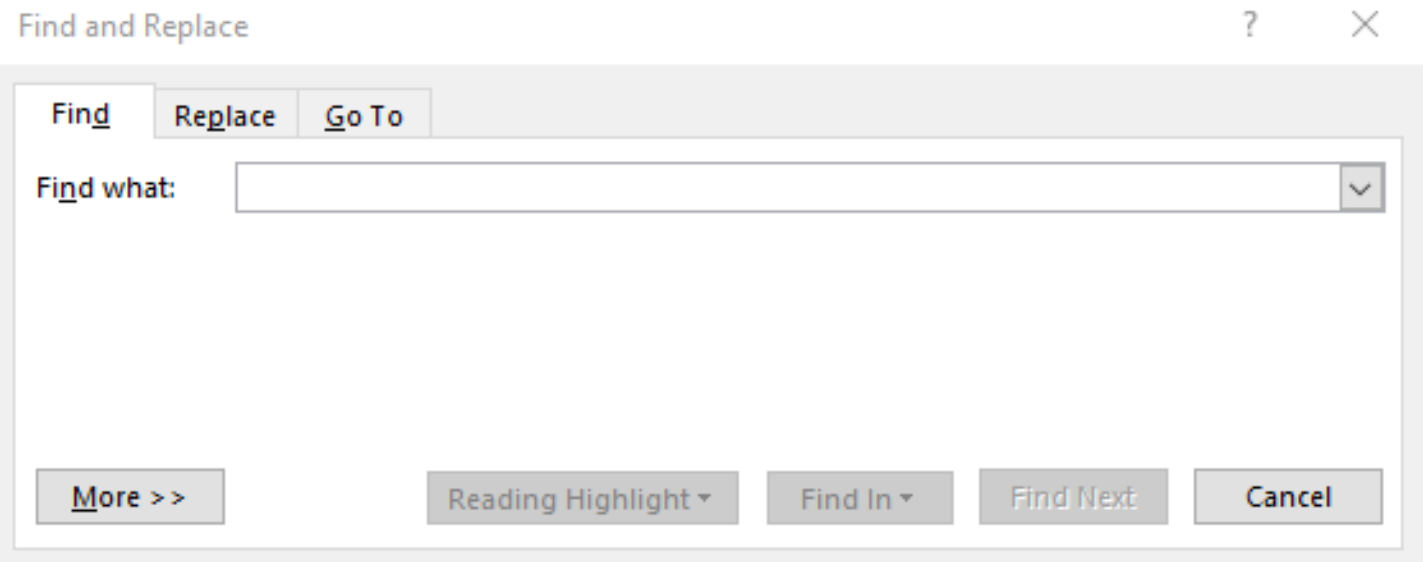
Also – Admission Clerking document is in live for people to review – feel free to feedback. There's nothing specific for stroke in there at the moment, it's geared to acute med

Cheers

Richard

Richard Yorke
Head of EPR
King's College Hospital NHS Foundation Trust
LONDON UK
Richard.Yorke@nhs.net

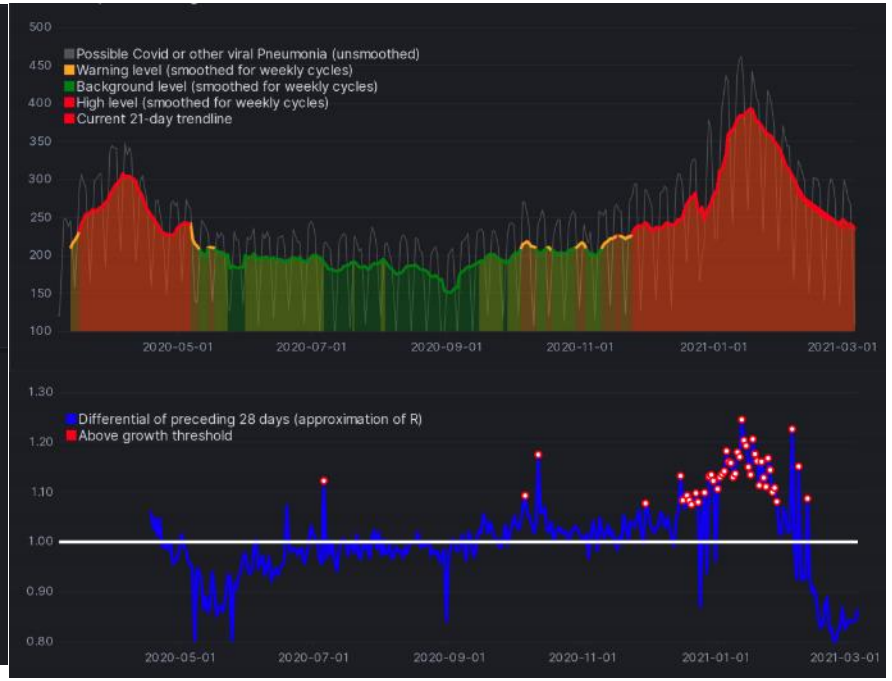
“All you did is press Ctrl-F”



Freetext signal from clinical notes

KCH (900 beds)
PRUH (600 beds)

Guys & St Thomas (850 beds)



npj | Digital Medicine

www.nature.com/npjdigitalmed

BRIEF COMMUNICATION OPEN

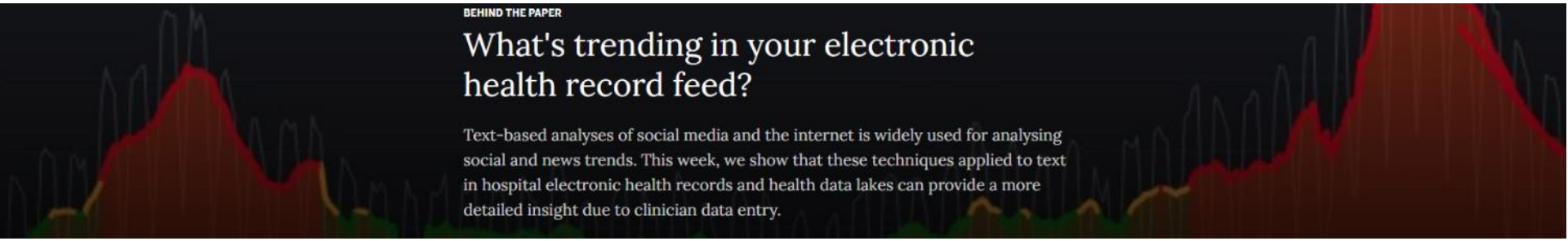
Check for updates

Real-time clinician text feeds from electronic health records

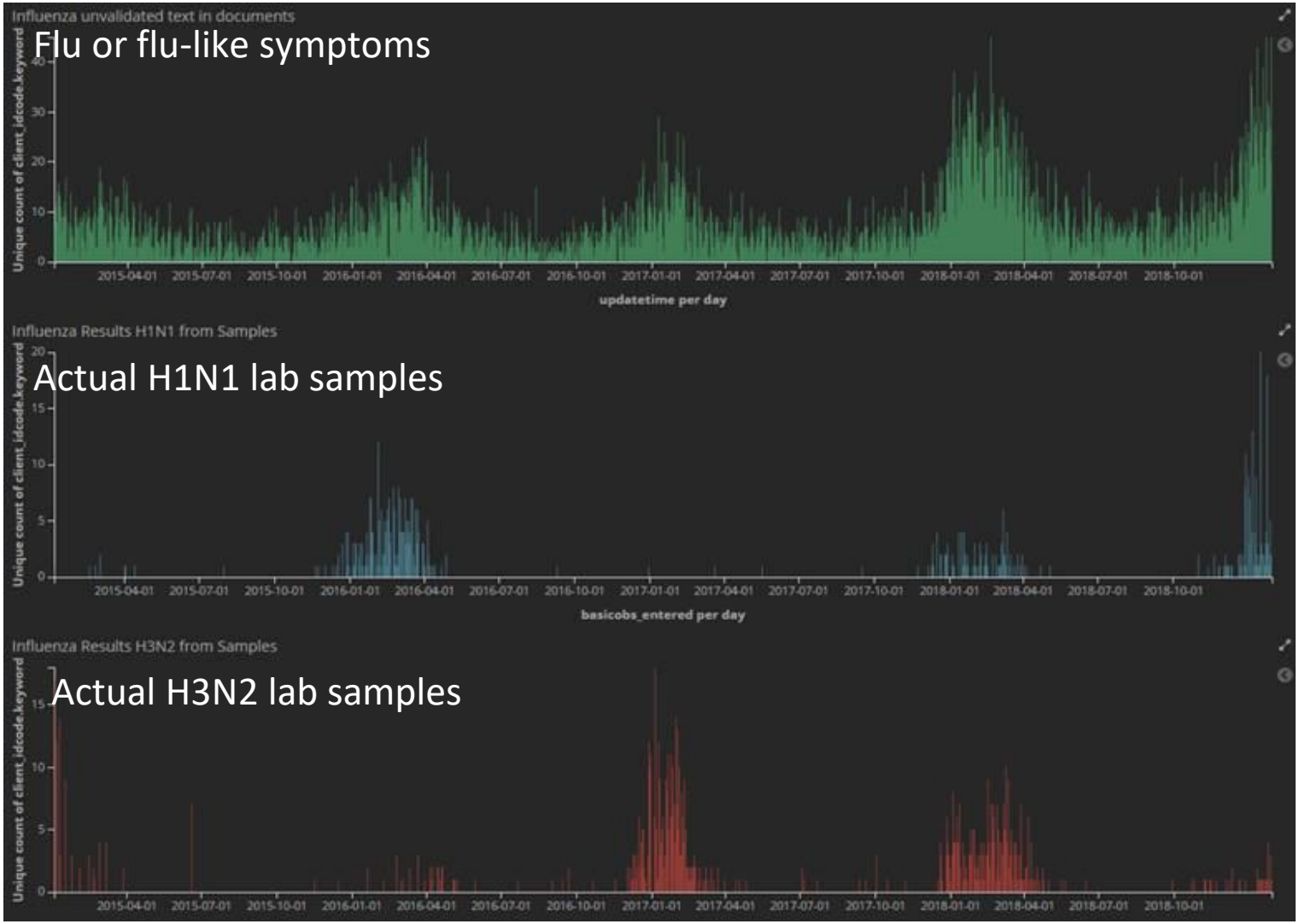
James T. H. Teo[1,2,3✉], Vlad Dinu[3], William Bernal[1], Phil Davidson[1], Vitaliy Oliynyk[2], Cormac Breen[2], Richard D. Barker[1] and Richard J. B. Dobson[3]

Analyses of search engine and social media feeds have been attempted for infectious disease outbreaks, but have been found to be susceptible to artefactual distortions from health scares or keyword spamming in social media or the public internet. We describe an approach using real-time aggregation of keywords and phrases of freetext from real-time clinician-generated documentation in electronic health records to produce a customisable real-time viral pneumonia signal providing up to 4 days warning for secondary care capacity planning. This low-cost approach is open-source, is locally customisable, is not dependent on any specific electronic health record system and can provide an ensemble of signals if deployed at multiple organisational scales.

npj Digital Medicine (2021)4:35 ; https://doi.org/10.1038/s41746-021-00406-7



Real-time data feed of "flu or flu-like illnesses"



Unstructured Data

Easy for humans to input
Easy for humans to read
Contained in documents and variety of formats
Agnostic to ontologies and can capture non-health concepts
Particular to language

Problem: DIRTY & MESSY

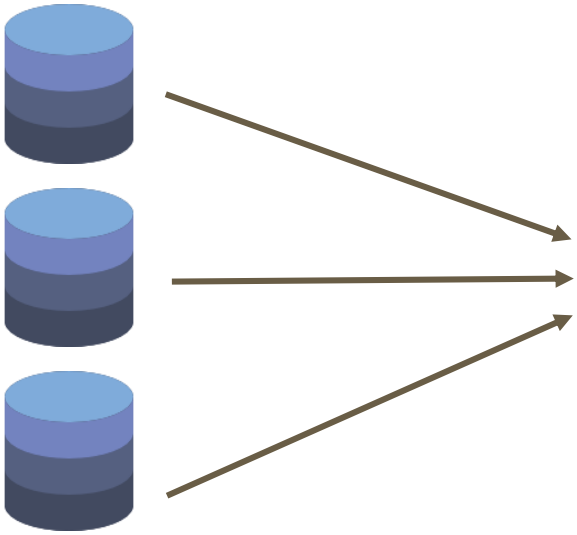
"Mrs Smith is a 65 year old woman with atrial fibrilation had a CVA in March. She had a past history of a #NOF and OA. She has a family history of breast cancer. She has been prescribed apixiban. She has no history of haemorrhage."

- Spelling / Typo
- Nomenclature
- Acronyms
- Negative terms
- Family history terms


King's
NLP
Natural
Language
Processing
Search and Semantic

An evolution waiting to happen

Clinical Record source systems



Unstructured, heterogeneous data

CogStack

Inferred structure e.g. SNOMED, ICD10

MedCAT

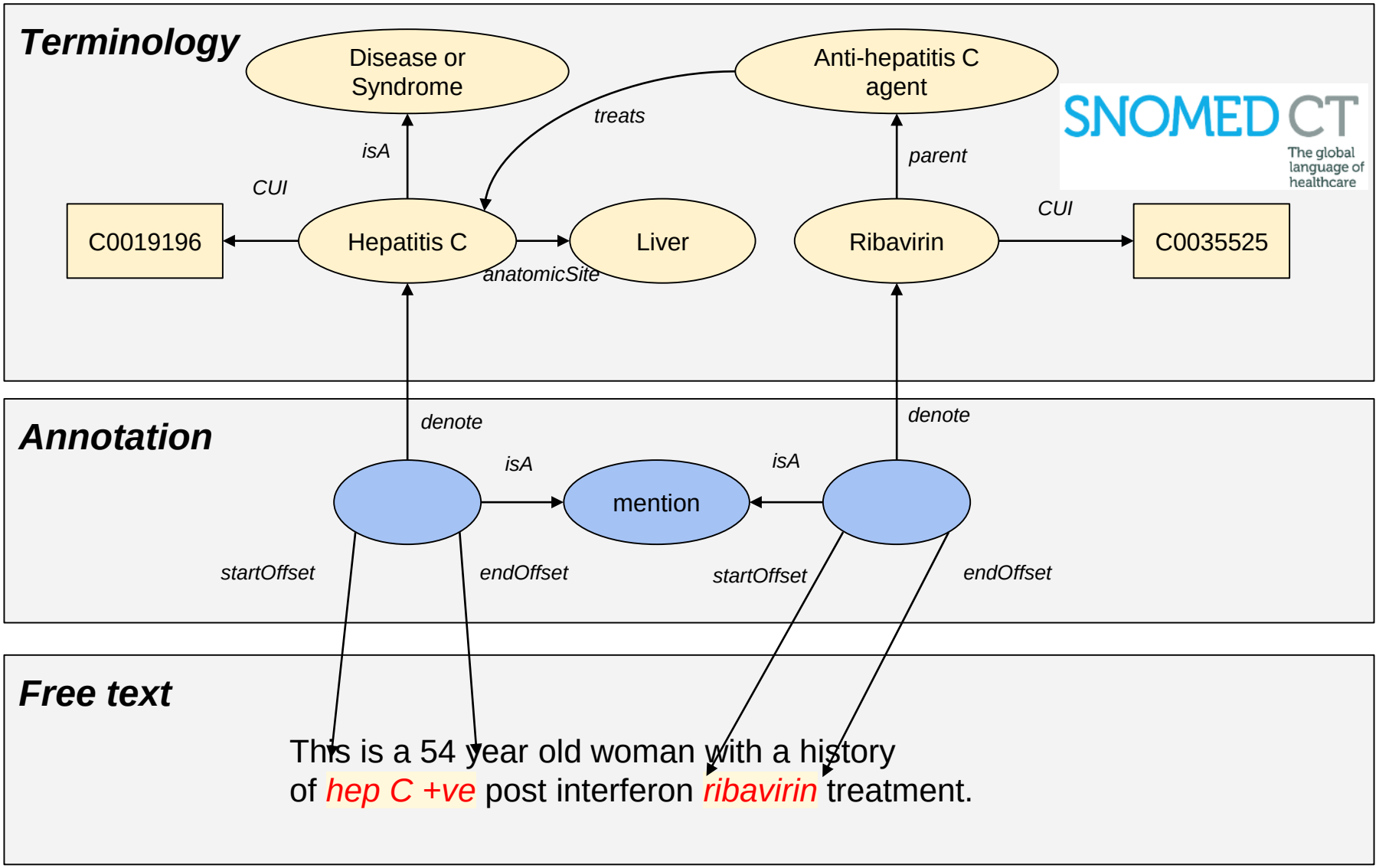
AI-based language processing

Reference Number Lookup and hand-gathered lists

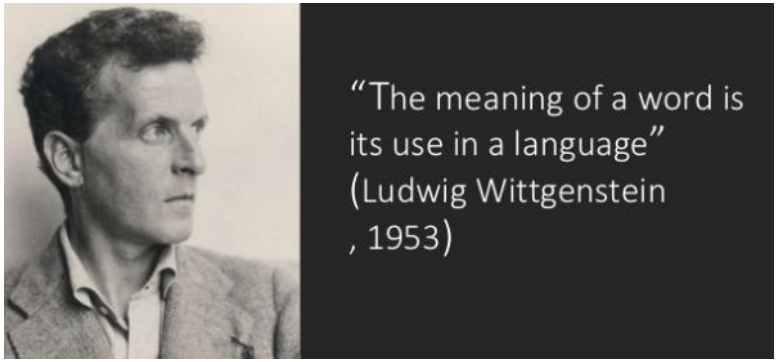
Data pooling Keyword Search Engine

Searching on meaning

Tag words with a standardised label
Standardised labels need to sit in a Knowledge Graph



Teaching AI to read medical text using syntatic context



During C0347984 the night C0240526 HR C0018810 was in the 40s-50s. The patient C0030705 was at 8mg/ HR C0439227

Correctly detected HR and tagged it with **CUI: C0018810 - Heart Rate**

Correctly detected HR and tagged it with **CUI: C0439227 - Hour**

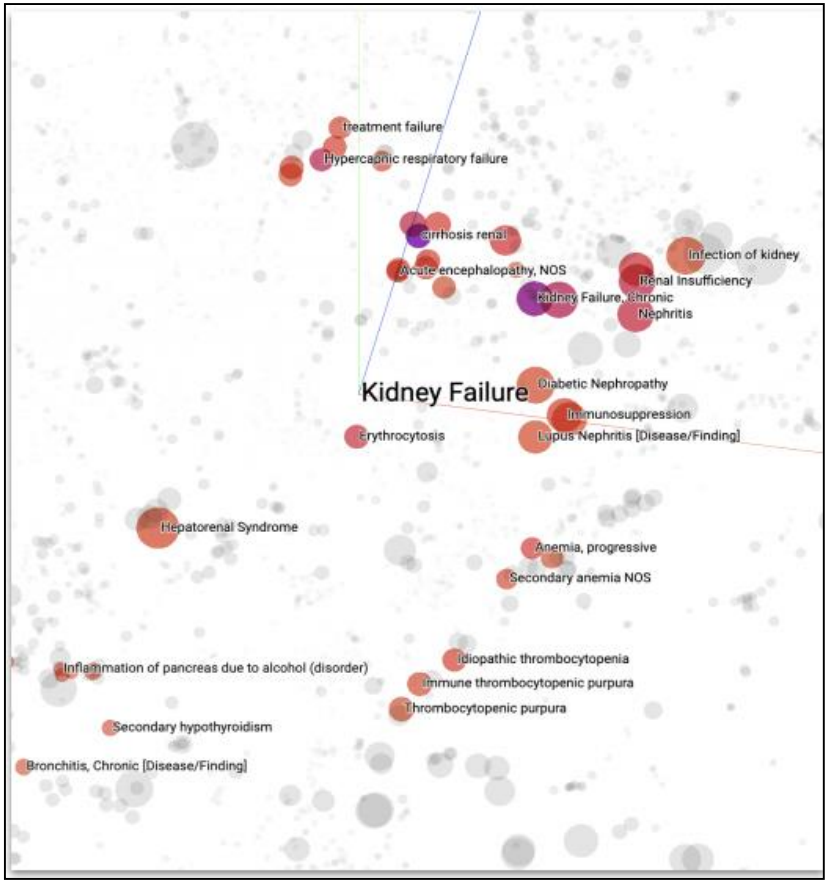
Machine Learning of Clinical Language

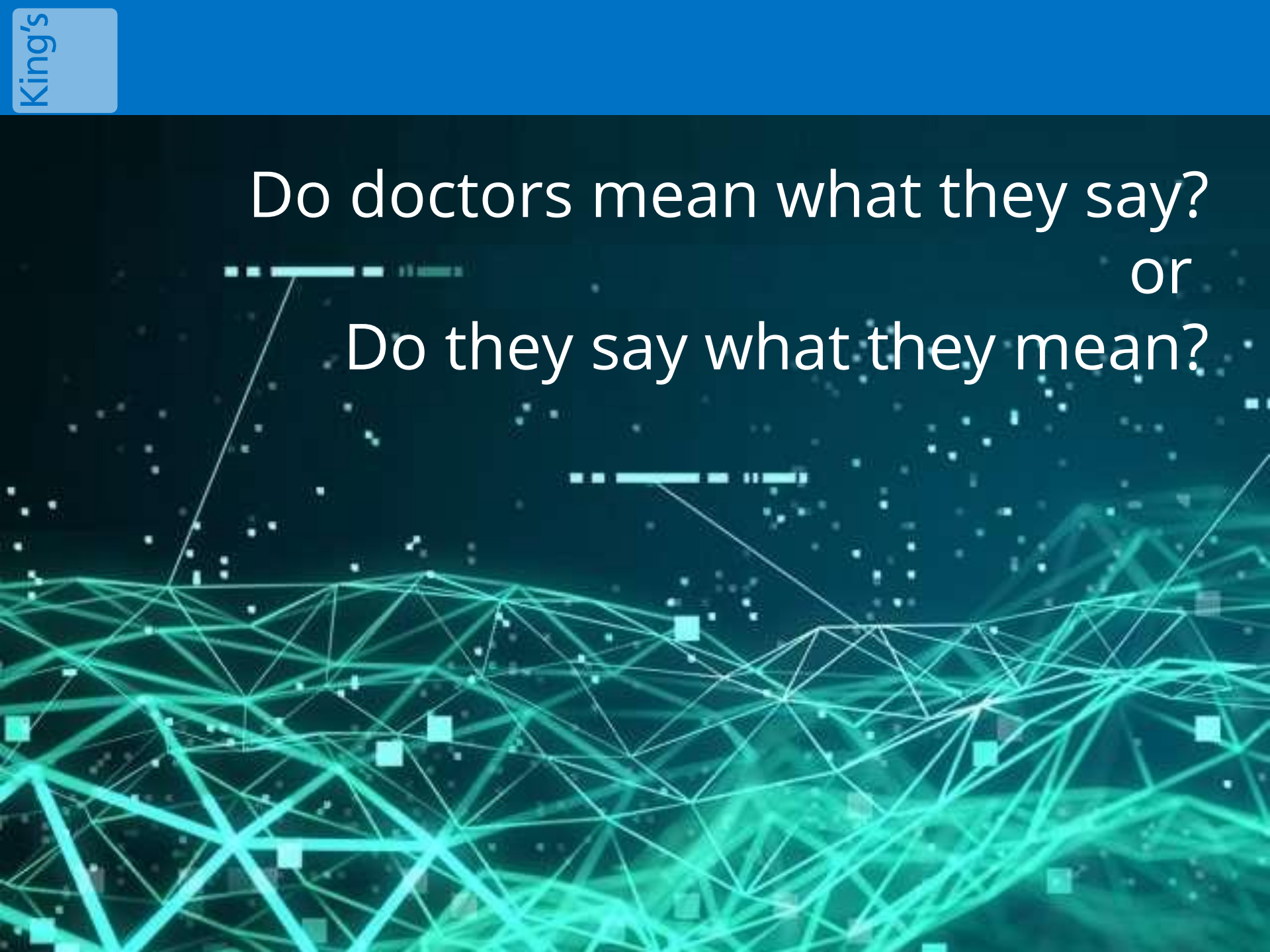
Deep Learning Neural Network learning to read medical text

- Shown 17 million KCH medical documents
- Looks at the surrounding words for context
- Learns synonyms, acronyms and local jargon
- Maps them to international medical dictionaries
- Improves when taught by humans

Kidney Failure	Paracetamol	Duloxetine
Chronic kidney disease	Acetaminophen	Paroxetine
Kidney Failure, Acute	Codeine	Buspirone
Kidney Failure, Chronic	Aspirin	Olanzapine
Renal Insufficiency	Ibuprofen	Aspirin
Acute kidney injury	Naproxen sodium	Fluoxetine
Chronic Kidney	Hydrocodone	

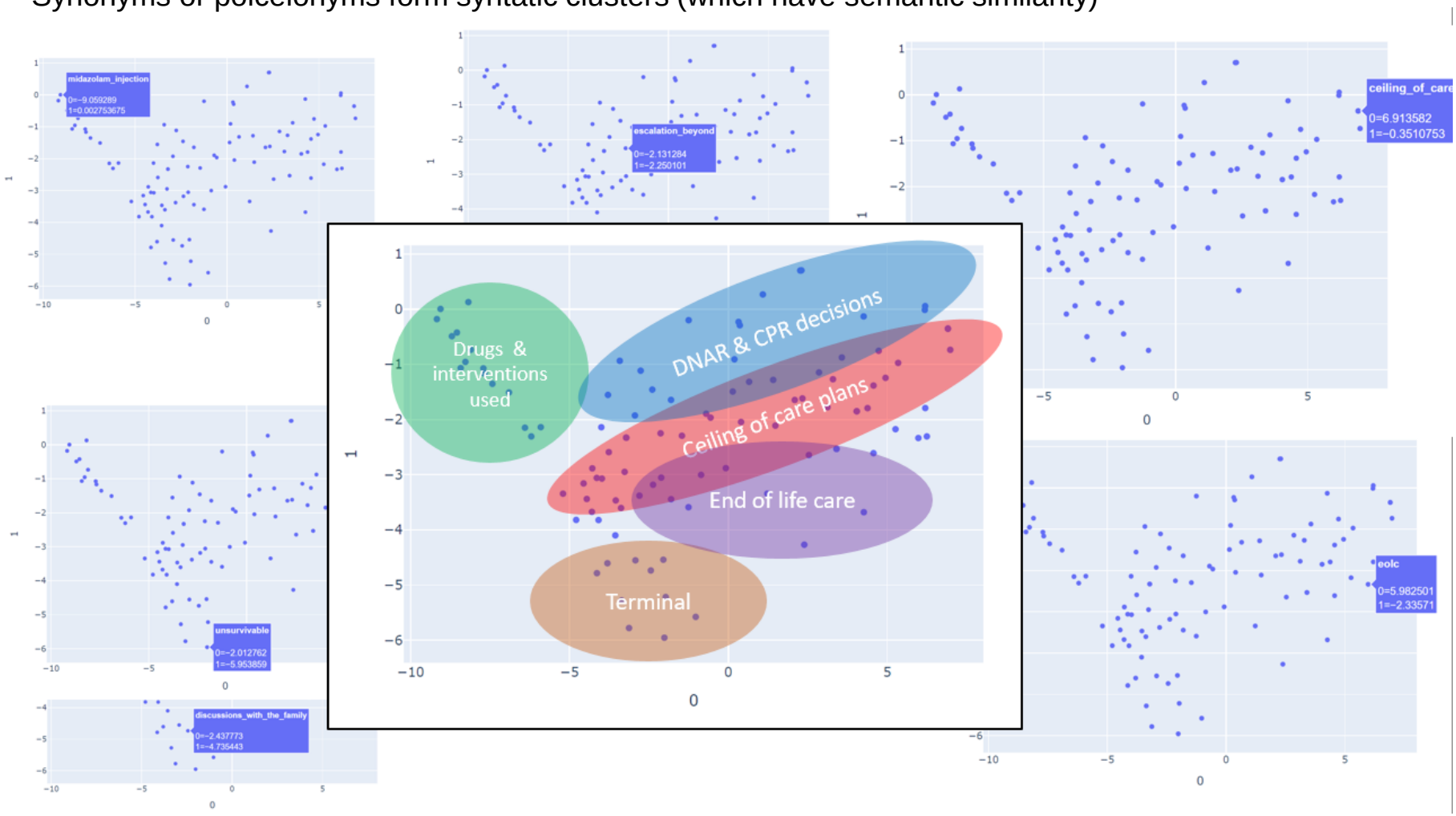
Most similar concepts based on Vector embeddings




Do doctors mean what they say?
or
Do they say what they mean?

Hunting for synonyms
t-distributed stochastic neighbor embedding (TSNE) plots

X-gram (up to 4) word clusters closest to phrase "Ceiling of Care" or "Ceiling of Treatment"
Synonyms or poicelonyms form syntatic clusters (which have semantic similarity)



Correlation with outcome

	Key phrases showing up in documents from Oct-2019 to Sept-2019	Any inpatients with the phrase in Sunrise EPR	Any inpatients with the phrase and Death Dates within 7 days	%	Relative risk vs annual control
Ceiling of Care Cluster	"not for inotropes"	20	3	15.0%	6.75
	"not for hdu"	39	7	17.9%	8.07
	"currently for full"	83	17	20.5%	9.21
	"ceiling of rx" OR "ceiling of care" OR "ceilings of care" OR "ceilings of treatment" OR "ceiling of treatment"	1,254	203	16.2%	7.28
	"ceiling of care" OR "ceilings of care" OR "limit of care" OR "limits of care"	910	169	18.6%	8.35
	"ceilings of treatment" OR "ceilings of treatment" OR "ceiling of rx"	431	54	12.5%	5.64
	"Treatment Escalation Plan"~2	3,181	55	1.7%	0.78
	"not for intubation" OR "not suitable for intubation" OR "not appropriate for intubation"	184	51	27.7%	12.47
	"not for itu" OR "not for icu" OR "not suitable for itu" OR "not appropirate for itu" OR "not for escalation to itu" OR "not for critical care"	284	99	34.9%	15.68
	"ward based ceiling of care" OR "'ward based care only"	140	53	37.9%	17.03
	"not for escalation" OR "escalation beyond"	193	75	38.9%	17.48
	"unsurvivable"	59	34	57.6%	25.92
End of Life Care Cluster	"palliative treatments only" OR "palliative input" OR "palliative medications" OR "palliation"	1165	390	33.5%	15.06
	"withdrawal of care" OR "withdrawal of treatment" OR "withdrawal of intensive"	67	38	56.7%	25.52
	"terminal care" OR "end of life care" OR "eol care" OR eolc	2138	1230	57.5%	25.88
	"liverpool care pathway" OR "liverpool care pathway verified" OR "secretions liverpool care pathway"	0	0	0.0%	0.00
Control	None of the above phrases in either cluster	424905	3406	0.8%	
	None of the above phrases in either cluster AND deceased date in 12 months		9445	2.2%	

Optimism prevails

Content-less headers

Similar rates of mortality

Combining Structured and Unstructured Data for rapid analysis

Building a multi-feature risk-predictor with National Early Warning Score (NEWS2) for Covid

Model trained on >40 variables

- Demographics, vital signs, comorbidities, common blood tests
- Many correlated features, but top 5 features captures most of the variance

		NEWS2 + age Mean (95%% CI)	All features Mean (95%% CI)
14-day ICU/death	AUC	0.700 [0.680, 0.722]	0.735 [0.715, 0.757]
	Brier score	0.192 [0.186, 0.197]	0.183 [0.177, 0.189]
	Sensitivity[1]	0.778 [0.747, 0.815]	0.735 [0.702, 0.772]
	Specificity[1]	0.478 [0.445, 0.509]	0.592 [0.562, 0.621]

8 hospital sites, 3 countries
KCH
GSTT
UCLH
University Hospitals Birmingham
University College London Hospitals
University Hospital Bristol Westmead
Oslo University Hospital, Norway
Wuhan Sixth Hospital & Taikang Tongji Hospital, Wuhan hospital, China validation

Multi-site international validation on real world data from 1st Wave (n=6k) Carr et al., 2021

Evaluation and improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study

Ewan Carr[1*†], Rebecca Bendayan[1,2†], Daniel Bean[1,3], Matt Stammers[4,5,6], Wenjuan Wang[7], Huayu Zhang[8], Thomas Searle[1,2], Zeljko Kraljevic[1], Anthony Shek[9], Hang T. T. Phan[4,5], Walter Muruet[7], Rishi K. Gupta[10], Anthony J. Shinton[6], Mike Wyatt[11], Ting Shi[8], Xin Zhang[12], Andrew Pickles[1,2], Daniel Stahl[1], Rosita Zakeri[13,14], Mahdad Noursadeghi[15], Kevin O'Gallagher[13,14], Matt Rogers[11], Amos Folarin[1,3,16,17], Andreas Karwath[18,19,20], Kristin E. Wickstrøm[21], Alvaro Köhn-Luque[22], Luke Slater[18,19,20], Victor Roth Cardoso[18,19,20], Christopher Bourdeaux[11], Aleksander Rygh Holten[23], Simon Ball[20,24], Chris McWilliams[25], Lukasz Roguski[3,16,19], Florina Borca[4,5,6], James Batchelor[4], Erik Koldberg Amundsen[21], Xiaodong Wu[26,27], Georgios V. Gkoutos[18,19,20,24], Jiaxing Sun[26], Ashwin Pinto[6], Bruce Guthrie[8], Cormac Breen[7], Abdel Douiri[7], Honghan Wu[3,16], Vasa Curcin[7], James T. Teo[9,13†], Ajay M. Shah[13,14†] and Richard J. B. Dobson[1,2,3,16,17†]

ARTICLE INFORMATION

doi https://doi.org/10.1101/2020.04.24.20078006

History September 30, 2020.

ARTICLE VERSIONS

Version 1 (April 29, 2020 - 04:54).

Version 2 (May 3, 2020 - 09:48).

Version 3 (June 11, 2020 - 14:42).

You are viewing Version 4, the most recent version of this article.

Data studies at speed

Comments (1)

Impact of ethnicity on outcome of severe COVID-19 infection. Data from an ethnically diverse UK tertiary centre

James TH Teo, Daniel M Bean, Rebecca Bendayan, Richard JB Dobson, Ajay M Shah

doi: https://doi.org/10.1101/2020.05.02.20078642

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

Abstract | Full Text | **Info/History** | Metrics | Preview PDF

ARTICLE INFORMATION

doi https://doi.org/10.1101/2020.05.02.20078642

History May 25, 2020.

ARTICLE VERSIONS

Version 1 (May 6, 2020 - 02:12).

Version 2 (May 13, 2020 - 21:06).

Version 3 (May 21, 2020 - 07:23).

You are viewing Version 4, the most recent version of this article.

Comments (4)

Treatment with ACE-inhibitors is associated with less severe disease with SARS-Covid-19 infection in a multi-site UK acute Hospital Trust

Daniel M Bean, Zeljko Kraljevic, Thomas Searle, Rebecca Bendayan, Andrew Pickles, Amos Folarin, Lukasz Roguski, Kawsar Noor, Anthony Shek, Kevin O'Gallagher, Rosita Zakeri, Ajay M Shah, James TH Teo, Richard JB Dobson

doi: https://doi.org/10.1101/2020.04.07.20056788

Now published in *European Journal of Heart Failure* doi: 10.1002/ejhf.1924

Abstract | Full Text | **Info/History** | Metrics | Preview PDF

ARTICLE INFORMATION

doi https://doi.org/10.1101/2020.04.07.20056788

History April 11, 2020.

ARTICLE VERSIONS

You are currently viewing Version 1 of this article (April 11, 2020 - 05:30).

The line of peer review

RESEARCH PAPER | VOLUME 28, 100574, NOVEMBER 01, 2020

A case-control and cohort study to determine the relationship between ethnic background and severe COVID-19

Rosita Zakeri • Rebecca Bendayan • Mark Ashworth • Daniel M. Bean • Hiten Dodhia • Stevo Durbaba • et al.

Show all authors

Open Access • Published: October 09, 2020 • DOI: https://doi.org/10.1016/j.eclinm.2020.100574

European Journal of **Heart Failure**

Research Article | Open Access

Angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers are not associated with severe COVID-19 infection in a multi-site UK acute hospital trust

Daniel M. Bean, Zeljko Kraljevic, Thomas Searle, Rebecca Bendayan, O'Gallagher Kevin, Andrew Pickles, Amos Folarin, Lukasz Roguski, Kawsar Noor, Anthony Shek, Rosita Zakeri ... See all authors

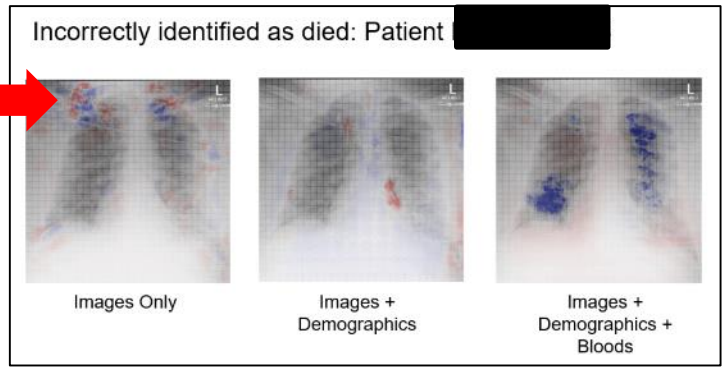
First published: 02 June 2020 | https://doi.org/10.1002/ejhf.1924 | Citations: 45


King's
Combining with other kinds of data
- Images

Machine Vision Saliency Maps (what is the AI looking at?)

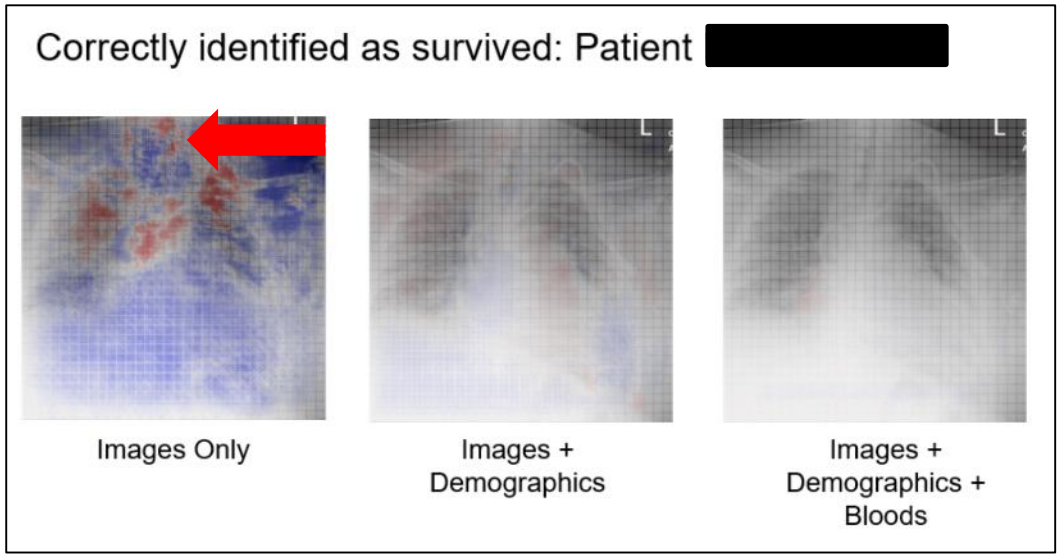
Blue = correlated good outcome
Red = correlated bad outcome

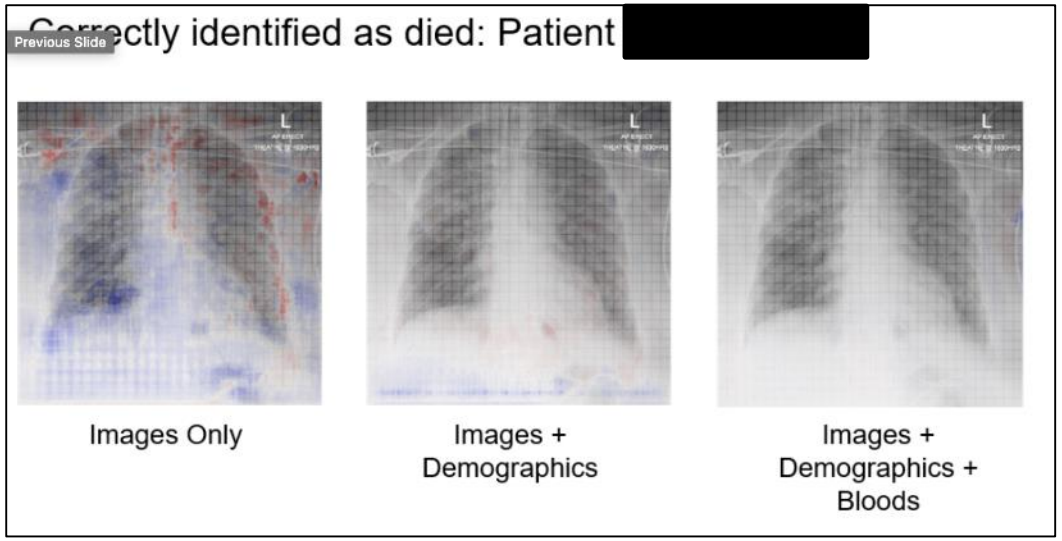
BELOW: Computer vision AI trained on images only easily captures signs of medical intervention and uses that as their marker of severity. AI cheats.



RIGHT: Adding in demographics, bloods and unstructured data stops the cheating but ends up showing limited value for the CXR signals

CONCLUSION: Most of relevant signal is already captured in bloods, vital signs and text, rather than Chest X-rays






Questions?