Advances in HEOR

New Frontiers Based on Developments in Artificial Intelligence

Moderator: Eric Wu
Managing Principal, Analysis Group, Inc.

November 14, 2018

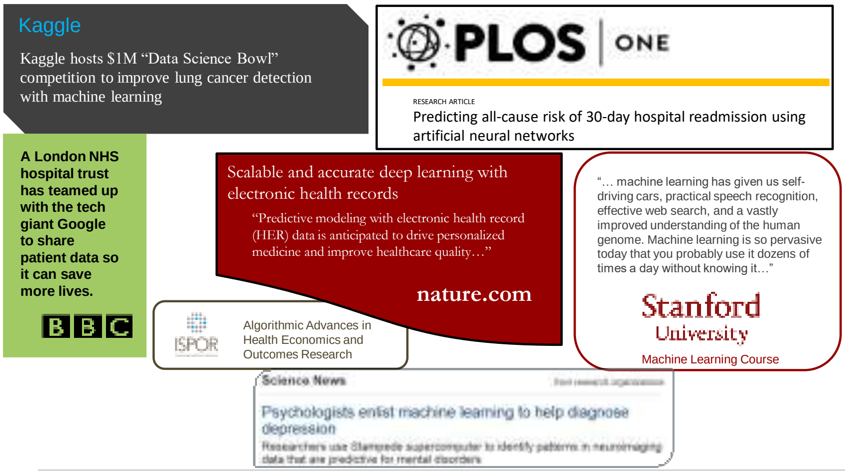
What Is Artificial Intelligence?

Artificial Intelligence borrows from many different fields

- Artificial Intelligence (AI) is an umbrella term usually referring to new methodological advances in the fields of Machine Learning (ML) and Natural Language Processing (NLP)
- ML focuses on pattern recognition and computational learning, and is used to either create predictive algorithms or to make classifications based on data
- Unlike traditional statistical methods, ML methods are capable of analyzing data and exploring unknown patterns without prior knowledge of possible relationships
- NLP combines statistics, computer science, and data management to process and analyze large amounts of language data

ANALYSIS GROUP

Wide Range of Real-Life Applications



ANALYSIS GROUP

Wide Range of Real-Life Applications

- Using unstructured Data – Robert Stewart MD, King's College London
  - Using NLP in combination with electronic health records for disease identification and diagnosis
- Using structured and unstructured data – Eric Q. Wu PhD, Analysis Group, Inc.
  - Using ML and NLP to improve performance and efficiency in literature reviews
- Using structured Data – Jimmy Royer PhD, Analysis Group, Inc.
  - Using AI algorithms to estimate treatment effect in retrospective studies

Developing the electronic mental health record for AI applications: the CRIS experience

Rob Stewart

Lead, Clinical and Population Informatics
SLAM Biomedical Research Centre for Mental Health

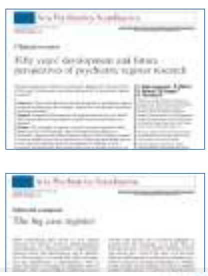
Professor of Psychiatric Epidemiology and Clinical Informatics
King's College London

Consultant in Liaison Old Age Psychiatry, South London & Maudsley NHS Foundation Trust

The trade-off …

Research data

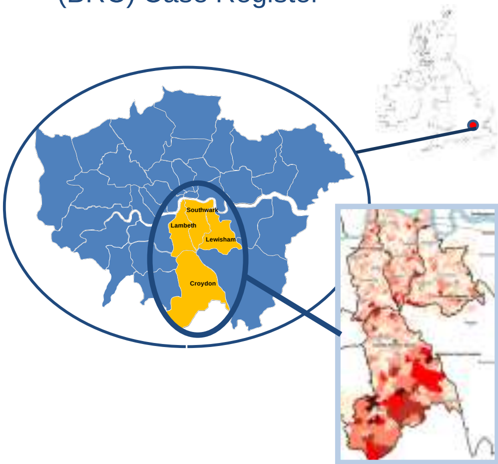
Administrative data



South London and Maudsley Biomedical Research Centre (BRC) Case Register

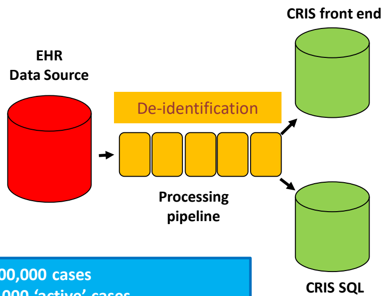
King's College London (KCL)

South London and Maudsley (SLAM)



NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

CRIS at the Maudsley – core functionality



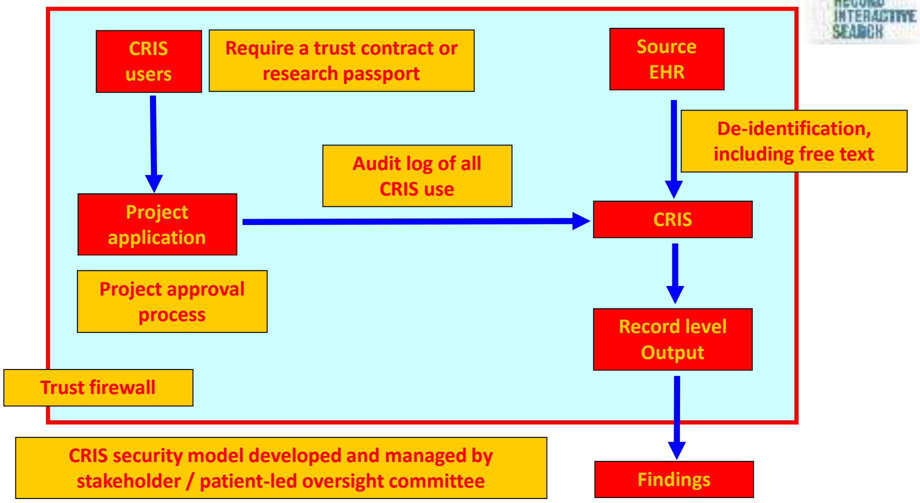
>400,000 cases
35,000 'active' cases
125 tables
6500 fields
30m documents

Set up in 2007-08 (NIHR funding)
Re-build and enhancement in 2017
Exported successfully to other UK Trusts
>120 research papers to date

brc.slam.nhs.uk

NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

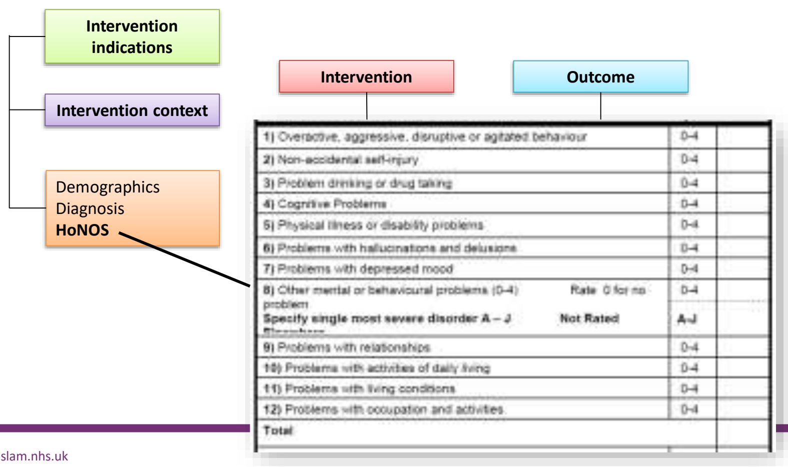
CRIS Security Model – service user led governance



Research ethics approval 2008, 2013, 2018
Numerous amendments for data linkages

NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

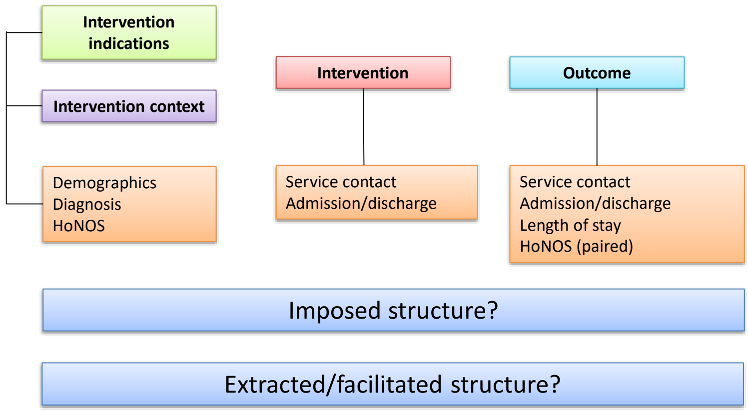
The 'patient journey' from the unmodified mental health EHR



NHS National Institute for Health Research

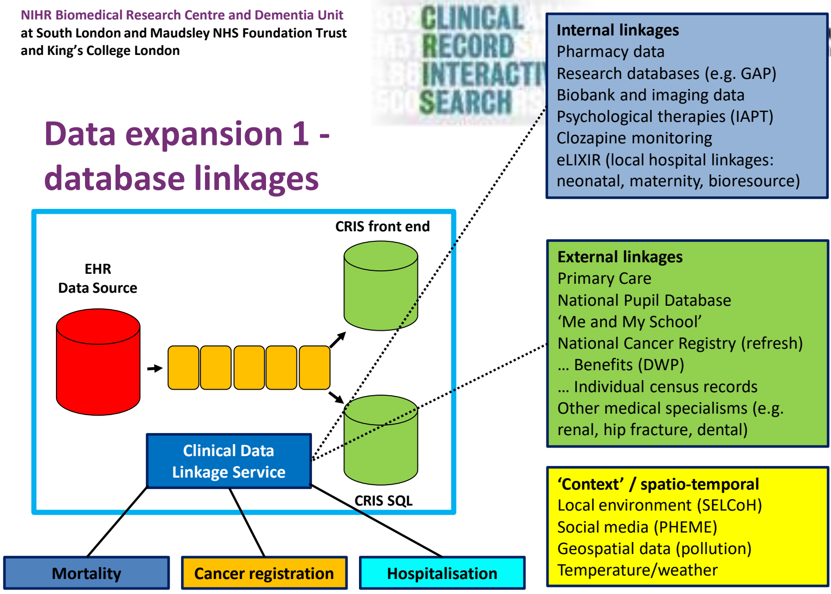
CLINICAL RECORD INTERACTIVE SEARCH

The 'patient journey' from the unmodified mental health EHR



NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

Data expansion 1 - database linkages



Applications – hospitalisation data linkage


ELSEVIER

Alzheimer's & Dementia

Featured Article

Predictors of emergency department attendance by people with dementia in their last year of life: Retrospective cohort study using linked clinical and administrative data

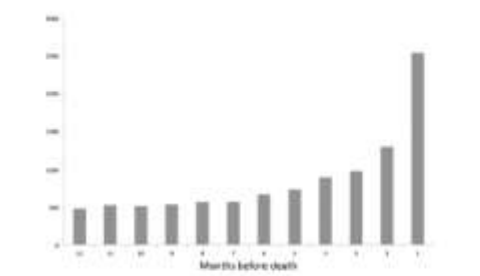


Fig. 1. Frequency of ED attendance by proximity to death in months (n = 10,361 attendances); 24.6% (n = 2546) of ED attendances in the last year of life occurred in the month before death. Abbreviation: ED, Emergency Department.

 https://www.ncbi.nlm.nih.gov/pubmed/28838779

CLINICAL RECORD INTERACTIVE SEARCH

The 'patient journey' from the mental health EHR

Intervention indications

Intervention context

Demographics
Diagnosis
HoNOS
Linked data

Intervention

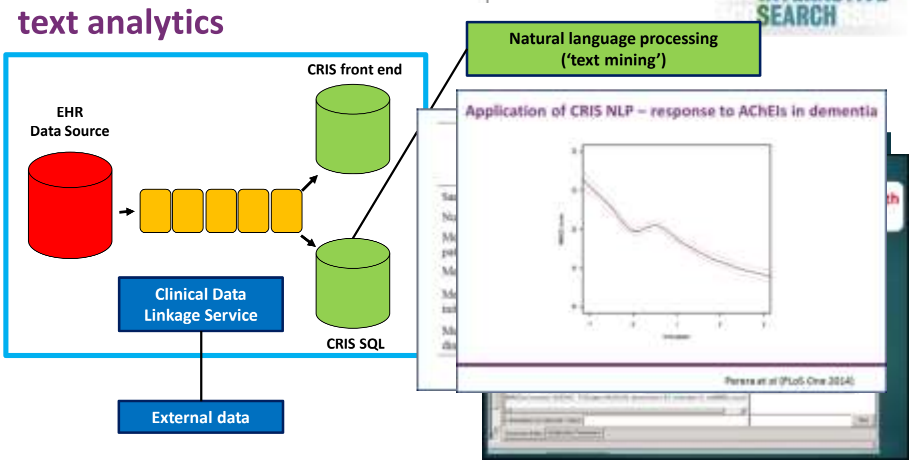
Service contact
Admission/discharge

Outcome

Service contact
Admission/discharge
Length of stay
HoNOS (paired)
Linked data

NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

Data expansion 2 - text analytics



NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

CRIS with text-mining (CRIS-CODE)

Intervention indications

Intervention context

Intervention

Outcome

☑ = complete
✍ = in progress

Symptoms
Psychosis
- Positive ☑
- Negative ☑
- Disorganisation ☑
- Manic ☑
- Catatonic ☑

Affective
- Mood ☑
- Biological symptoms ☑
- Instability ☑
- Anxiety ✍
- Obsessive/compulsive ☑
- Somatic ✍

Behaviour
- Agitation/withdrawal ☑

Other
- Insight ☑

Context
- Cognitive function ☑
- Social care ☑
- Living alone ☑
- Diagnosis ☑
- Social status ✍
- Education ☑ ✍
- Physical disorders ☑
- Investigations ✍

Substances
- Smoking ☑
- Alcohol ✍
- Cannabis ☑
- Mephedrone ☑
- Amphetamine ☑
- Cocaine ☑
- Others ✍

Interventions
Pharmacotherapy ☑
- Polypharmacy ☑
- Adherence/compliance ☑

Psychotherapy
- CBT ☑ (receipt, offer, etc.)
- DBT ✍
- CAT ✍
- Family ✍
- Supportive/behavioural ✍

Outcomes
Adverse drug events
- Extrapyramidal ☑
- Other ☑ ✍

Symptom trajectories ☑ ✍
- Improvement
- Deterioration

General mental health ☑ ✍
- Improvement
- Deterioration

NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

NHS National Institute for Health Research

Symptoms/phenotyping NLP

CLINICAL RECORD INTERACTIVE SEARCH

Poor motivation
Blunted / flat affect
Diminished eye contact
Emotional withdrawal
Poor rapport
Social withdrawal
Poverty of speech
Apathy
Concrete thinking
Poverty of thought

Elation
Euphoria
Elevated mood
Insomnia
Disturbed sleep
Irritability
Grandiosity
Pressured speech

Mood instability
Affective instability
Emotional instability

Hallucinations
Delusions
Hostility
Arousal
Aggression
Agitation
Suspicious
Paranoia
Persecutory ideas

Catalepsy
Echolalia
Echopraxia
Immobility
Mannerism
Rigidity
Posturing
Perseverance
Stupor
Mute
Waxy flexibility

Reduced coherence
Formal thought disorder
Circumstantial speech
Tangential speech
Derailment
Flight of ideas
Thought block

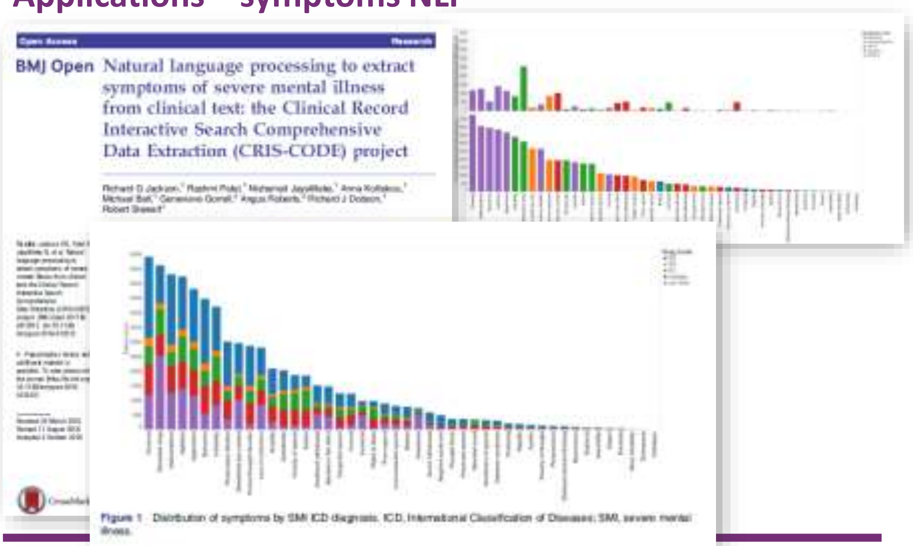
Low mood
Anhedonia
Guilt
Hopelessness
Reduced appetite
Suicidality
Poor concentration
Weight loss
Lowered energy / anergia
Helplessness
Psychomotor retardation
Worthlessness
Tearfulness

28,000+ annotations

NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

NHS National Institute for Health Research

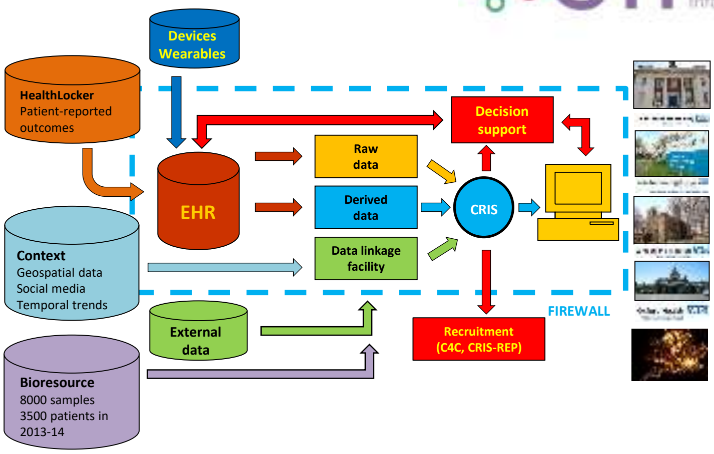
Applications – symptoms NLP



 https://bmjopen.bmj.com/content/7/1/e012012.long

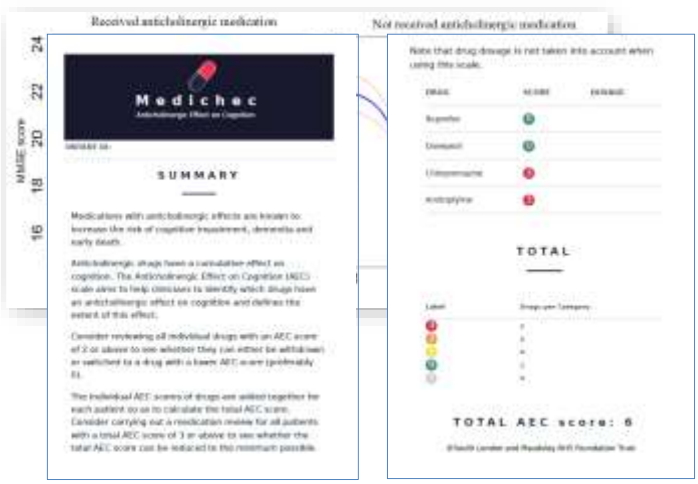
Integrated informatics

CTI Centre for Translational Informatics



Decision support: co-prescribing in dementia

www.medichec.com



- CRIS
- Co-prescribing
- CRIS/linkages
- Outcomes
- Medichec
- Information and advice
- GPs, public, NHS etc.

NIHR Biomedical Research Centre and Dementia Unit
at South London and Maudsley NHS Foundation Trust
and King's College London

CTI Centre for Translational Informatics

NHS National Institute for Health Research

CLINICAL RECORD INTERACTIVE SEARCH

Clinical informatics – the multi-disciplinary team

- **Leads**
  - Matthew Broadbent, Richard Dobson, Stephen Docherty, Rob Stewart
- **Administration / management**
  - Debbie Cummings, Anna Kolliakou, Megan Pritchard
- **Technical team**
  - Amelia Jewell (data linkage), Shanmukha Gudiseva (compute), Hitesh Shetty (data extraction), Jyoti (NLP support)
- **Epidemiology / Clinical**
  - Craig Colling, Lauren Carson, Lauren Cross, Johnny Downs, Rina Dutta, Sophie Epstein, Daniela Fonseca de Freitas, Emma Francis, Richard Hayes, Giouliana Kadra, Christoph Mueller, Rashmi Patel, Gayan Perera, Kate Polling, Katherine Sleeman, Brendon Stubbs
- **Bioinformatics / Computer Science / NLP**
  - Elizabeth Baker, Daniel Bean, Andre Bittar, David Chandran, Amos Folarin, Karen Hodgson, Zina Ibrahim, Ehtesham Iqbal, Julia Ive, Daniel Leightly, Stephen Newhouse, Angus Roberts, Hegler Tissot, Sumithra Velupillai, Natalia Viali, Honghan Wu
- **PhD studentships**
  - Delia Bishara, Andrea Fernandes, Nikeysha Bell, Katrina Davies, Usha Gungabissoon, Richard Jackson, Leo Koeser, Alice Wickersham,
- **Oversight and Governance**
  - Felicity Callard, Patrick Green, Jenny Liebscher, Sean Maskey, Katharine Rimes, Murat Soncul
- **SLAM partners**
  - Nicola Byrne, Fiona Gaughran, Anthony Schnarr
- **University of Sheffield collaborators**
  - Kalina Bontcheva, Genevieve Gorrell, Ian Roberts

brc.slam.nhs.uk

ANALYSIS GROUP

Artificial Intelligence in Literature Reviews

Case Studies

Eric Wu
Managing Principal, Analysis Group, Inc.

BOSTON CHICAGO DALLAS DENVER LOS ANGELES MENLO PARK NEW YORK SAN FRANCISCO WASHINGTON, DC • BEIJING • BRUSSELS • LONDON • MONTREAL • PARIS

ANALYSIS GROUP

Introduction

- Literature reviews, both systematic (SLRs) and targeted (TLRs), are a critical component of various decision-making processes in healthcare
- Literature reviews aim to be unbiased and have high recall (i.e. capture all relevant articles)
- In both TLRs and SLRs, screening of a large number of articles is typically required
  - Only a small fraction of articles are typically selected
- The amount of medical literature has been growing rapidly – with approximately 46% increase of new MEDLINE articles each year
- We need to be ready to handle and review this every growing body of medical literature

ANALYSIS GROUP

Introduction

- Currently, most literature reviews are 100% human based, which are both resource intensive and lacking transparency
- The rapidly increasing body of medical literature will continue to put pressure on future literature review and will potentially have several detrimental consequences:
  - Presenting challenges to support time sensitive decision making
  - Cost prohibitive to research with limited resources
  - Encourages conservative search criteria to reduce review time, which increases bias and reduces recall
  - Leads to human reviewer fatigue, which may also increase bias and reduce recall

ANALYSIS GROUP

Objectives

- We are presenting two case studies to illustrate how AI can help address challenges related to quality, transparency, and labor-intensiveness in literature reviews

ANALYSIS GROUP

Artificial Intelligence in Literature Reviews

- Overview of general approach for AI-based literature reviews

Initial abstract identification

↓

Text feature extraction - NLP

↓

Manual screening for a subset - Training

↓

Predictive model development - ML

↓

Abstract relevance prediction

↓

AI assisted relevance-based screening

Text Feature Extraction

- Text feature extraction (NLP) includes:
  - Customized features: e.g., presence of specific keywords, created explicitly by the reviewer
  - Automated feature generation
    - Topic modeling (LDA)
    - Bag-of-words based approach
    - Neural network-based approach
    - Syntactic feature generation

Predictive model development

- Predictive modeling (ML) include:
  - Logistic regression
  - Neutral networks
  - Support vector machines
  - Naïve Bayers classifiers
  - Decision trees or random forests
  - Assemble classifiers combining several of the above

Analysis Group

Case Study 1: Systemic Review

- Manual review vs. AI-based approaches

Manual approach	AI-based approach
All abstracts are screened by 2 human reviewers + 1 reconciler	Subset of abstracts screened by human reviewers
	Predictive model is developed
	Remaining abstracts are screened by algorithm and a human reviewer
	Discrepancies are resolved

Analysis Group

Case Study 1: Systematic Review

- AI-based review process

Total abstracts downloaded: **3,800**

↓

Manual screening and prediction model training (2 reviewers + 1 reconciler): **8%**

↓

Predictive models retrain (1 Reviewer + AI + 1 reconciler): **8%**

↓

Additional screening (1 Reviewer + AI + 1 reconciler): **10%**

↓

Relevant full-texts identified: **150**

Case Study 1: Systemic Review

- Differences in resources used:

Component	Manual review	AI-based review
Reviewer 1	95 hours	95 hours
Reviewer 2	95 hours	6 hours
Reconciler	16 hours	10 hours
Programmer	-	18 hours

Case Study 2: Targeted Review

- Manual vs. AI-based approaches

Manual approach

All abstracts are screened by 1 human reviewer

AI-based approach

Subset of abstracts screened by 1 human reviewer

Predictive model is developed

Abstracts meeting a relevance cutoff are screened by a human reviewer

ANALYSIS GROUP

Case Study 2: Targeted Review

- AI-based review process

Total abstracts downloaded: **2,200**

↓

Manual screening and prediction model training (1 reviewer): **21%**

↓

Additional articles screened based on relevance cutoff: **14%**

↓

Relevant full-texts identified: **130**

ANALYSIS GROUP

Case Study 2: Targeted Review

- Differences in resources used:

Component	Manual review	AI-based review
Reviewer	65 hours	17 hours
Programmer	-	18 hours

Conclusions

- AI-based approaches permit wide scope of literature reviews to be conducted within strict time constraints
- The relative efficiency of AI-based reviews compared to human-reviewed studies increases with the number of studies initially identified
- An often overlooked aspect of literature reviews is reproducibility and transparency. When based on a clearly specified approach, AI-based literature reviews can be reproduced more easily than fully manual reviews. Existing algorithm can also be used as the base for future literature search.
- By removing the need to narrow down search strategies excessively, AI-based reviews may reduce bias and improve recall
- AI-based reviews may be extended to the screening of full-text articles with caution.

Artificial Intelligence and Treatment Effects

Using Claims Data and Propensity Score Models in a High Dimensional Setting

Jimmy Royer
Principal, Analysis Group, Inc.

ANALYSIS GROUP

Estimating Treatment Effect using Non-Experimental Retrospective Data

Using Claims Data to Create a High Dimensional Dataset

- Physicians' treatment choices are non-random and usually undocumented in retrospective studies
- There is a need to find variables that can proxy patients' unobservable characteristics yielding to such treatment decisions
- Although Claims Data do not usually include many of such patient characteristics, historical claims (diagnoses, procedures, drugs) can be used to create a High Dimensional dataset
  - For example, past Rx, diagnoses, and procedures may explain why someone is treated with a particular drug therapy today

ANALYSIS GROUP

Estimating Treatment Effect using Non-Experimental Retrospective Data

How to Minimize Risk of Overfitting with High Dimensional Datasets

- High Dimensional Claims Datasets provide thousands of possible confounding variables to use in estimating treatment effects
- There is a real risk of overfitting models and having lack of common support, especially in estimating the treatment decision equation. In that case, multivariable models will not necessarily perform better than a univariate model
  - Traditionally the solution was to use clinical insights to select a limited number of variables
  - AI provides an unbiased and agnostic framework (i.e. regularization – or a penalty for "too many" variables) to address the risk of overfitting on the one hand, and providing potentially unbiased measures on the other hand

ANALYSIS GROUP

Estimating Treatment Effect using Non-Experimental Retrospective Data

Possible Estimation Strategies

- Univariate Analysis
  - Uncommon and generally biased because of the potential omission of important confounding variables
- Multivariable Linear Regression (GLM, Proportional Hazard Model)
  - Confounding variables (correlated to the treatment) enter linearly in the conditional expectation
  - Common, but generally too restrictive
- Two-Step Procedures (Propensity Score, IPTW, tMLE)
  - Confounding variables (correlated with the treatment) enter non-linearly in the conditional expectation – i.e. more flexible and less restrictive
  - Provide a framework to estimate unbiased treatment effects when used in combination with the appropriate data and models

ANALYSIS GROUP

Estimating Treatment Effect using Non-Experimental Retrospective Data

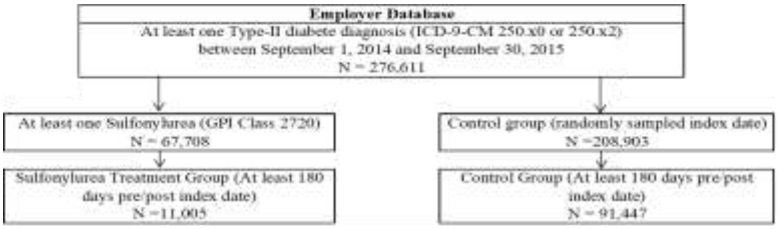
Suggested Approach

- An employer insurance claims database is used to simulate an outcome (Survival rate) assuming a treatment effect of 1.5 (RR)
- The claims database provide a framework to build a high dimensional database with thousands of potential predictors
- Machine Learning Models are used to estimate the treatment selection equation
- A proportional hazard model is used to estimate the treatment effect

ANALYSIS GROUP

Simulation Experiment using Claims Data

Data Selection and Outcome



Outcome Measures	Unweighted Number of Events		Unweighted Person-Years		Unweighted Incidence Rate (per 100 person-years)	
	Case	Control	Case	Control	Case	Control
Myocardial infarction (MI)[1]	178	1301	11246.9	91303.09	1.58	1.42

Note:
[1]ICD-9-CM (410.xx)

ANALYSIS GROUP

Treatment Selection Model and AI Algorithms

Brief Outline of AI Algorithms

- Logistic Regression with LASSO Penalty
  - Covariates *marginally* related to the outcome are selected out
- Random Forest
  - A tree-based methodology where a large number of trees are estimated, each from a bootstrap sample
  - The final prediction is the mean of all the predicted propensity scores computed over all trees
- Gradient Descent Boosting
  - A tree-based methodology where each tree is estimated using the same training sample (no bootstrap)
  - There is an iterative process with an updating of parameters at each iteration
- Deep Neural Networks
  - A neural network structure is organized in layers of neurons connected by synapses

Model Estimation and Results

Real Effect = 1.5

Model	Estimated ATE
Univariate Model	1.311 ± 0.098
Standard Propensity Score Approach	1.374 ± 0.104
Logistic Regression with Lasso (HD)	1.496 ± 0.119
Random Forest (HD)	1.533 ± 0.123
Boosting (HD)	1.501 ± 0.118
Deep Learning (HD)	1.502 ± 0.119

Conclusions

Key Takeaways

- AI algorithms can be useful to estimate causal relationships in retrospective studies that use Propensity Score models when:
  - There is a very large potential number of covariates
  - The number of observations is large
  - The studied outcome is a rare event
- Logistic LASSO, Gradient Descent Boosting and Neural networks perform very well
- Random Forest seems to slightly underperform

Analysis Group

Backup and References

Analysis Group

Model Comparisons

Variable Importance Top 10 based on mean rank

Rank	Lasso	Random Forest	Boosting	Deep Learning
1	Drug Class 2725 (Biguanides)	Drug Class 2725 (Biguanides)	Drug Class 2725 (Biguanides)	Drug Class 2725 (Biguanides)
2	Drug Class 2710 (Insulin)	dx25002 (Uncontrolled Diabetes Mellitus)	Drug Class 2755 (Dipeptidyl Peptidase-4 Inhibitors)	Region (Mid-Atlantic)
3	dx25002 (Uncontrolled Type II Diabetes Mellitus)	Drug Class 9705 (Parenteral Therapy Supplies)	Drug Class 2799 (Antidiabetic Combinations)	dx25000 (Type II Diabetes Mellitus not stated as uncontrolled)
4	Drug Class 2799 (Antidiabetic Combinations)	Drug Class 2710 (Insulin)	Drug Class 2710 (Insulin)	dx25003 (Uncontrolled Type I Diabetes Mellitus)
5	Region (Mid-Atlantic)	Age (40-49)	dx25002 (Uncontrolled Type II Diabetes Mellitus)	Drug Class 3940 (HMG COA Reductase Inhibitor Combinations)

ANALYSIS GROUP

Model Comparisons

Variable Importance Top 10 based on mean rank

Rank	Lasso	Random Forest	Boosting	Deep Learning
6	Drug Class 2755 (Dipeptidyl Peptidase-4 inhibitors)	dxV7612 (Mammogram Screening)	Region (Mid-Atlantic)	dx25001 (Type I Diabetes Mellitus not stated as uncontrolled)
7	Age (50-59)	Age (50-59)	Drug Class 9705 (Parenteral Therapy Supplies)	Drug Class 2710 (Insulin)
8	Age (40-49)	Drug Class 3610 (ACE Inhibitors)	Drug Class 3940 (HMG COA Reductase Inhibitor Combinations)	dx7999 (Unspecified Cause of Morbidity)
9	dxV7612 (Mammogram Screening)	dx25003 (Uncontrolled Type I Diabetes Mellitus)	Age (70-79)	dx79021 (Nervousness)
10	Age (30-39)	Region (Mid-Atlantic)	Age (40-49)	dx25002 (Uncontrolled Type II Diabetes Mellitus)

ANALYSIS GROUP

Sensitivity Analysis

Model	30 Least Correlated Variables Removed	Only 20 Most Correlated Variables Included (LASSO only)
Lasso	1.496 ± 0.119	1.436 ± 0.11
Random Forest	1.456 ± 0.115	-
Boosting	1.504 ± 0.118	-
Deep Learning	1.499 ± 0.117	-

Selected References

[1] Imbens GW, Rubin DB. Causal Inference in Statistics, Social, and Biomedical Sciences. Cambridge University Press; 2015.

[2] Franklin JM, Schneeweiss S, Polinski JM, and Rassen JA. Plasmode Simulation for the Evaluation of Pharmacoepidemiologic Methods in Complex Healthcare Databases. Computational Statistics and Data Analysis 2014; 72: 219-226. https://doi.org/10.1016/j.csda.2013.10.018

[3] Karim ME, Pang M, and Platt RW. Can We Train Machine Learning Methods to Outperform the High-dimensional Propensity Score Algorithm? Epidemiology 2018; 29 (2). https://doi.org/10.1097/EDE.0000000000000787.

[4] McCaffrey, Daniel F, Ridgeway G, Morral AR. Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies. Psychological Methods. 2004; 9 (4): 403. https://doi.org/10.1037/1082-989X.9.4.403

Selected References

[5] Westreich D, Lessler J, Jonsson-Funk M. 2010. Propensity Score Estimation: Machine Learning and Classification Methods as Alternatives to Logistic Regression. Journal of Clinical Epidemiology. 2010; 63 (8): 826. https://doi.org/10.1016/j.jclinepi.2009.11.020

[6] Efron B, and Hastie T. Computer age statistical inference: Algorithms, evidence, and data science. Vol. 5. New York: Cambridge University Press; 2016.

[7] James RG, Witten D, Hastie T., and Tibshirani R. An introduction to statistical learning with applications. New york: Springer; 2013.

[8] Dadson N, Pinheiro L, Royer J. Decision Making with Machine learning in Our Modern, Data-Rich Health Care Industry. In Birnbaum HG, Greenberg PE, editors. Decision Making in a World of Comparative Effectiveness Research. Adis, Singapore; 2017. pp. 277-289. https://doi.org/10.1007/978-981-10-3262-2_21

ANALYSIS GROUP

Selected References

[9] Casella G, Berger RL. Statistical inference. 2nd edition. Australia: Thomson Learning; 2002.

[10] H2O.ai Team. R Interface for H2O.R. In package version 3.10.4.6. 2017. h2o: https://github.com/h2oai/h2o-3

[11] Therneau TM. A Package for Survival Analysis in S. Version 2.38. 2015. https://CRAN.R-project.org/package=survival.