

Towards machine learning informed early detection of dementia in UK primary care

Modeling Patient Data for Diagnosis and Risk

S109

Dr Elizabeth Ford and Dr Johannes Starlinger

Brighton and Sussex Medical School

Twitter: @drelizabethford

#AMIA2019



Disclosure

I and my spouse/partner have no relevant relationships with commercial interests to disclose.

Learning Objectives

After participating in this session the learner should be better able to:

- Understand the opportunities and challenges for early detection of dementia using routinely collected clinical data from primary care.

UK primary care in the NHS

- 98% of the population in the UK are registered with an NHS general practitioner.
- Clinical data is collected in commercial electronic record systems using a clinical coding system called Read codes
- Some GP clinics supply their data to centrally housed research databases.



Dementia

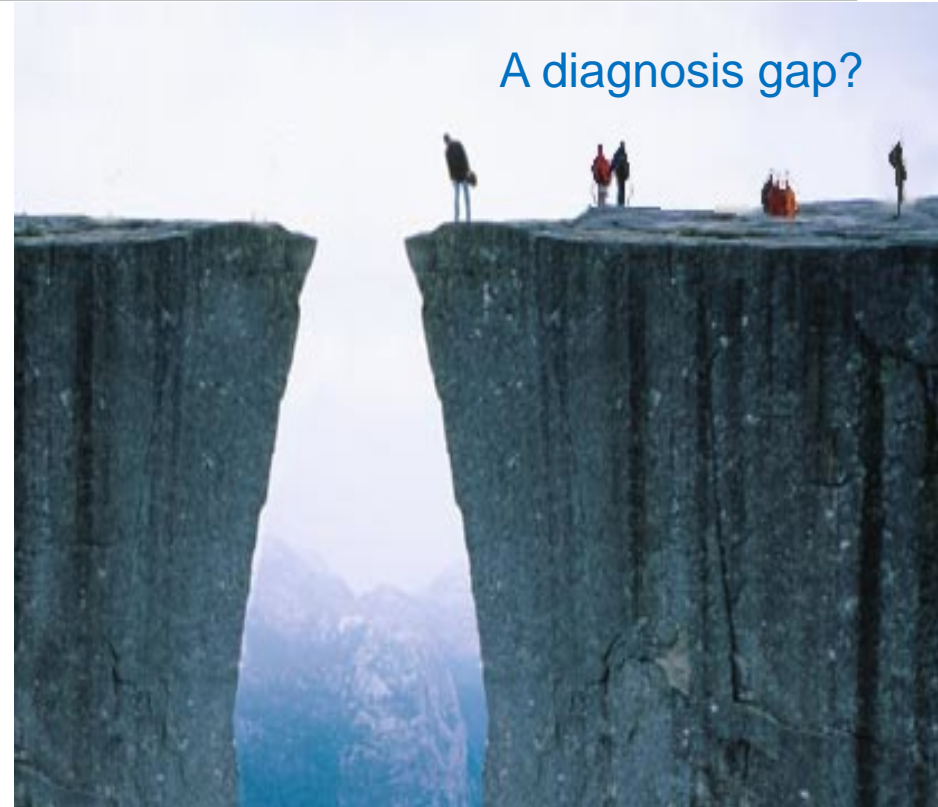
- A disease characterized by progressive loss of memory and cognitive function
- A devastating impact and a pressing public health concern
- Global economic impact US \$604 billion
 - estimated to increase to \$1tr by 2030



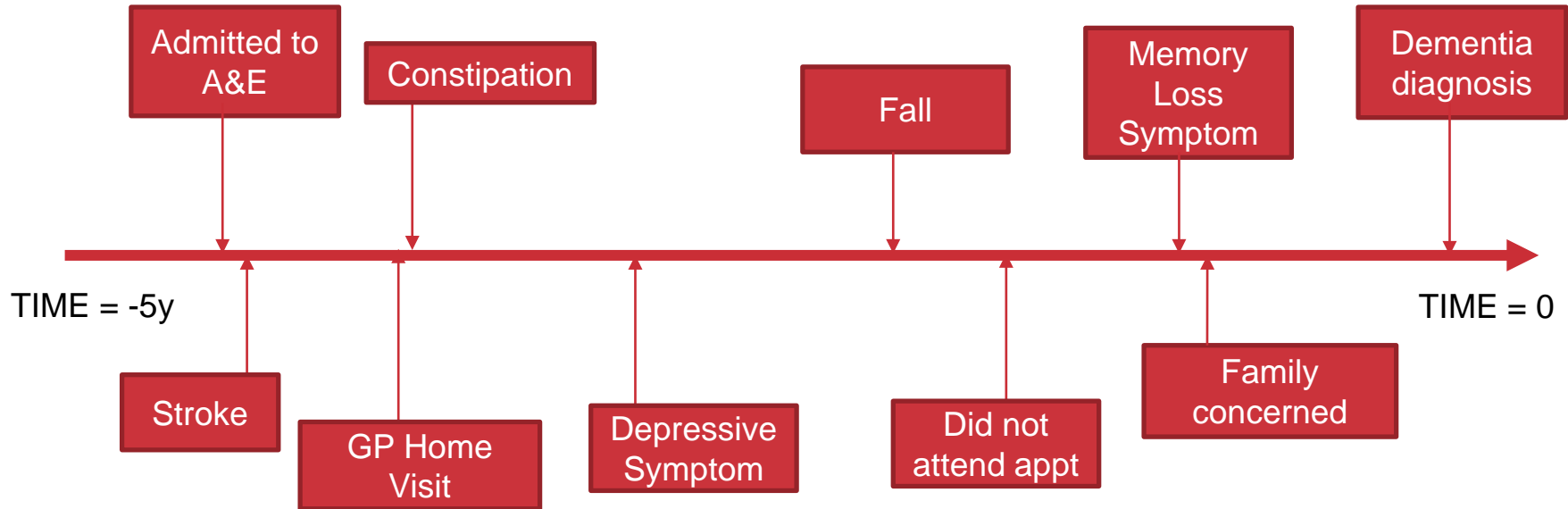
Dementia

Currently, diagnosis is often made:

- At a time of crisis (e.g. following a burn or a fall)
- When the family are struggling to cope with care needs
- Too late for the person with dementia to express their financial and care preferences and make plans.



A primary care clinical record in the 5 years before dementia diagnosis



Motivation for this study

- A clearer picture of dementia accrues along the patient timeline prior to diagnosis, in the form of codes recorded by the GP.
- We may be able to take this patient pathway and read it like a sentence where the last word in the sentence is “Dementia”.
- We can use machine learning techniques determine patients who have a high risk of having a diagnosis of dementia in the future.
- If we can detect the condition earlier, we may be able to flag this to GPs and help them to make the diagnosis in a more timely way, before a crisis occurs.

Research objectives

- 1) Investigate how early before current diagnosis dementia could be detected accurately.
- 2) Determine a minimum set of best features for discriminating between dementia cases and controls.
- 3) Investigate if there are any differences between predictive models for Alzheimer's disease and Vascular dementia.

- Clinical Practice Research Datalink
 - Represents 8% of UK population.
 - Not for profit organisation, part of UK Government.

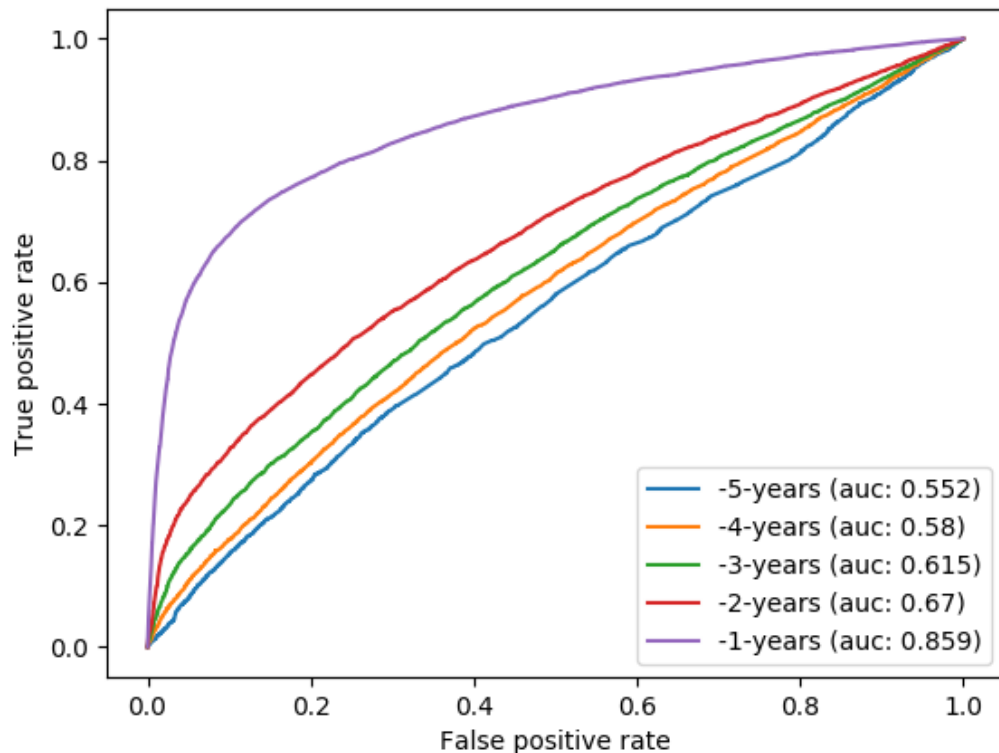


- Case-control design
 - 47,000 English patients aged 65+y with dementia diagnosis code entered in time period 2000-2012 (any cause dementia).
 - 1:1 matched controls (age, sex, GP practice).
- Full record (all Read Codes) at least 3 years and up to 5 years prior to dementia diagnosis or matched date. Codes partitioned into year-long blocks.
- Feature list created after meta-analysis and GP survey – 77 separate predictive features, each operationalised by a code list drawn up *a priori*.

- Random Forest Classifiers
 - For each year before diagnosis (-5y only, -5 & -4, -5 to -3, -5 to -2, all 5 years)
 - Training set 67%; testing set 33%
 - Random Forest intrinsically includes feature weighting and selection.
- Divided features into three groups:
 - Long standing risk factors e.g. high BMI, smoking, hypertension, stroke
 - Increasing frailty or prodromal symptoms e.g. falls, repeated infections, wounds or burns, depression.
 - Indicators that the GP has noticed memory loss symptoms; initiation of dementia diagnosis pathway

Results – all cause dementia

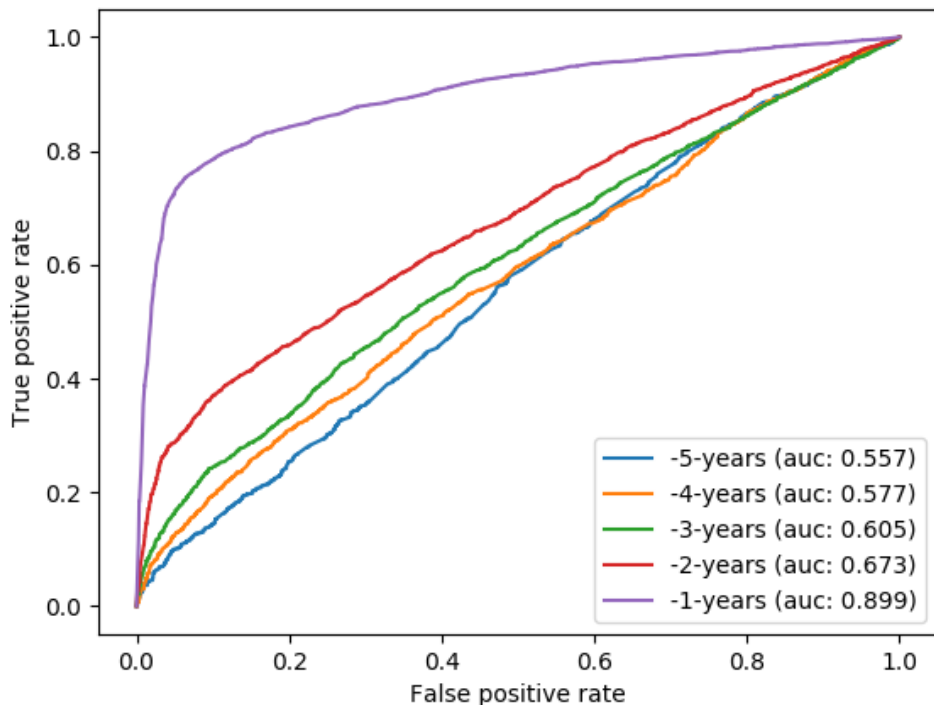
ROC curve showing discrimination between cases and controls for each of 5 years before dementia diagnosis



Top 10 predictive features for each year before diagnosis

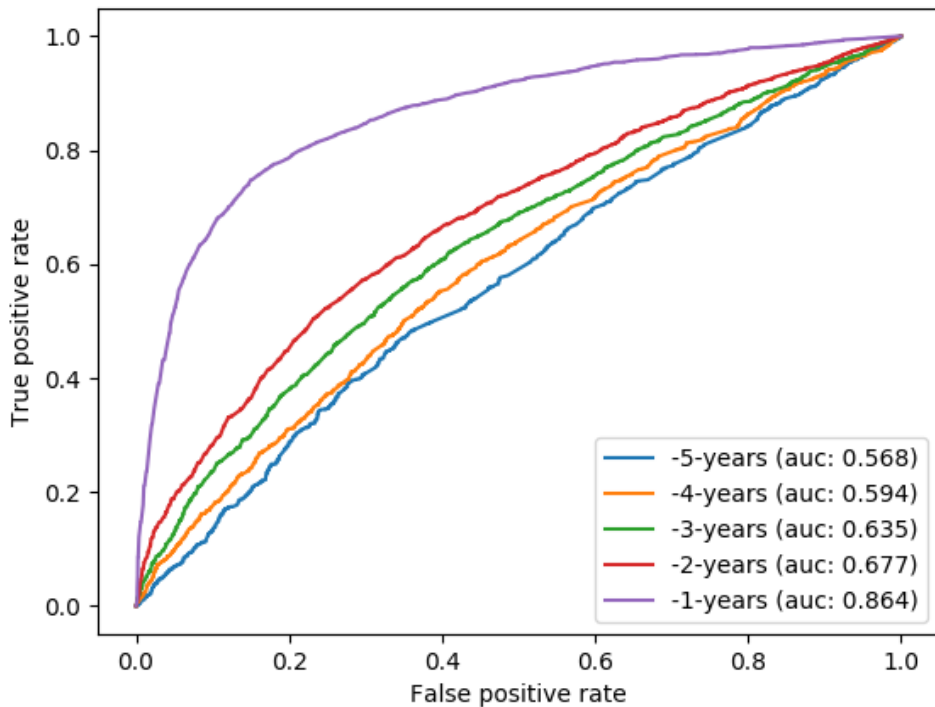
	Year -5	Year -4	Year -3	Year -2	Year -1
1	Anti-depressants	Anti-depressants	Dressing of wounds burns or ulcer	Memory loss codes	Memory loss codes
2	Dressing of wounds burns or ulcer	Dressing of wounds burns or ulcer	Memory loss codes	Dressing of wounds burns or ulcer	Referral to Psychiatrist Neurologist or Geriatrician
3	GP Home visit	GP Home visit	GP Home visit	GP Home visit	GP Home visit
4	Antipsychotics	Is a smoker	Is a smoker	Anti-depressants	Anti-depressants
5	Is a smoker	Antipsychotics	Anti-depressants	Is a smoker	Cognitive Screening (MMSE)
6	Visit to Emergency Dept	Infections	Infections	Infections	Did not attend code
7	Hospital admission	Hospital Admission	Antipsychotics	Antipsychotics	Antipsychotics
8	Infections	Urinary Tract Infection	Hospital Admission	Visit to Emergency Dept	Infections
9	Urinary Tract Infection	Visit to Emergency Dept	Visit to Emergency Dept	Did not attend code	Dressing of wounds burns or ulcer
10	Hypertension	Hypertension	Urinary Tract Infection	Hospital Admission	Is a smoker

Alzheimer's disease



	Year -5	Year -4	Year -3	Year -2	Year -1
1	Anti-depressants	GP Home visit	Memory loss codes	Memory loss codes	Memory loss codes
2	GP Home visit	Dressing of wounds burns or ulcer	GP Home visit	GP Home visit	Referral to Psychiatrist Neurologist or Geriatrician
3	Dressing of wounds burns or ulcer	Infections	Dressing of wounds	Dressing of wounds	GP Home visit
4	Anti-psychotics	Anti-depressants	Infections	Anti-depressants	Anti-depressants
5	Infections	Anti-psychotics	Anti-depressants	Is a smoker	Cognitive Screening (MMSE)

Vascular dementia



	Year -5	Year -4	Year -3	Year -2	Year -1
1	Anti-depressants	GP Home visit	Dressing of wounds	Memory loss codes	Memory loss codes
2	GP Home visit	Anti-depressants	GP Home visit	GP Home visit	Referral to Psychiatrist Neurologist or Geriatrician
3	Dressing of wounds	Infections	Infections	Infections	GP Home visit
4	Is a smoker	Dressing of wounds burns or ulcer	Is a smoker	Dressing of wounds burns or ulcer	Anti-depressants
5	Anti-psychotics	Is a smoker	Anti-depressants	Is a smoker	Cognitive Screening (MMSE)

Results summary

- Average to poor discrimination more than 1 year before diagnosis.
- Features which enable accurate discrimination are those that indicate GP has already initiated dementia diagnosis pathway.
- Secondary features which enable discrimination are indicators of increasing frailty, possible prodromal symptoms (e.g. depression) and care needs.
- Few differences between all cause, Alzheimer's and vascular dementia.

Implications and future directions

- Implication: Automated detection of dementia in routinely collected clinical data is unlikely to outperform the clinician who created the data.
- Next steps for improvement:
 - Identify temporal patterns in data which may have gone unnoticed by clinicians (e.g. worsening indicators of memory problems)
 - Link in other sources of data – wearables, patient report, genomic and other clinical sources e.g. Emergency Department
- Limitations
 - Possible misclassifications in dataset
 - Sparse data and long gaps between consultations for some patients
 - Elimination of age as a predictor due to matched design

Thanks to the Astrodem team:



Dr Johannes Starlinger



Prof Jackie Cassell



Dr Philip Rooney



Prof Seb Oliver

Please contact me for further information: e.m.ford@bsms.ac.uk

ASTRODEM


is supported by
The Wellcome Trust
The University of Sussex
Santander Universities



University of Sussex



Santander



brighton and sussex
medical school

Find out more:

www.bsms.ac.uk/astrodem



ASTRODEM