A pilot study of the use of routinely held NHS clinical data to identify undiagnosed dementia

WITH PLYMOUTH UNIVERSITY

Athanasios Anastasiou
Javier Escudero
Emmanuel Ifeachor

NHS National Institute For Health Research

Camille Carroll
Steven Pearson
John Zajicek
William Henley

NHS Devon Commissioning Services

Peter Turnbull
Graham Sykes

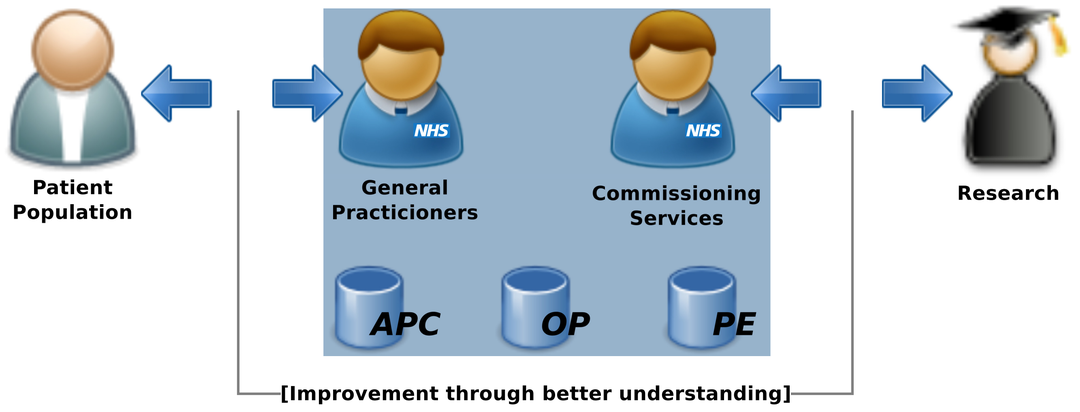
[Data]

Routinely collected NHS data from 18 General Practice surgeries across Devon were obtained via the NHS Devon Commissioning Services department.

The data covered a two year time period (2010-2012) and included all senior patients at or above 65 years of age.

The data were composed of three different datasets describing healthcare received by patients at the levels of Admitted Patient Care (APC), Outpatient Departments (OP) and Practice Events (PE).

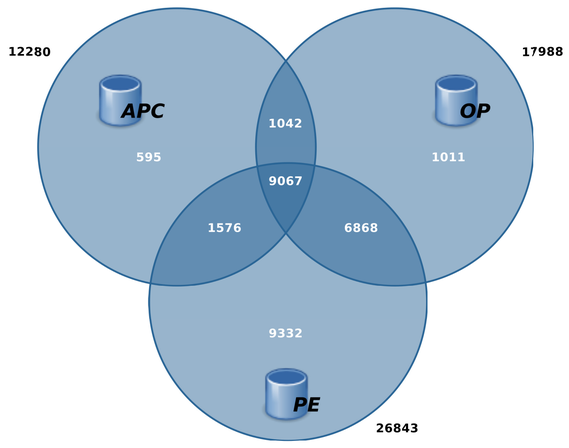
Taken together the three datasets contain a multitude of parameters for tenths of thousands of patients.



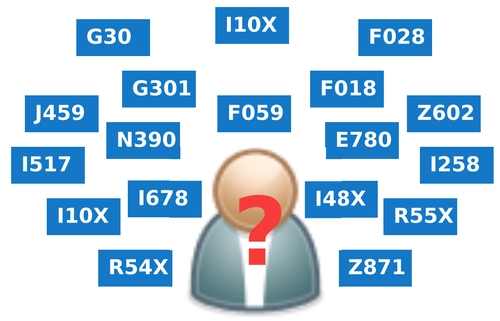
Out of all these parameters, the research team focused initially on the set of diagnostic codes assigned to patients during their visits to their local surgeries.

These codes contain characterisations of the overall health condition of a subject at different points in time such as Hypertension, Atrial Fibrilation, Fall Event and others.

Therefore, the primary objective of identifying missdiagnosed cases of dementia now became a matter of answering the question: "Is it possible to use the diagnostic codes assigned to each patient with a diagnosis to construct a profile of key dementia characteristics and then evalute the similarity of the profiles of non-demented patients against this profile?".



Patient populaton sizes distributed amongst the three available datasets



Is it possible to infer a case of missdiagnosed dementia just by examining a subset of the patients ICD diagnostic codes?

[Methodology]

An extract of the ICD diagnosis codes of 10345 patients was constructed by collecting all diagnostic codes assigned to them within the available two year time-interval. This extract was then split into the subpopulations of "Control" and "Demented".

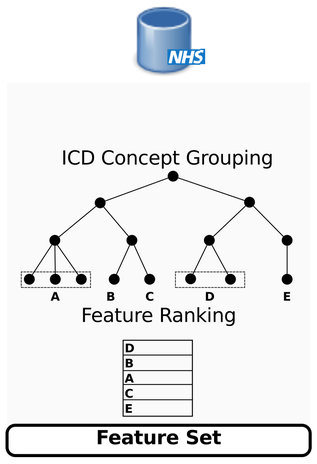
A Dementia characterisation was assigned to patients with specific diagnostic codes associated with a specific dementia such as Alzheimer's Disease, Multi-infract dementia, Subcortical vascular dementia, Mild cognitive disorder and others (i.e. Codes F00-F10 and G30-G31 in the ICD disease classification standard).

This resulted in a large array of 3289 different codes, describing various clinical conditions in extreme detail.

In order to create representative profiles of the subpopulations, a clinically meaningful way of grouping together codes into group features was devised.

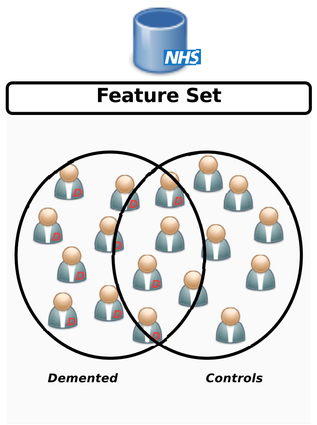
In this way, each patient was assessed according to a set of clinical concept features (e.g. cardiovascular factors) rather than individual codes (e.g. Hypertension.)

These grouped features where used by a Bayesian classifier trained on the dataset extract to classify cases of "Demented" and "Control" individuals.



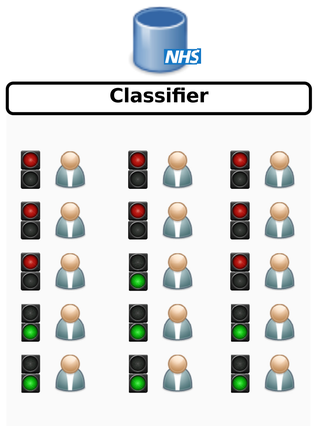
Clinically relevant code grouping by exploiting the structure of the ICD classification

Feature ranking by maximising cross correlation with diagnosis while minimising feature inter-correlation.



Naive Bayes Classifier estimating the probability density function of each feature.

The classifier is based on computing the probability that each code appears in each of the two diagnostic groups.



Ten-fold cross validation

Subjects are assigned a score representing their probability of being "Demented" according to the profile discovered by the Classifier.

[Results]

Concept Grouping	Controls	Demented	Condition	Sensitivity	Specifity	Accuracy	AROC
Without	8226	1750	Controls	80.00%	82.46%	82.37%	0.873
	74	296	Demented				
With	6684	3292	Controls	83.78%	67.00%	67.60%	0.846
	60	310	Demented				

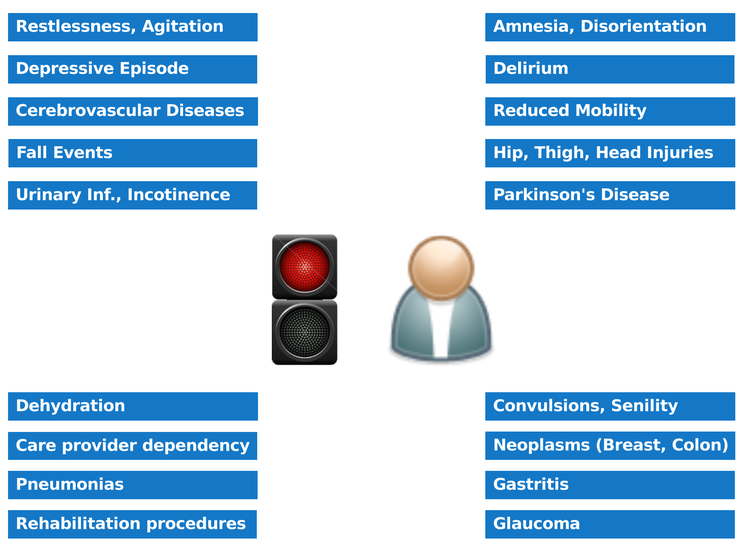
Routinely collected NHS data were succesfuly employed to identify Dementia profiles automatically.

Classification rates remained constantly above 80%

Exploiting the structure of the ICD-10 encoding to group diagnostic codes into concept groups was found to have a marked increase in the sensitivity of the classification.

Automatically grouping the diagnosis codes into clinical concept groups leads to higher sensitivity and therefore a lower chance for misdiagnosed patients to go unnoticed while reviewing a surgery's patient register.

Ongoing research is focusing on investigating alternative classification techniques over routinely collected NHS data as well as investigating the full potential of the automatic code grouping method in order to increase classifier sensitivity and provide a tool to help clinicians deal effectively with large ammount of encoded clinical data.



The automatic grouping of the patient's codes that was automatically discovered by the feature selection process and was used to assess the Dementia risk of the patients in the given dataset.