Developing clinical informatics for mind and brain health in Cambridge

Rudolf Cardinal
University lecturer in clinical informatics, University of Cambridge
Honorary consultant liaison psychiatrist, Cambridgeshire & Peterborough NHS FT / Cambridge University Hospitals NHS FT

Thu 6 Dec 2018, KCL BRC / CRIS 10th anniversary; 15′

NHS Constitution for England (2013, 2015)

"The NHS also commits: …

to anonymise the information collected during the course of your treatment and use it to support research and improve care for others (pledge); ...

to inform you of research studies in which you may be eligible to participate (pledge)"

Overview

CPFT clinical records

Visible to clinicians.

Researchers can only see records of patients who have given their permission for this.

Mrs Sarah JONES
Born 28 August 1937, female
6 The Mews, Fictionville
CPFT number 54321

Alzheimer's disease
Taking donepezil 5 mg at night
Memory score 23 this week

Mr John SMITH
Born 21 April 1970, male
13 Westport Lane, Doodletown
CPFT number 65432

Depression
Taking venlafaxine 150 mg per day
Much better, but one panic attack last month

Removing information that could identify an individual (such as names, addresses, exact dates of birth, and hospital/NHS numbers)

CPFT research database

Accessible only to researchers conducting research approved by CPFT.

XXXXXXXX
Born **XXX** August 1937, female
XXXXXXXX
CPFT number **XXXXXXXX**
Research ID 68926
Alzheimer's disease
Taking donepezil 5 mg at night
Memory score 23 this week

XXXXXXXX
Born **XXX** April 1970, male
XXXXXXXX
CPFT number **XXXXXXXX**
Research ID 57265
Depression
Taking venlafaxine 150 mg per day
Much better, but one panic attack last month

Opt-out system.

Within NHS secure network.

Our phase 1 (2013→): CRIS

- **Software**
- CRIS (Clinical Records Interactive Search), from NIHR Biomedical Research Centre at KCL/South London & Maudsley NHS FT (Stewart 2009).
- **Safeguards**
- NHS ethics, Caldicott Guardian, and CPFT R&D approvals.
- Research Database Oversight Committee with service user representation.
- **Technical**
- CPFT records 2005–2012 (“CRS/CDL”) anonymised into the research database; then newer electronic records system (RiO data 2013–).
- Research Database manager appointed.
- **Research and testing**
- Pilot studies testing different aspects of CRIS functionality.
- Epidemiological research.
- Audit.
- “Consent-to-contact” methods.
- Feasibility enquiries for commercial studies.

Drug treatment of schizophrenia and admissions, 2005–2012

- Search on CPFT's de-identified clinical records
- Letters to GPs, discharge summaries, care plans etc.
- Free text plus simple demographic standardised data (age, sex), diagnoses, admissions

- Schizophrenia was defined as the presence of an ICD-10 code of 'F20*' at any point in a patient's recorded diagnoses

- Database queries to examine admission records
- **Natural language processing** tool (Sultana 2014) to generate drug histories

- **~150,000** patients in the CPFT database for that time period
- **~15,000** (about 10%) had at least one coded ICD-10 diagnosis.

- **1,485** patients had a coded diagnosis of schizophrenia.
- Sex ratio 1.73 male : 1 female.
- Age range 13–93

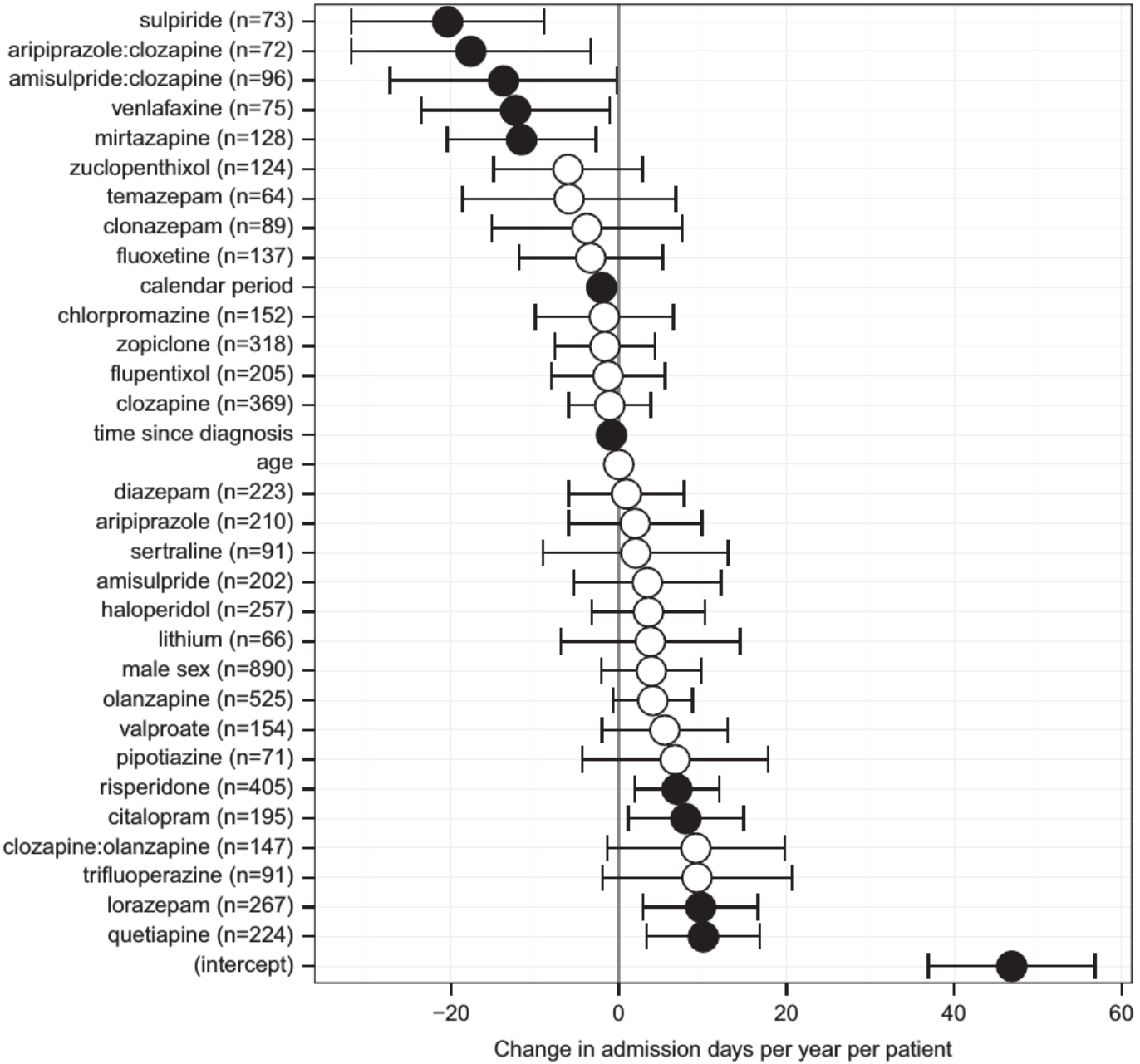
ARTICLE OPEN

Association between antipsychotic/antidepressant drug treatments and hospital admissions in schizophrenia assessed using a mental health case register

Rudolf N Cardinal[1,2], George Savulich[1], Louisa M Mann[2] and Emilio Fernández-Egea[1,2]

npj | Schizophrenia

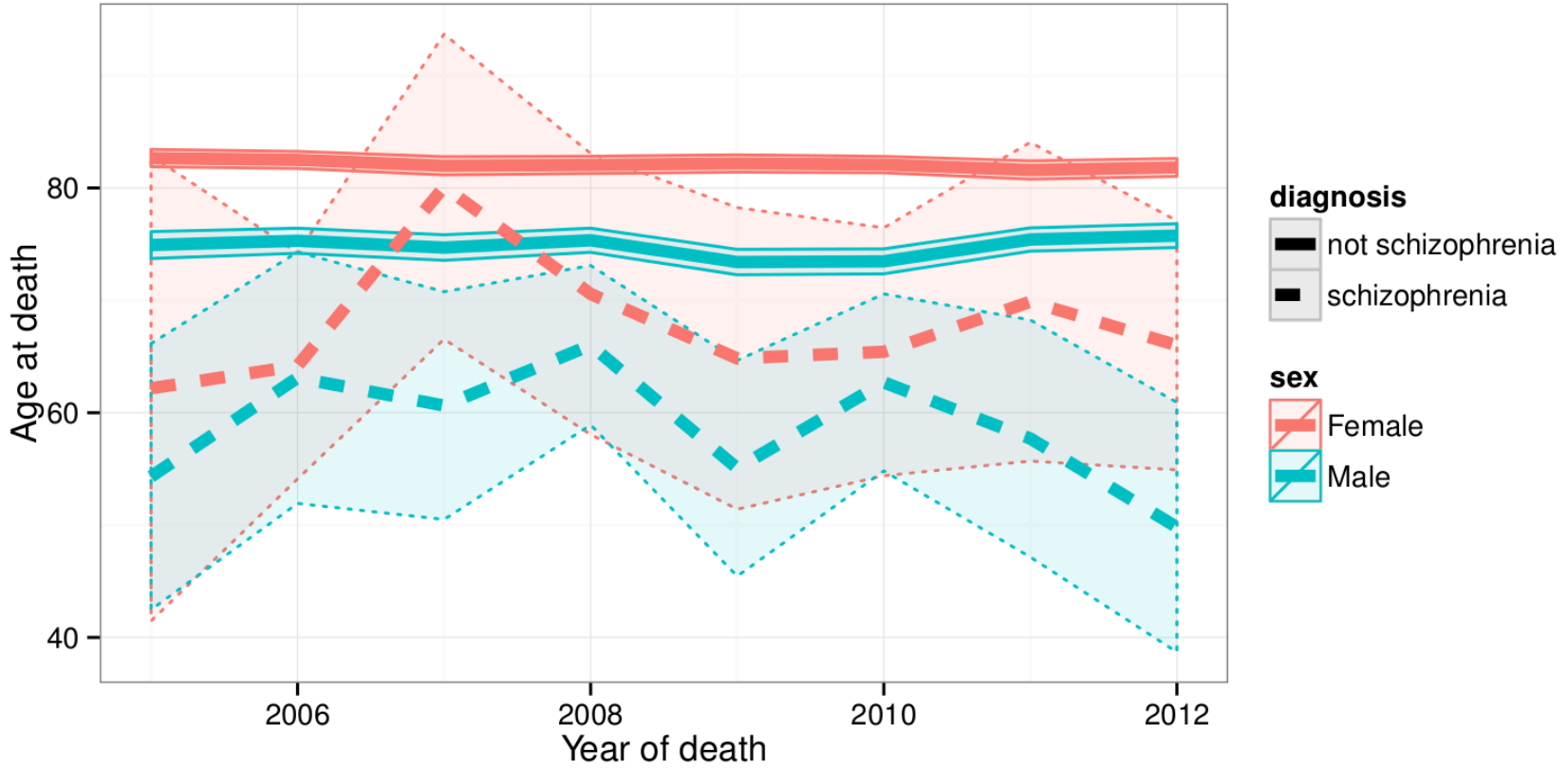
- Divide time into 6-month periods.
- Dependent variable = #days spent in hospital.
- Predictors include: drug Y/N in preceding period [via NLP]; subject; time; sex; age.



Highlighting public health concerns: life expectancy in schizophrenia

- Average age at death for all other CPFT service users:
  - male 74.8, female 82.1
- Average age at death for CPFT patients with schizophrenia:
  - male 59.0, female 67.5

- Overall, life expectancy is reduced by about 16.8 years for people with schizophrenia compared to other CPFT service users

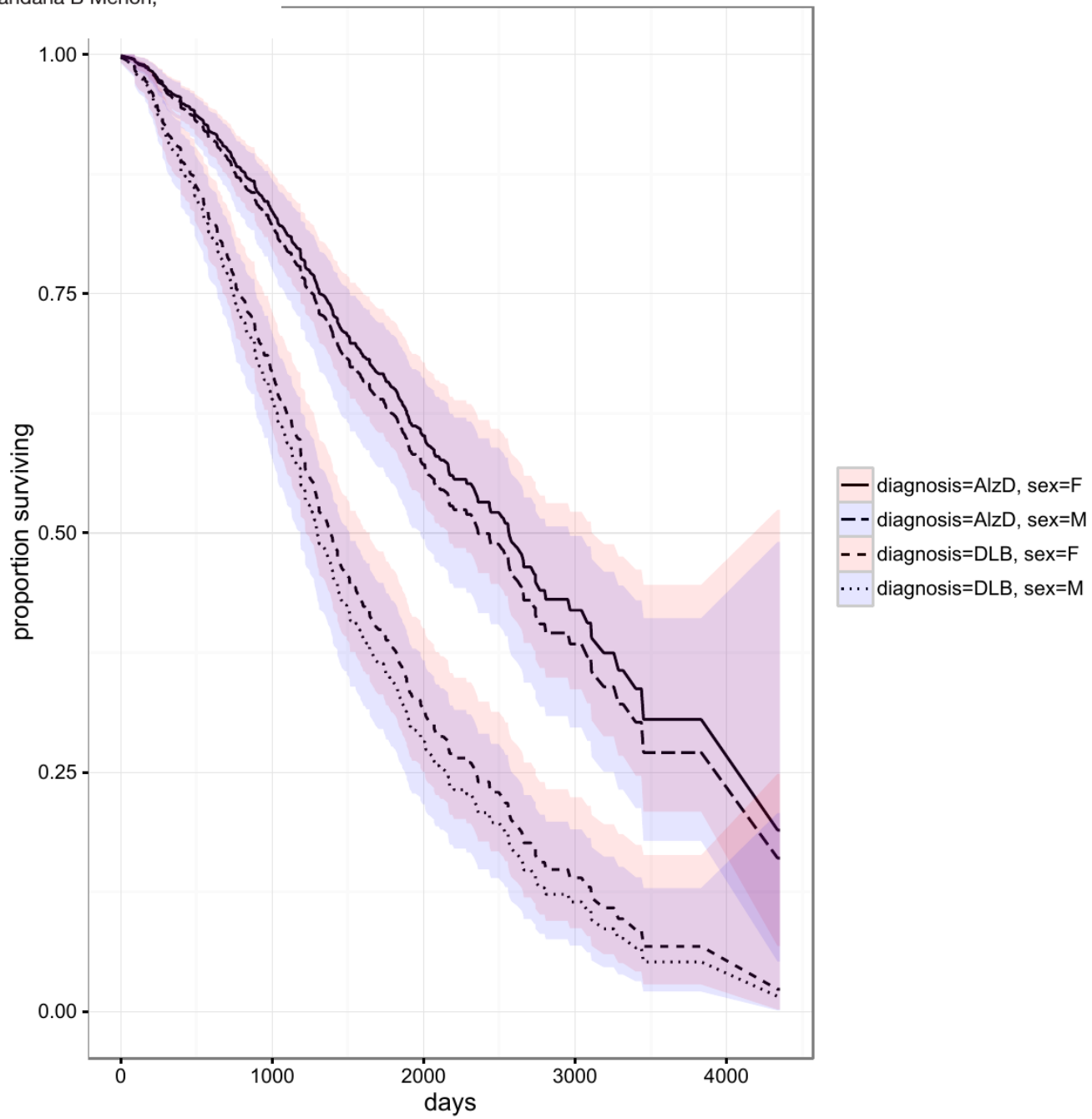


BMJ Open

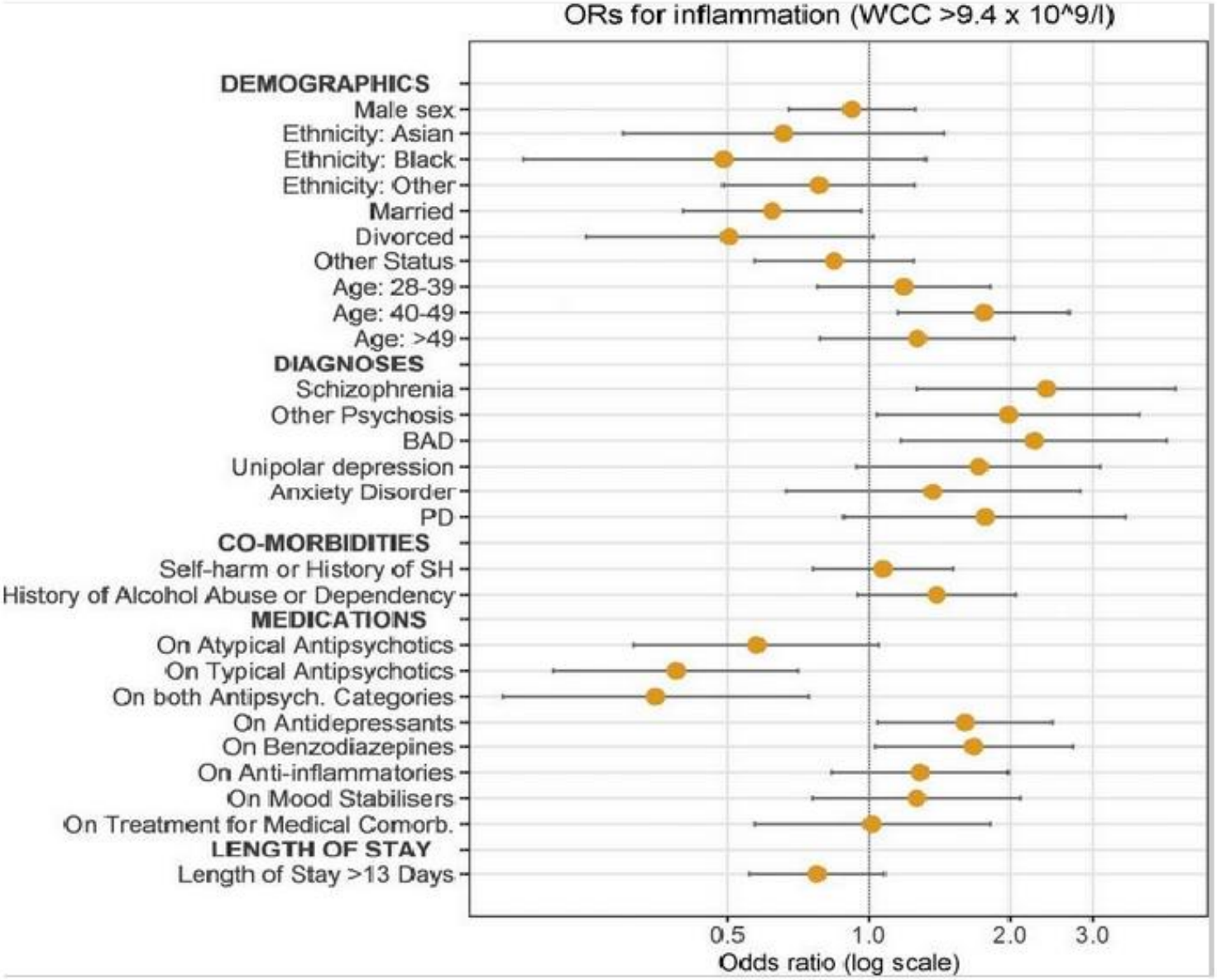
Mortality in dementia with Lewy bodies compared with Alzheimer's dementia: a retrospective naturalistic cohort study

Annabel Price,[1,2] Redwan Farooq,[3] Jin-Min Yuan,[3] Vandana B Menon,[4] Rudolf N Cardinal,[1,2] John T O'Brien[1,2]

start = time of presentation with cognitive impairment (if unknown: date of diagnosis)



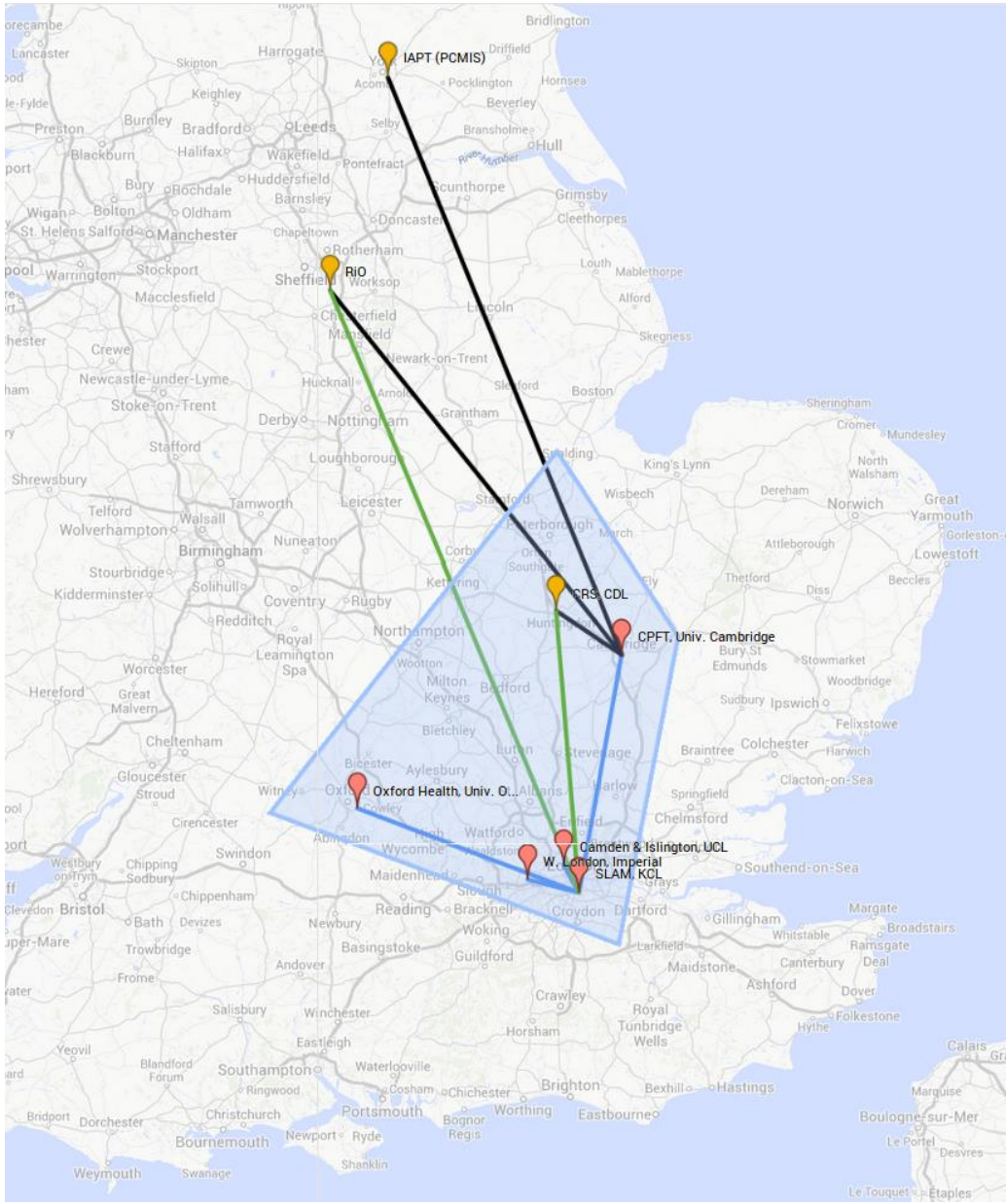
▼ Low-grade inflammation in adult inpatients on psychiatric wards (Osimo et al. 2018, *Psychoneuroendocrinology*).



Broader initiatives: universities/mental health trusts

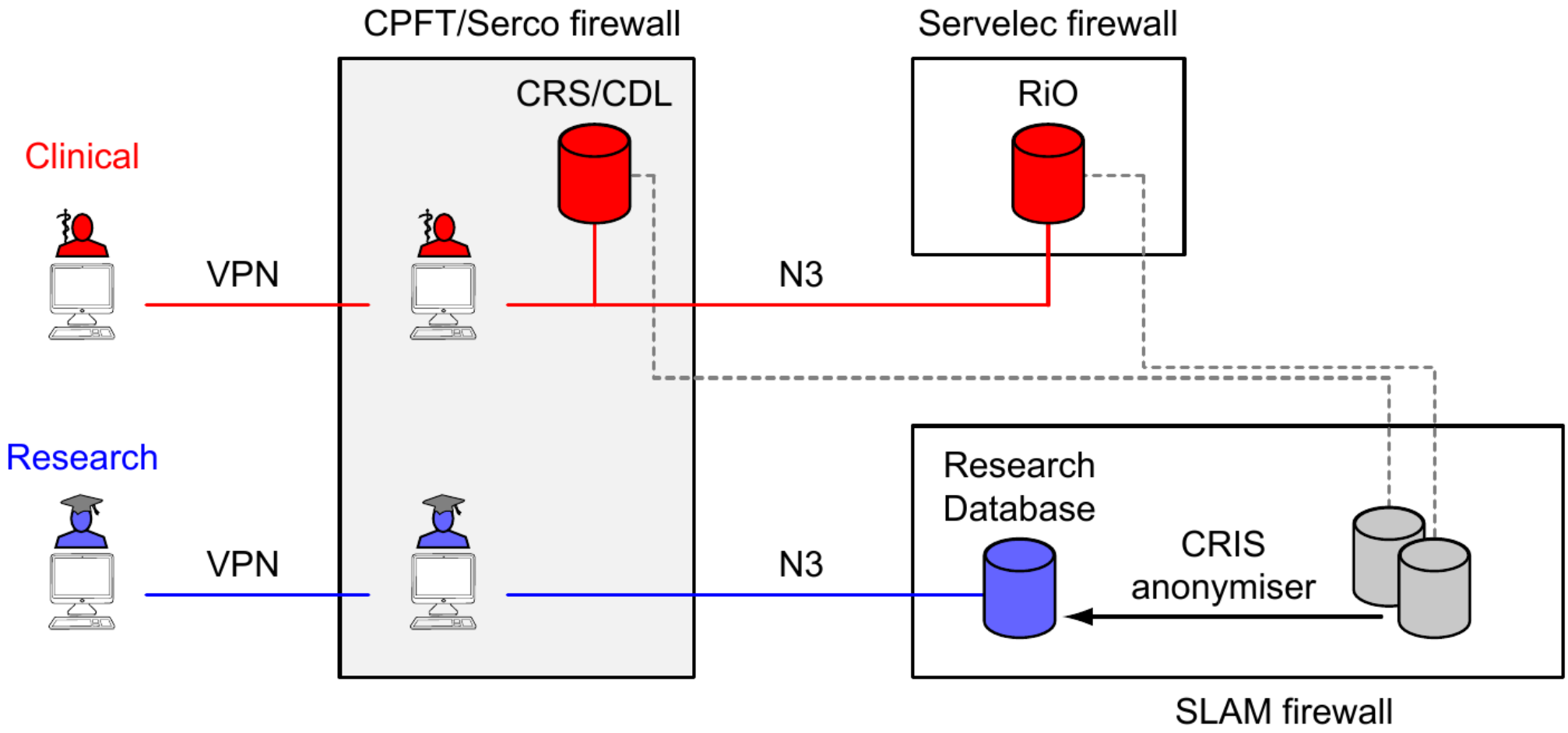
D-CRIS (Nov 2013→)

- SLaM/KCL, CPFT/Cambridge, Oxford, UCL and Imperial
- CPFT the first CRIS adopter outside SLaM
- £1m NIHR funding to increase database capacity and enhance infrastructure connections between centres



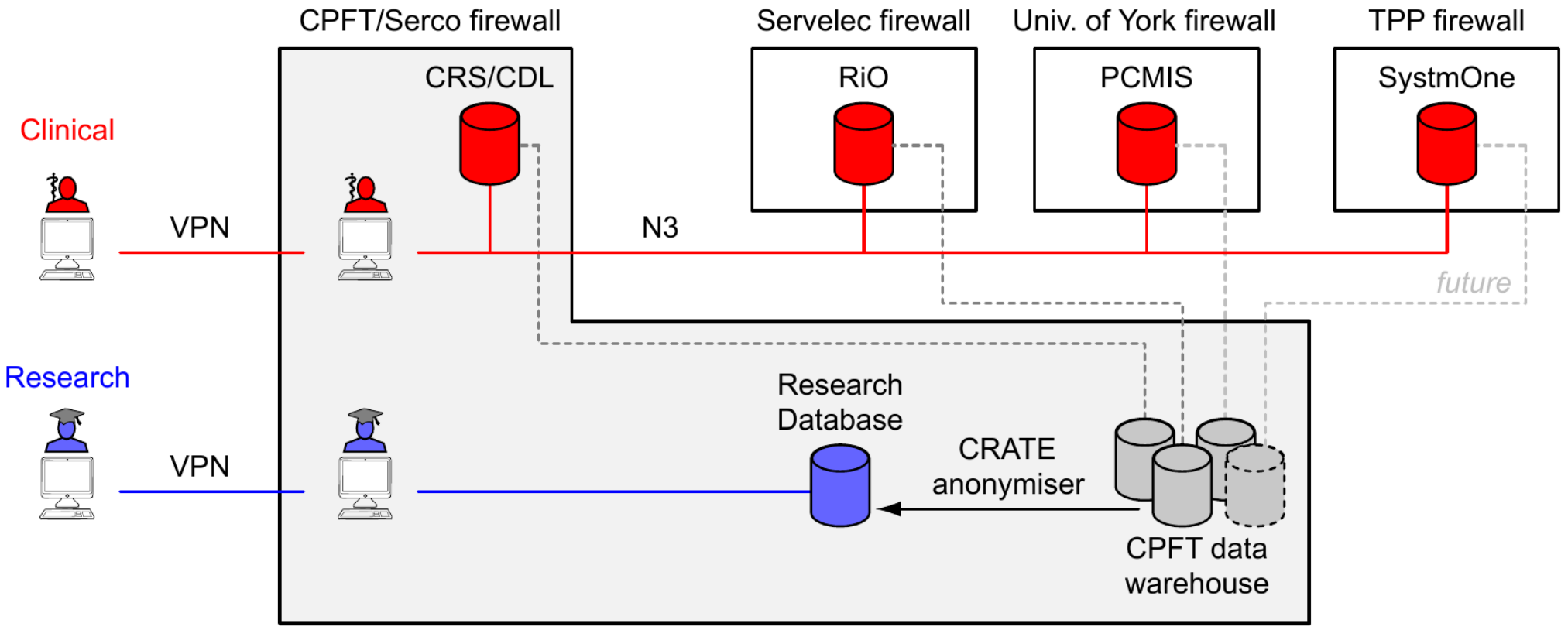
Phase 1 data flows

Via SLAM and CRIS



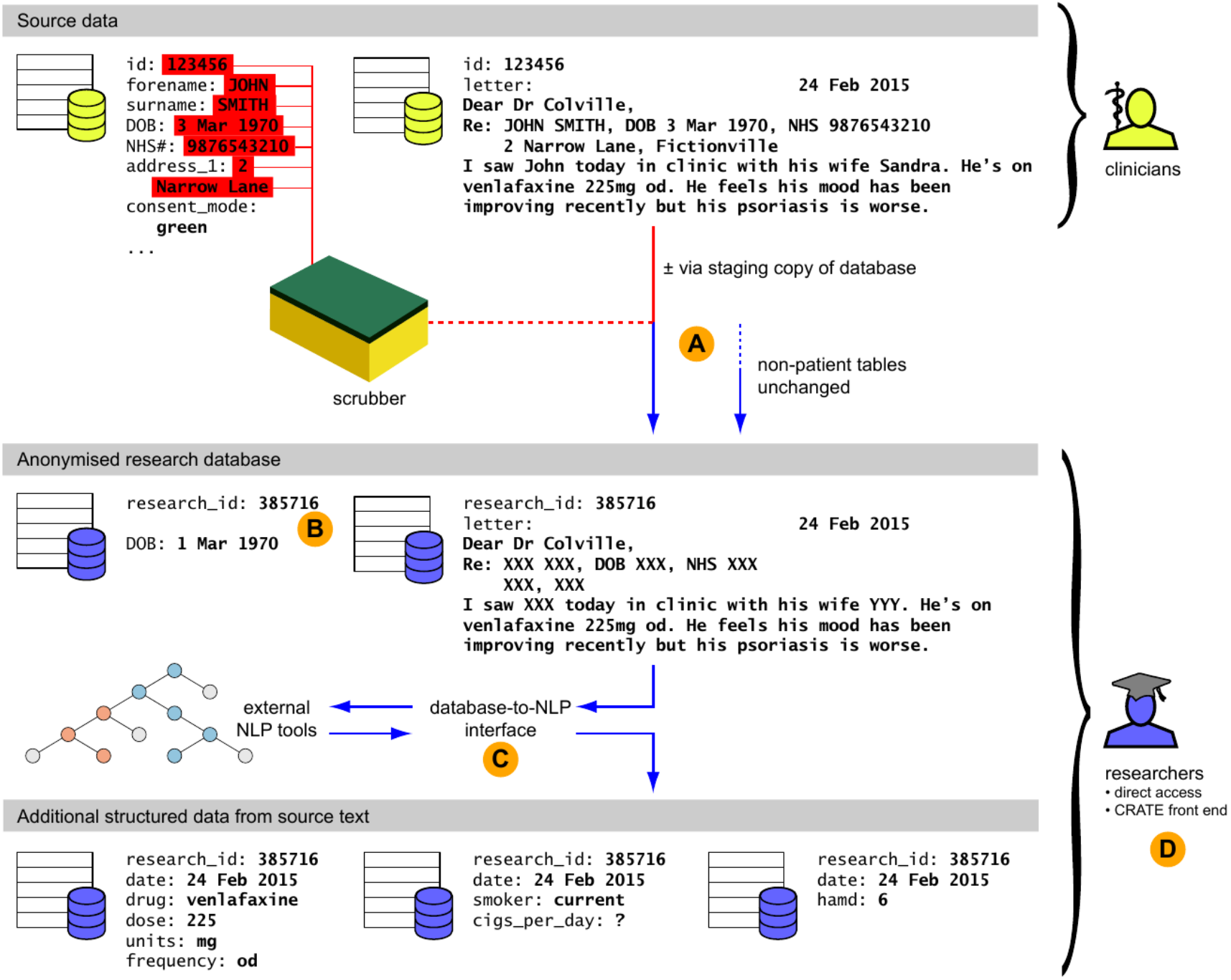
Our phase 2 (2017→): CRATE

In house via CRATE



Clinical records anonymisation and text extraction (CRATE): an open-source software system

Cardinal *BMC Medical Informatics and Decision Making* (2017) 17:50
DOI 10.1186/s12911-017-0437-1



CRATE

- Open-source. Python; some Java to talk to external Java NLP tools.
- Relational database, agnostic – e.g. SQL Server, MySQL, PostgreSQL.
- Preprocessors (e.g. for RiO) and auto-drafting of data dictionaries.
- Text extraction from attached documents – e.g. DOCX, PDF.

- Cryptographically secure research ID generation
- Standard NULL handling.
- Copes with typographical errors.
- Parallel processing.

- NLP interfaces and some custom NLP.
- Researchers can use (1) web front end with automatic SQL query builder, (2) raw SQL via web front end; (3) direct SQL connection.
- Our consent-to-contact system.

- Modular (e.g. anonymiser / NLP / research queries / consent-to-contact).

Table 2 De-identification performance

Metric	Condition 1	Condition 2	Condition 3
Number of words in source text (*n*)	50,274	50,274	50,274
Hits	1,392	1,326	1,116
False alarms	275	132	25
For known identifiers (those recorded as structured information in the source database):			
Misses	1	0	0
Correct rejections	48,606	48,816	49,113
Sensitivity = recall	0.999	1	1
Precision	0.835	0.909	0.978
For all identifiers (including those not recorded as structured information in the source database):			
Misses	127	125	128
Correct rejections	48,480	48,691	49,005
Sensitivity = recall	0.916	0.914	0.897
Precision	0.835	0.909	0.978

Performance of the de-identifier on the same corpus of clinical documents, with three different specimen configurations. The conditions differed in the definition of "identifying information" used, in whitelisting of geographical location, and in the method used for detecting fragments of addresses (see text; these differences lead also to variation in the number of hits counted, for example whether successful masking of an address such as "29 Acacia Avenue" was counted as one hit, if masked to "[___]", or several hits, if masked to "[___] [___] [___]"). A miss was defined as any identifier appearing in the destination text and identifiers were defined very liberally, including a single initial, so appearance of a single identifier in the destination text does not equate to identifying the patient concerned [13]

De-identification of unstructured text by CRATE.

Performance similar to published figures for CRIS (and similarly: primarily depends on source accuracy!).

Can already remove some kinds of generic information (e.g. postcodes, phone numbers).

Other techniques developed: e.g. "all names minus medical eponyms" blacklist.

Current databases

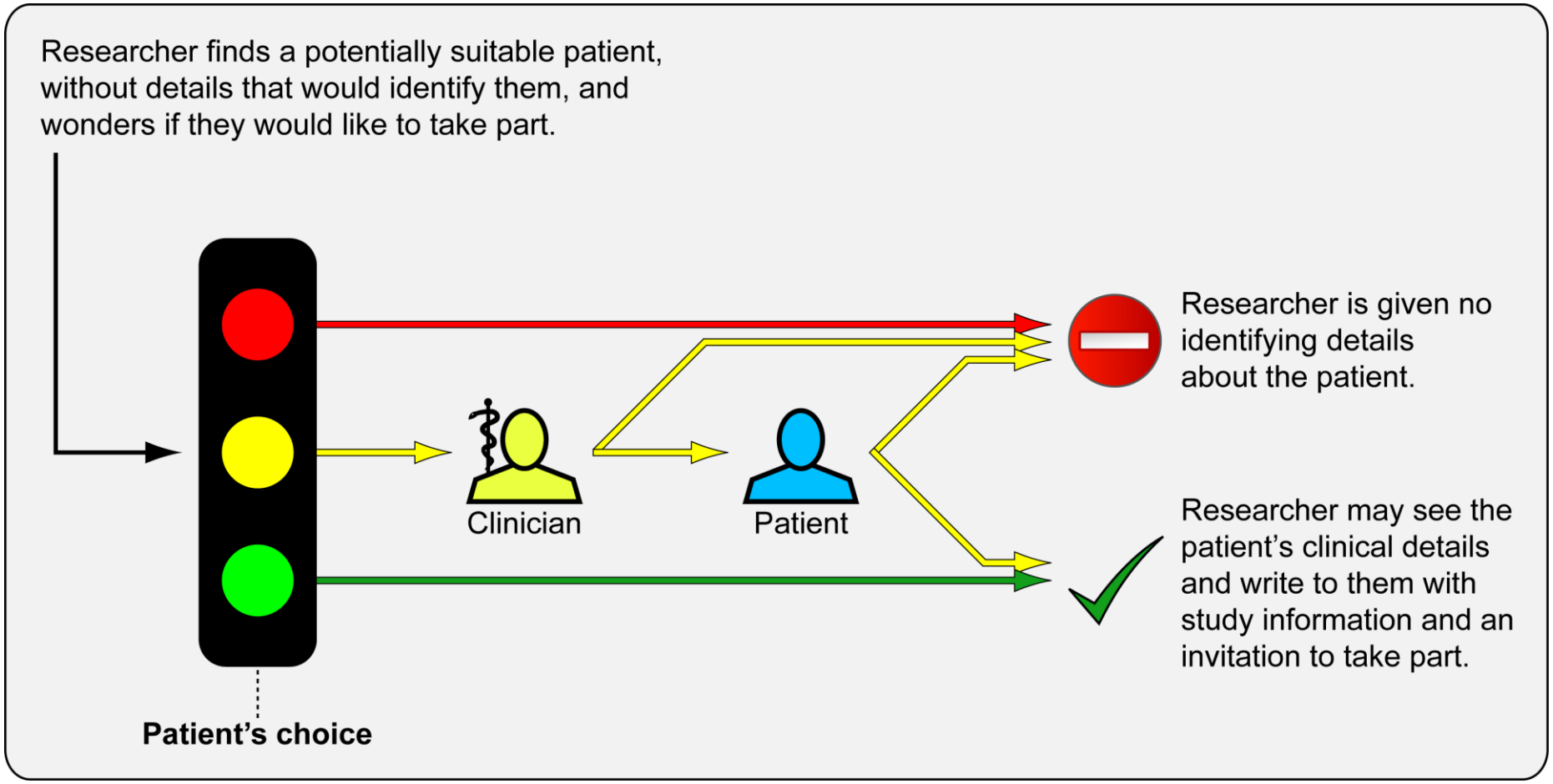
- CRS/CDL (Serco), 2005–2012
- ~160,000 patients; mental health secondary care
- 7 Gb anonymised; primarily demographics, admissions, documents

- RiO, 2013–
- ~185,000 patients and growing; mental health secondary care
- 120 Gb source data plus 662 Gb documents (not all useful!), growing fast; 36 Gb output

- PCMIS, 2008–
- >53,000 patients from psychological treatment service (70,000 referrals); mild to moderate depression/anxiety
- 392 Mb source; well structured

- SystmOne, 2016–
- *coming soon? Community services; high volume; supplier difficulties*

Patient-contact research



"Unknown" treated as YELLOW (and patient asked their traffic-light preference).
Special methods for children and those lacking capacity.
Safeguards, e.g. automatically exclude patients who have died; additional patient preferences.
Research study requires its own ethics approval. Study consent a later step.

CamCOPS: Cambridge Cognitive and Psychiatric Assessment Kit

- Clients (Android, Windows, Linux; soon iOS, OS/X) + institution-hosted web server.
- **Offline** operation, secure info transfer.
- **Clinical assessment:** clerking, progress notes, photos, referrals, diagnostic coding, discharge information...
- **Questionnaire-style tasks** (e.g. self-administered and clinician-scored rating scales, cognitive assessments).
- More complex **animated tasks** (e.g. 3D intradimensional / extradimensional set-shifting).
- Methods to deal with different kinds of copyright restrictions.
- Operates in **clinical and research environments** (e.g. identifiable versus pseudonyms).
- Definable group-level ID policies and security structure.
- Free (speech, beer). Cost = IT infrastructure (tablets, server, etc.).
- In use for clinical research and coming into routine clinical practice.

Client/server views

PATIENT, TEST (M). DOB: 23 Nov 1983. NHS: 2.

CamCOPS
Cambridge Cognitive and Psychiatric Assessment Kit
www.camcops.org

PATIENT, TEST (M)
Date of birth: **23 November 1983**
NHS number: **2**

Psychiatric clerking (Clerking)
Created: **28 November 2013, 23:15 +0000** (patient

Clinician's specialty:
Clinician's name:
Clinician's professional registration:
Clinician's post:
Clinician's contact details:
Current contact
Location
Transplant HDU
Contact type (e.g. admission, referral, outpatients, e
Referred by hepatology
Reason for contact (e.g. patient's reason, professior
Low mood following liver transplant
Presenting issue(s) (history of presenting complaint)
Mr Patient felt well prio OLT (donor/recipient C 6. Graft function is goo He reports low mood ar leave hospital. He think energy. He continues t occasionally by ward n some difficulty concent He has had no thought
Review of symptoms/sys
Unremarkable except s
Collateral history
Transplant team: doing Wife: never anxious be
Background
Diagnoses — psychiatric
Nil
Diagnoses — medical (p
Primary sclerosing cho Cirrhosis, portal hypert Eczema
Operations and procedur
Orthotopic liver transpl Ingrowing toenail excis
Allergies and adverse rea
NKDA

Page 1. Clerking created 28 N

PATIENT, TEST (M). DOB: 23 Nov 1983. NHS: 2.

CamCOPS
Cambridge Cognitive and Psychiatric Assessment Kit
www.camcops.org

Cambridgeshi

Underst

PATIENT, TEST (M)
Date of birth: **23 November 1983**
NHS number: **2**

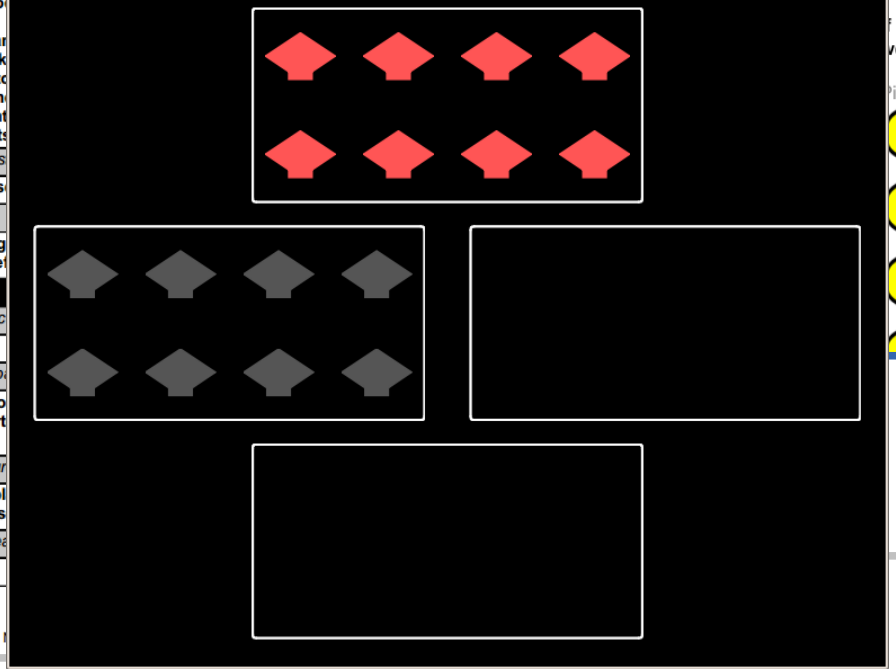
Patient Health Questionnaire-9 (PHQ-9)
Created: **28 November 2013, 23:14 +0000** (patient aged 30)

Completed?	**Yes**
Total score [1]	**6** / 27
PHQ9 depression severity: [2]	**mild**
Number of symptoms: core [3], other [4], total	**1**/2, **1**/7, **2**/9
PHQ9 major depressive syndrome? [5]	**No**
PHQ9 other depressive syndrome? [6]	**Yes**

Ratings are over the last 2 weeks.

Question	*Answer*
1. Little interest or pleasure in doing things	**1 — Sev**
2. Feeling down, depressed, or hopeless	**2 — Mor**
3. Trouble falling or staying asleep, or sleeping too much	**0 — Not**
4. Feeling tired or having little energy	**1 — Sev**
5. Poor appetite or overeating	**0 — Not**
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	**0 — Not**
7. Trouble concentrating on things, such as reading the	**2 — Mor**

camcops



Patient Health Questionnaire (PHQ-9)

Over the last 2 weeks, how often have you been bothered by any of the following problems?

	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	○	●	○	○
2. Feeling down, depressed, or hopeless	○	○	●	○
3. Trouble falling or staying asleep, or sleeping too much	●	○	○	○
4. Feeling tired or having little energy	○	●	○	○
5. Poor appetite or overeating	●	○	○	○
6. Feeling bad about yourself — or that you are a failure or have let yourself or your family down	●	○	○	○
7. Trouble concentrating on things, such as reading the newspaper or watching television	○	○	○	○
8. Moving or speaking so slowly that other people could have noticed? Or the opposite — being so fidgety or restless that you have been moving around a lot more than usual	○	○	○	○
9. Thoughts that you would be better off dead or of hurting yourself in some way	○	○	○	○

f you checked off any problems, how difficult have these problems made it for you to do your vork, take care of things at home, or get along with other people?

ick 1:

○ Not difficult at all

○ Somewhat difficult

○ Very difficult

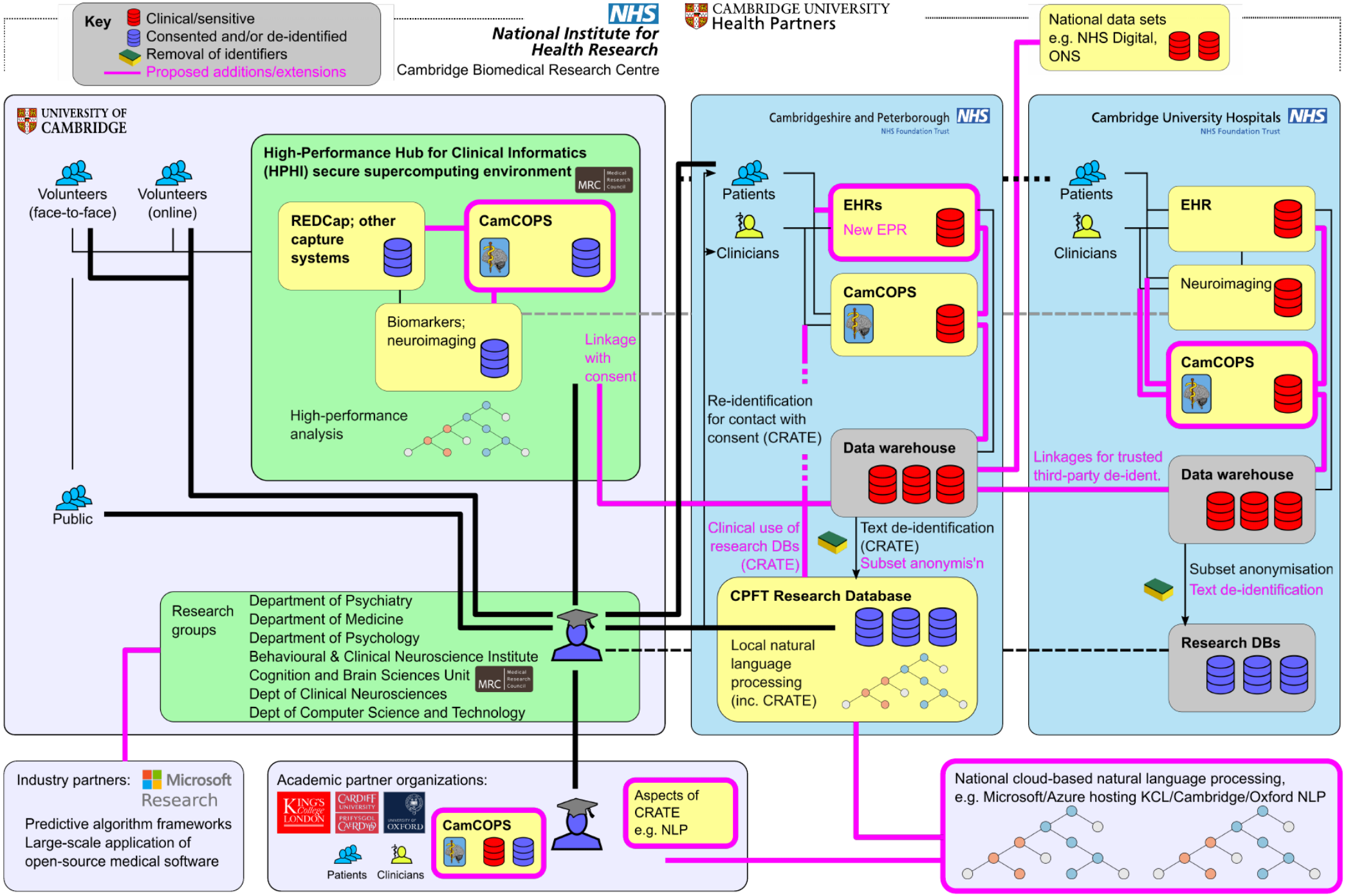
► **Alzheimer's Society** (O'Brien). Using medical records to diagnose dementia with Lewy bodies.
► **MRC Mental Health Data Pathfinder** (Cardinal). Cambridge themes:

1. Consolidating/extending the CPFT Research Database
► Linkage to ongoing cohort studies, and across mental/physical health.
► Cloud-based natural language processing (with KCL, Microsoft).

2. Improving the range and depth of structured clinical/cognitive data
► Integration of rapid, secure, structured data capture across research and clinical settings, ± large-scale online testing.
► Characterization of specimen cohorts, e.g. immunopsychiatry.

3. Tackling the mortality gap in serious mental illness
► Predict e.g. mortality using classical statistical and machine-learning approaches (with: Computer Lab, Microsoft Research UK).
► Inform interventions.

4. Democratizing mental health research
► Improved query tools/data visualization, inc. for clinician-/patient-led research.
► Patient/public involvement re best national system for consent for data use (for research and clinical purposes) and contact.


Key
Clinical/sensitive
Consented and/or de-identified
Removal of identifiers
Proposed additions/extensions
NHS
National Institute for Health Research
Cambridge Biomedical Research Centre
CAMBRIDGE UNIVERSITY
Health Partners
National data sets
e.g. NHS Digital, ONS
UNIVERSITY OF CAMBRIDGE
Cambridgeshire and Peterborough NHS
NHS Foundation Trust
Cambridge University Hospitals NHS
NHS Foundation Trust
High-Performance Hub for Clinical Informatics (HPHI) secure supercomputing environment
MRC
Volunteers (face-to-face)
Volunteers (online)
REDCap; other capture systems
CamCOPS
Biomarkers; neuroimaging
High-performance analysis
Linkage with consent
Public
Patients
Clinicians
EHRs
New EPR
CamCOPS
Re-identification for contact with consent (CRATE)
Data warehouse
Clinical use of research DBs (CRATE)
Text de-identification (CRATE)
Subset anonymis'n
CPFT Research Database
Local natural language processing (inc. CRATE)
Linkages for trusted third-party de-ident.
Patients
Clinicians
EHR
Neuroimaging
CamCOPS
Data warehouse
Subset anonymisation
Text de-identification
Research DBs
Research groups
Department of Psychiatry
Department of Medicine
Department of Psychology
Behavioural & Clinical Neuroscience Institute
Cognition and Brain Sciences Unit
MRC
Dept of Clinical Neurosciences
Dept of Computer Science and Technology
Industry partners:
Microsoft Research
Predictive algorithm frameworks
Large-scale application of open-source medical software
Academic partner organizations:
KING'S College LONDON
CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD
UNIVERSITY OF OXFORD
CamCOPS
Patients
Clinicians
Aspects of CRATE e.g. NLP
National cloud-based natural language processing, e.g. Microsoft/Azure hosting KCL/Cambridge/Oxford NLP